Décompositions en ondelettes redondantes pour le codage par descriptions multiples des images fixes et des séquences vidéo

Claudia Teodora Petrisor

École Doctorale
d'Informatique,
Télécommunications
et Électronique de Paris

Thèse

présentée pour obtenir le grade de docteur
de l'Ecole Nationale Supérieure des Télécommunications
Spécialité : Signal et Images

Claudia Teodora PETRIŞOR

Décompositions en ondelettes redondantes pour le codage par descriptions multiples des images fixes et des séquences vidéo

Soutenue le 28 Septembre 2009 devant le jury composé de

James Fowler	**Rapporteurs**
Marc Antonini	
Amel Benazza-Benyahia	**Examinateurs**
Gemma Piella	
Marco Cagnazzo	
Béatrice Pesquet-Popescu	**Directeurs de thèse**
Jean-Christophe Pesquet	

Remerciements

Comme tout doctorant devenu docteur, voici venu le moment d'exprimer ma gratitude à tous ceux qui y ont contribué. La liste est très longue, car, avant tout, ces années ont représenté pour moi une experience humaine.

J'adresse mes premières remerciements à mes deux directeurs de thèse, Béatrice Pesquet-Popescu et Jean-Christophe Pesquet sans qui ce travail n'aurait jamais abouti. Un immense merci Béatrice pour les conseils scientifiques, pour la confiance accordée, pour la valeur humaine de nos relations, pour ta grande patience et surtout pour m'avoir donné l'occasion de réaliser et de finir cette thèse. Monsieur Pesquet, je ne saurai pas exprimer en seulement quelques mots la gratitude que je vous porte pour tout ce que j'ai pu apprendre à votre côté, pour votre pédagogie et votre rigueur scientifique, ainsi que pour votre patience. Merci !

Merci à Amel Benazza pour avoir accepté de presider mon jury et à mes deux rapporteurs, James Fowler et Marc Antonini, qui m'ont fait l'honneur de lire le manuscrit et de me faire part de leur commentaires et questions fort intéressants. Merci également à Gemma Piella et Marco Cagnazzo pour leur questions et leur disponibilité.

Je dois, bien évidemment, des remerciements à Yves Grenier pour m'avoir accueilli à TSI durant toutes ces années, ainsi qu'à la direction de l'Ecole Doctorale pour les multiples re-inscriptions. Tout particulièrement, un grand merci à Florence Besnard pour l'organisation impeccable et pour l'accueil chaleureux que vous m'avez réservé à chaque occasion. A ce point, je ne peux m'empêcher d'avoir une pensée reconnaissante pour Stéphane Bonenfant. Ce fut un honneur.

Je ne peux oublier non plus Laurence Zelmar et Patricia Friedrich sans qui ma vie administrative aurait été bien plus difficile. Merci pour votre aide et votre disponibilité dans des moments clé.

Ensuite, merci à toute l'équipe du L2TI de l'Institut Galilée pour m'avoir accueilli parmi eux durant mes deux années d'ATER. Plus particulièrement je voudrais remercier Henri-François Raynaud, Caroline Kulcsar et Jean-Pierre Astruc pour notre collaboration intéressante et agréable, et pour votre convivialité.

Et, restant toujours dans les multiples milieux professionnels de mes années de thèse, il faut que j'adresse mes remerciements à tous les membres du Laboratoire des Systèmes Embarqués de THALES Research and Technologies pour leur excellent accueil et la transition facile de l'académie vers l'industrie. En particulier, un grand merci à mon co-bureau Sami Yehia pour son soutien, ses conseils et son amitié. Merci aussi à Arnaud, Jimmy, Sylvain et Nicolas S. pour m'avoir soutenu en direct.

Durant ces nombreuses années de thèse mes collègues de Télécom et de Marne la Vallée sont devenus mes amis. Bien que cette page est trop étroite pour exprimer la sympathie que j'ai pour

tous et que chacun mériterait son propre paragraphe, je vais vous prier de m'excuser d'avance pour tout oubli et je vais remercier en désordre au moins à un sous-ensemble[1]: Greg, Maria, Lionel, Yiol, Tyze, Ismaël, Thomas M., Claudio, Laurent O., Cyril C., Benoît, Bertrand, Joujou, Nancy, Valentin, Nathalie, Sarah, Cléo, Chloé, Rémi, Nelly, Mounir, Ileana, Lotfi, Laurent D., Miguel, Loïs, Zaïd, Clara,... ainsi que Aurelia, Caro et Jérôme - mon trio Télécom-MLV-ien et ma cellule de soutien psychologique et scientifique préférée. Un énorme merci à vous tous pour n'avoir un seul instant cessé de m'encourager !

Quant à vous, vous le savez déjà, mais je n'aurais de toute façon pas fini de vous remercier : merci Christophe pour les quantités phénoménales de mousse de canard, merci Fabricio pour tous les rires et la garde de Sarmisegetuza, merci Mathieu pour nos longues discussions et ton amitié, merci Tomtom et Arnaud pour votre soutien, merci Hachef pour avoir été là quand j'ai eu le plus besoin.

Un énorme merci à mes compatriotes préférés pour tous les repas copieux et pour ne pas m'avoir abandonné malgré toutes les fois ou j'ai manqué le rendez-vous.

Et puis le grand merci à mes parents Constantin et Mariana, grace à qui la vie est si facile. Merci à mon papa pour nos longues discussions nocturnes et pour m'avoir ouvert l'intérêt pour ce chemin et à ma maman pour le soutien inconditionnel, la joie de vivre et la passion apportés depuis le début. Merci également à ma deuxième famille, notamment Marie-Hélène et Pierre, pour vos nombreux encouragements et votre aide.

Pour finir, merci Jean, pour ton éternelle bonne humeur et confiance durant cette interminable "dernière ligne droite"... et pour tout le reste.

[1]pour ceux qui se retrouvent sous "..." je préparerai des remerciements personnalisés sur simple demande.

Contents

Glossary

4CIF: Four CIF: image resolution format of 706×576 pixels.

ARQ: Automatic Repeat on Request.

CIF: Common Interchange Format: image resolution format of 352×288 pixels.

DCT: Discrete Cosine Transform.

DWT: Discrete Wavelet Transform.

EBCOT: Embedded Block Coding with Optimized Truncation: still image coder.

EZBC: Embedded Zeroblock Coder: still image coder.

FB: Filter Bank.

FEC: Forward Error Correction.

FIR: Finite Impulse Response: class of filters with no internal feedback, whose impulse response goes to zero within a finite number of samples.

fps: Frames per second (measuring unit for the frame rate of a video sequence).

GOF: Group of Frames in a video sequence.

GOP: Group Of Pictures in a video sequence.

H.264: Video compression standard

HVBSM: Hierarchical Variable Size Block Matching - motion estimation algorithm.

IIR: Infinite Impulse Response.

JPEG: Joint Photographic Experts Group: organization and compression standard.

JPEG2000: new JPEG compression standard based on wavelets.

MC: Motion Compensation.

MC-EZBC: Motion Compensated EZBC: video coder.

MDC: Multiple Description Coding.

MDSQ: Multiple Description Scalar Quantization.

MDVQ: Multiple Description Vector Quantization.

MDCT: Multiple Descriptions with Correlating Transform.

MDTC: Multiple Description Transform Coding.

ME: Motion Estimation.

MIMO: Multiple-Input Multiple-Output.

MSE: Mean Squared Error.

MV: Motion Vector.

OFB: Oversampled Filter Bank.

PCT: Pairwise Correlating Transform.

pel: Pixel (abbreviation used in motion estimation/compensation operations to indicate pixel precision, such as full pel, half-pel or 1/8-pel. The latter two indicate sub-pixel accuracy).

pixel: Picture element.

PSNR: Peak Signal to Noise Ratio.

PZW: Packetized Zerotree Wavelet.

QCIF: Quarter CIF: image resolution format of 176×144 pixels.

QoS: Quality of Service.

RDWT: Redundant Discrete Wavelet Transform.

SPIHT: Set Partitioning Into Hierarchical Trees.

SVC: Scalable Video Coding

UEP: Unequal Error Protection.

Y-PSNR: PSNR on the luminance component, Y, of a signal.

Abstract

The increasing usage of the Internet and other best-effort networks for diverse multimedia communications, brings with it a stringent need for reliable transmission. For a long time, the research efforts have been concentrated on enhancing the existing error correction techniques, but during the last decades an alternative solution has emerged and gained more and more popularity. This solution mainly answers to the situation in which immediate data retransmission is either impossible (network congestion or broadcast applications) or undesirable (e.g. in conversational applications with very low delay requirements). We are referring to a specific joint source-channel coding technique known as Multiple Description Coding (MDC).

Multiple description coding builds several correlated but independently decodable (preferably with equivalent quality) bitstreams, called descriptions, that are to be sent over as many independent channels. In an initial scenario these channels are working in a binary manner, in other words, if an error occurs on one channel this is considered entirely damaged and the conveyed bitstream is unusable at the so-called side decoder end. As in other robust coding methods, some amount of redundancy has to be added to the source signal, such that an acceptable reconstruction can be achieved from any of the bitstreams. Then, similar to layered coding, the reconstruction quality will be enhanced with every bitstream received, maximal reconstruction quality being attained at the so-called central decoder. The major difference with layered coding is that all "layers" have equal importance in MDC.

This thesis focuses on new approaches to Multiple Description Coding in low-redundancy scenarios. We will present their application to the transmission of still images and video sequences. To this end, we have proposed new schemes based on wavelet frame decompositions, which, for computational convenience, are implemented in a lifting form.

We first study new methods of building two descriptions in the temporal axis of a "t+2D" video codec. The redundancy of the schemes is inherent to the wavelet frame transform which is equivalent to an oversampled filter bank. However, keeping the whole set given by this decomposition would yield a redundancy of a factor of 2 which could be highly inefficient if both paths were error-less. In our schemes we perform an additional subsampling of the detail subbands while keeping the obtained approximation subbands entirely. Thus the redundancy is tuned to the size of an approximation subband in a classical wavelet decomposition. However this raises a new problem which is the perfect reconstruction of such a scheme. In this part we have proven the perfect reconstruction for certain schemes and we have established choice criteria among them based on the minimization of the quantization noise. We have compared the performances of several schemes among the efficient ones in a scalable video coding context provided by the MC-EZBC (Motion Compensated - Embedded Zero-trees Block Coding) codec. Two scenarios - losing a whole descriptions versus losing only packets in each description - have been implemented and the results have been compared to the classical critically

sampled decomposition.

A second direction that we have explored in this thesis refers to the MDC of still images, which is viewed as an extension of the temporal schemes developed earlier. The problem of structure invertibility is not trivial in the two-dimensional schemes and an exhaustive study has been conducted in order to select the efficient schemes among all possible combinations based on the proposed subsamplings. Moreover, we have explored the possibility of improving the decoding by a post-processing based on a priori information on the system. This information is given by the quantization steps which can be viewed as convex constraints. The reconstruction problem has thus been formulated as the optimization of a quadratic function under convex constraints and the decoded image gains several dB in terms of Peak Signal to Noise Ratio (PSNR) both when a whole description is lost and when random pixels in each description are destroyed.

Finally, we have approached the Multiple Description Problem from a completely different angle, by considering the problem as a rate-distortion optimization in which a certain maximal distortion is allowed, and the best transmission rate (linked to the quantization step) is searched. In general this is a non-convex problem hard to solve, therefore some approximations have been formulated in order to be able cast it as a convex optimization. The scheme has also been generalized to an arbitrary number of descriptions. By doing this the complexity is shifted to the encoder whereas the decoding becomes a simple linear process. We have tested this theory for still images. This method is inspired from the recently developped theory named "Compressed Sensing" aiming at exploiting the sparsity present in signals in order to solve inverse problems.

Résumé de la thèse

Introduction, Motivations

Considérons le schéma de transmission simplifié présenté dans la Figure 1. Ce schéma illustre la transmission d'un signal numérique, x, sur un réseau. Habituellement, entre le transmetteur et le récepteur se trouvent plusieurs blocs de traitement, comme : une transformée ayant le rôle de mieux préparer le signal pour la transmission (en réalisant une meilleure localisation de son énergie, par exemple), un quantificateur qui impose un nombre fini de valeurs et, finalement, un encodeur allouant une séquence binaire à chacun des échantillons à transmettre. Ensuite les bits sont regroupés dans des paquets, en formant de cette manière le flux de données, qui sera transmis sur un canal réseau. Le canal est habituellement sujet à des perturbations pouvant altérer le flux de données et donc empêchant la réception correcte de celui-ci. Le signal reçu est dénoté ici par $\hat{x}$. Bien évidemment, le but majeur de la transmission est de minimiser l'erreur $\epsilon = \|x - \hat{x}\|$ entre le signal envoyé et celui décodé. Idéalement les deux signaux devraient être identiques, ou alors leur difference doit se situer dans une marge de tolérance.

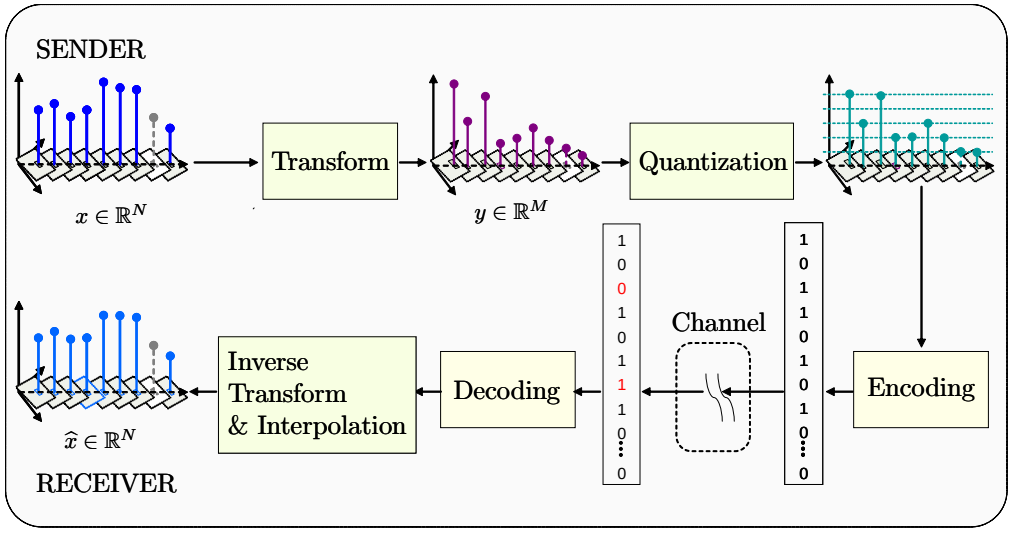


FIGURE 1: Schéma de transmission de données simplifié.

Il existe plusieurs stratégies pour aborder ce problème. Une première consiste à retransmettre les paquets corrompus, mais cette méthode présente le désavantage d'introduire d'importants délais qui ne sont pas acceptables dans toutes les circonstances. Une autre technique consiste à envoyer un flux de données plus important, englobant un code correcteur d'erreur quelconque. Néanmoins, ceci est limité à quelques bits erronés par paquet unique-

ment. Cette limitation est palliée avec une stratégie de codage en couches ("layered coding" en anglais) qui consiste à transmettre un flux de données affinant successivement une couche de base. La couche de base est tout de même essentielle pour pouvoir reconstruire les données avec une qualité minimale. En conséquence, si la couche de base est touchée par des erreurs de transmission, nous sommes ramenés au problème initial. Bien sûr, un mélange de ces méthodes est facilement envisageable, et des techniques de protection pour la couche de base ont été employées.

L'usage intensif de l'Internet et/ou d'autres réseaux à pertes de paquets pour de la communication multimédia impose la nécessité d'une transmission fiable. Les recherches se sont pendant longtemps axées sur le perfectionnement des codes correcteurs d'erreurs existants, mais les dernières décennies ont connu l'apparition d'une solution alternative devenue de plus en plus populaire. Cette solution est amenée à répondre a un scénario de fonctionnement qui ne permet pas la retransmission de l'information erronée (par exemple, à cause de goulots d'étranglement ou pour des applications de type diffusion) ou alors cette dernière n'est pas désirée (par exemple pour des applications conversationnelles avec des exigences de délais très réduits). Cette technique est connue sous le nom de *Codage par Descriptions Multiples* (en anglais Multiple Description Coding - MDC), méthode qui s'apparente au codage conjoint source-canal.

La MDC se définit par la construction de flux de données, appelés *descriptions*, qui peuvent être décodés indépendamment les uns des autres et, de préférence, avec une qualité équivalente. Ces flux de données sont transmis sur des canaux séparés à fonctionnement binaire : ils sont considérés défaillants (et donc inutilisables) si une erreur est survenue lors de la transmission, autrement ils sont considérés à fonctionnement parfait. L'enjeux principal dans la construction des descriptions est de réaliser le meilleur compromis en termes de redondance (et donc débit supplémentaire) et distorsion obtenue après décodage. Une distorsion maximum acceptée peut être atteinte par chacune des descriptions, et cette distorsion doit diminuer avec chaque nouvelle description correctement reçue. Ceci a des similarités avec le codage en couches, mais la différence majeure vient du fait qu'il n'existe pas de couche de base qui soit vitale pour le décodage.

La formalisation de ces techniques est apparue à la fin des années '70 dans la communauté de la théorie de l'information. Les premiers résultats étaient à caractère théorique et ils donnait la région des distorsions atteignables pour une source sans mémoire, décrite avec plusieurs flux de données à débit fixé. Une vingtaine d'années plus tard l'utilité de ces résultats pour la transmission de données sur des réseaux à pertes à donné naissance aux premières méthodes pratiques de construction de descriptions multiples. Si l'on se réfère à la Figure 2, représentant les principales étapes lors de l'encodage d'une source, nous identifions autant de points d'entrées pour les méthodes par descriptions multiples.

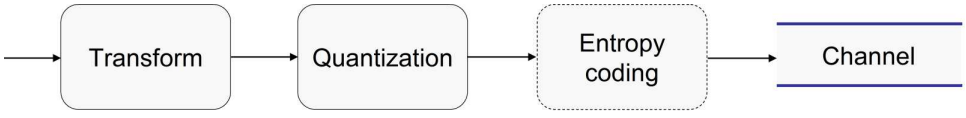


FIGURE 2: Points d'entrée pour le codage par descriptions multiples.

Plus précisément, il existe deux classes principales de méthodes, distinguées par l'endroit ou la redondance est introduite dans la chaîne d'encodage. Chronologiquement, la separation en descriptions a été introduite lors de l'étape de quantification, en décalant le intervalles de

manière à obtenir une representation quantifiée plus finement si tous le canaux ont été fonctionnels (ce que l'on appelle "décodeur central") et, respectivement plus grossièrement si seulement un sous-ensemble de canaux a pu être exploité (la reconstruction est faite au niveau du décodeur dit "lateral") [Vai93b]. La deuxième grande direction pour la construction des descriptions multiples vise à introduire de la diversité au niveau de la transformée [WOR97], [Goy98],[GKV99]. Ici, deux principales stratégies ont été utilisées : les transformées qui rajoutent de la corrélation et les décompositions en trame. Nos travaux se placent dans cette dernière direction, et ils sont motivés par la robustesse au bruit additif sur le signal (tel que le bruit de quantification) que ce type de transformée garantit.

L'objectif de cette thèse est de construire des représentations par descriptions multiples, à faible redondance. Nous allons présenter leur application à la transmission des images fixes et des séquences vidéo. Pour ce faire, nous avons considéré des schémas basés sur des trames d'ondelettes, implantées en "lifting".

Dans un premier temps nous avons considéré la construction des deux descriptions selon l'axe temporel d'un codeur vidéo $t + 2D$. La redondance introduite est inhérente à la transformation utilisée (une trame d'ondelettes qui peut également être vue comme un banc de filtres sur-échantillonnés), cependant elle est d'un facteur deux et donc trop élevé si les deux canaux ont fonctionné correctement (les deux descriptions ont été reçues). Nous avons proposé une réduction de cette redondance par une étape de sous-échantillonnage supplémentaire des sous-bandes de détails dans la décomposition. En procédant ainsi nous pouvons ajuster cette redondance à la taille d'une sous-bande d'approximation en ondelettes, issue d'une décomposition classique, échantillonnée critique. En revanche, ceci rajoute un problème de reconstruction parfaite que nous avons adressé en détail. En outre, nous avons établi des critères de choix entre les plusieurs schémas possibles, en introduisant un critère de minimisation du bruit de quantification. Nous avons applique ceci à la transmission de la vidéo avec un codeur MC-EZBC (Motion Compensated - Embedded Zerotrees Block Coding) via deux scénarios : la perte d'une description entière versus des pertes aléatoires de paquets dans chacune des deux descriptions.

Une deuxième direction explorée dans cette thèse est l'extension des méthodes ci-dessus au signal 2D, plus précisément à la transmission d'images. En employant une stratégie de décomposition similaire, suivie d'un sous-échantillonnage supplémentaire nous avons identifié beaucoup plus de schémas possibles. Pour tous ces schémas une étude d'inversibilité à été menée. En outre, nous avons proposé une méthode de post-traitement au niveau des décodeurs, visant à améliorer la qualité de la reconstruction en situation de pertes. Ce problème a été formulé comme la minimisation d'un critère quadratique sous des contraintes convexes.

Enfin, nous avons abordé le codage par descriptions multiples depuis un nouvel angle, en considérant un problème d'optimisation débit-distorsion imposant une distorsion maximale, tout en cherchant le débit (traduit par un pas de quantification) optimal pour la transmission. Ceci est une formulation non-convexe pour laquelle nous avons cherché des approximations permettant de la résoudre comme une optimisation convexe. Cette approche traite un cas de figure different des approches précédentes, en considérant la complexité de calcul uniquement possible à l'encodeur et en imposant des décodeurs très simples - linéaires. Cette méthode vise à exploiter le fait que la source admet une représentation creuse dans la trame considérée et elle suit des lignes similaires à la nouvelle théorie du "Compressed Sensing".

Dans la suite nous allons présenter schématiquement les grandes lignes de nos contributions ainsi que quelques résultats. Le détail des méthodes proposées se trouve dans les chapitres suivants.

Schémas à descriptions multiples temporelles

Comme annoncé précédemment, les schémas à descriptions multiples proposés dans cette thèse font partie des méthodes de codage basées sur transformée. En particulier, la transformée en ondelettes est un très bon candidat, grace à ses capacité de compression [FPP07] ainsi qu'à sa scalabilité. Dans le contexte des descriptions multiples, des transformées en ondelettes redondantes peuvent être mises en oeuvre.

Soit $L^2(\mathbb{R})$ l'espace des signaux réels à énergie finie. Nous allons construire des schémas à descriptions multiples basés sur une analyse multirésolution de $L^2(\mathbb{R})$ ou $l^2(\mathbb{Z})$ (pour le cas discret), [Mal89],[Mal98], [Mey90]. Ceci consiste à projeter le signal sur des bases de fonctions permettant d'obtenir des approximations de plus en plus grossières du signal. On définit une analyse multirésolution d'un signal comme la séquence des espaces vectorielles fermés : $\{V_j\}_{j\in\mathbb{Z}}$ of $L^2(\mathbb{R})$, satisfaisant:

$$\{0\} \subset \ldots \subset V_{j+1} \subset V_j \subset V_{j-1} \subset \ldots \subset L^2(\mathbb{R}), \tag{1}$$

$$\forall x \in L^2(\mathbb{R}) \quad \text{on a } x(t) \in V_j \Leftrightarrow x(\frac{t}{2}) \in V_{j+1}, \quad j \in \mathbb{Z}. \tag{2}$$

Alors, la projection d'un signal $x \in L^2(\mathbb{R})$ sur un sous-espace V_j représente une approximation de x au niveau de résolution j. En outre, il existe un fonction dite "d'échelle" $\phi \in L^2(\mathbb{R})$, telle que la famille des fonctions translatées $\{t \mapsto \phi(t-n)\}_{n\in\mathbb{Z}}$ forme une base orthonormale de V_0, et les versions dilatées et translatées de $\phi(t)$,

$$\left\{\phi_{j,n}(t), \quad n \in \mathbb{Z}\right\} \quad \text{avec } \phi_{j,n}(t) = 2^{-j/2}\phi(2^{-j}t - n)$$

forment une base orthonormale de V_j.

La difference entre les approximations du signal, obtenues dans deux sous-espaces successifs, V_j et V_{j+1}, représente l'information de détail perdue d'un niveau de résolution à l'autre. Cette information peut être obtenue à partir d'un espace de détail sous-jacent, W_{j+1}, orthogonal à V_{j+1}:

$$V_j = V_{j+1} \oplus W_{j+1}.$$

De même que pour les espaces d'approximation, une base orthonormale pour W_j est obtenue à partir des versions translatées et dilatées d'une fonction appelée "ondelette mère" $\psi \in L^2(\mathbb{R})$, données par :

$$\left\{\psi_{j,n}, \quad n \in \mathbb{Z}\right\} \quad \text{avec } \psi_{j,n}(t) = 2^{-j/2}\psi(2^{-j}t - n).$$

Ainsi, les coefficients d'approximation et de détail d'un signal $x \in L^2(\mathbb{R})$ sont donnés respectivement par :

$$a_{j,n} = \langle x, \phi_{j,n}\rangle = \int_{-\infty}^{+\infty} x(t)\phi_{j,n}(t)dt = \int_{-\infty}^{+\infty} x(t)2^{-j/2}\phi(2^{-j}t - n)dt,$$

$$d_{j,n} = \langle x, \psi_{j,n}\rangle = \int_{-\infty}^{+\infty} x(t)\psi_{j,n}(t)dt = \int_{-\infty}^{+\infty} x(t)2^{-j/2}\psi(2^{-j}t - n)dt.$$

A une résolution donnée l'ensemble de coefficients d'approximation est appelé "sous-bande" d'approximation et il en va de même pour les coefficients de détail, qui forment la "sous-bande" de détail.

Ainsi, un signal $x \in L^2(\mathbb{R})$ peut être représenté comme suit :

$$x(t) = \sum_{n\in\mathbb{Z}} a_{J,n}\phi_{J,n}(t) + \sum_{j\leq J}\sum_{n\in\mathbb{Z}} d_{j,n}\psi_{j,n}(t). \tag{3}$$

Notons la base engendrée par ces familles de fonctions par:

$$\mathcal{B}^{\mathrm{I}} = \{\phi_{J,n},\ n \in \mathbb{Z}\} \cup \bigcup_{j\leq J}\{\psi_{j,n},\ n \in \mathbb{Z}\},$$

où l'exposant I servira à distinguer les différentes bases employées par la suite. Dans cette thèse nous nous limitons au cas des ondelettes orthogonales ou biorthogonales.

Ce type de décomposition classique en ondelettes génère une représentation non-redondante et creuse du signal original, ce qui a été exploité dans des nombreux schémas de compression.

L'idée directrice pour la construction de nos schémas à descriptions multiples est d'employer une décomposition en trame d'ondelettes, obtenue comme une union de bases issue des différentes translations de $\mathcal{B}^{\mathrm{I}}$. Plus particulièrement, nous allons utiliser deux représentations, notées $\mathcal{B}^{\mathrm{I}}$ (comme ci-dessus) et $\mathcal{B}^{\mathrm{II}}$, où :

$$\mathcal{B}^{\mathrm{II}} = \{\psi'_{J,n}, n \in \mathbb{Z}\} \cup \bigcup_{j\leq J}\{\phi'_{j,n}, n \in \mathbb{Z}\}.$$

et

$$\{\psi'_{j,n} = \frac{1}{2^{j/2}}\psi(\frac{t}{2^j} + \frac{1}{2} - n), \quad n \in \mathbb{Z}\}$$
$$\phi'_{j,n} = 2^{-j/2}\phi(2^{-j}t + 1/2 - n), \quad \forall n \in \mathbb{Z}.$$

Il est important de remarquer que l'ensemble de coefficients d'approximation et de détail donné par $\mathcal{B}^{\mathrm{I}}$ et $\mathcal{B}^{\mathrm{II}}$ mène à une représentation à haute redondance - le nombre d'échantillons est doublé. En supposant que deux descriptions ont été formées chacune avec la moitié de cet ensemble, le supplément de débit engendré est souvent trop important par rapport à l'amelioration de la qualité de reconstruction quand les deux descriptions ont été reçues.

Pour pallier ce problème nous proposons de rajouter une étape de sous-échantillonnage supplémentaire sur les coefficients, à fin de réduire la taille de chaque description, et par consequent le débit total. Ce procédé permet d'ajuster la redondance jusqu'à la taille d'une sous-bande d'approximation (en terme de nombre de coefficients). Ainsi, la redondance introduite décroît avec le nombre de niveaux de décomposition employé.

Cette operation de sous-échantillonnage secondaire sera appliquée sur les sous-bandes de détail et nous allons l'indiquer par les caractères $\hat{.}$ and $\check{.}$, ce qui donne :

$$\hat{d}^{\mathrm{I}}_{J,n} = d^{\mathrm{I}}_{J,2n}, \tag{4}$$
$$\check{d}^{\mathrm{I}}_{J,n} = d^{\mathrm{I}}_{J,2n-1}. \tag{5}$$
$$\tag{6}$$

pour les coefficients de détail issus d'une décomposition sur $\mathcal{B}^{\mathrm{I}}$ au dernier niveau de résolution, J. Ici, n dénote l'indice temporel d'un signal discret mono-dimensionnel. Des notations similaires seront employées pour les coefficients issus de $\mathcal{B}^{\mathrm{II}}$.

Ainsi, pour construire un schéma à deux descriptions, la stratégie est la suivante :

1. Calculer $a^{\mathrm{I}}_{j,n}$ et $d^{\mathrm{I}}_{j,n}$ for $j \in \{1, \ldots, J-1\}$, comme auparavant,
2. Calculer $a^{\mathrm{I}}_{J,n}$, $d^{\mathrm{I}}_{J,n}$, $a^{\mathrm{II}}_{J,n}$ et $d^{\mathrm{II}}_{J,n}$ (au dernier niveau de résolution, J),
3. Sous-échantillonner d'un facteur 2 tous les sous-bandes de détail, mais garder les sous-bandes d'approximation intactes,
4. Former deux descriptions avec ce qui reste des coefficients.

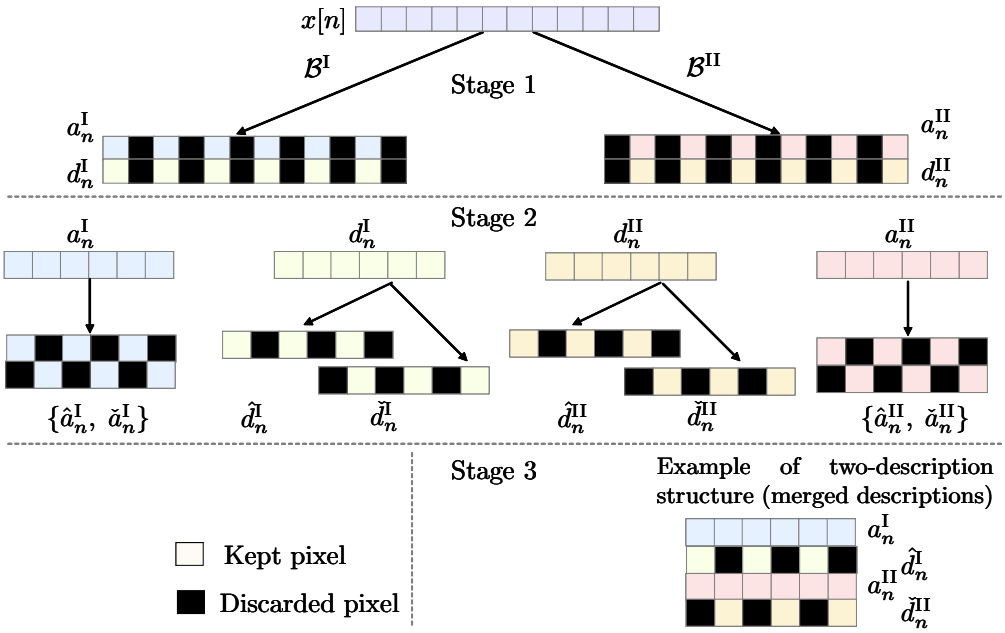


FIGURE 3: Construction des schéma à deux descriptions à redondance réduite. Première étape : construire la décomposition en ondelettes basée sur $\mathcal{B}^{\mathrm{I}}$ and $\mathcal{B}^{\mathrm{II}}$. Ensuite : échantillonnage d'un facteur 2 et selection d'un sous-ensemble de coefficients.

Dans la Figure 3, nous présentons la stratégie énoncée sur un niveau de décomposition. Le dernier étage dans cette figure donne une combination possible des sous-bandes pour former deux descriptions. Il faut souligner qu'il est important de garder l'intégralité des sous-bandes d'approximation (pas d'échantillonnage supplémentaire) à fin d'assurer une qualité de service minimale quand seulement une de deux descriptions peut être décodée.

Note: La décomposition sur une base d'ondelettes translatée, qui revient donc à diversifier le sous-échantillonnage classique, voir l'éliminer, peut être réalisée à n'importe quel niveau de résolution, non pas uniquement au dernier niveau. Cependant, plus le niveau de résolution est haut, plus la redondance introduite est importante et, par consequent, le post-traitement par sous-échantillonnage doit être important, menant en plus à un ensemble plus consequent de combinaisons possibles pour la création de descriptions. En effet, la plus faible redondance du schéma est obtenue en introduisant la transformée sur-échantillonnée au dernier niveau de

décomposition. En outre, l'étude de reconstruction parfaite peut être mené uniquement sur cet ensemble de coefficients. Nous avons donc favorisé ce type de décomposition à la fois pour sa plus grande simplicité et pour sa plus faible redondance.

En outre, il est bien connu que la décomposition en trame d'ondelette proposée est équivalente pour des signaux numériques à une décomposition en bancs de filtres suréchantillonnés. Avec cette considération, et en prenant $(h_n)_{n\in\mathbb{Z}}$ et $(g_n)_{n\in\mathbb{Z}}$ comme réponses impulsionnelles des filtres d'analyse, ainsi que le signal numérique $(x_n)_{n\in\mathbb{Z}}$ comme auparavant, nous pouvons définir plusieurs schémas à deux descriptions.

Plus particulièrement, nous avons identifié les cas suivants qui couvrent toutes les combinaisons possibles dans le contexte de la Figure 3 :

- **Schéma R** - composé des séquences de coefficients :

$$\{\underbrace{\hat{a}_n^{\mathrm{I}},\ \check{a}_n^{\mathrm{I}},\ \hat{d}_n^{\mathrm{I}},}_{\text{1ère description}}\ \underbrace{\hat{a}_n^{\mathrm{I}},\ \check{a}_n^{\mathrm{I}},\ \check{d}_n^{\mathrm{I}}}_{\text{2ème description}}\ \}.$$

- **Schéma MD1** - donné par :

$$\{\underbrace{\hat{a}_n^{\mathrm{I}},\ \check{a}_n^{\mathrm{I}},\ \hat{d}_n^{\mathrm{I}},}_{\text{1ère description}}\ \underbrace{\hat{a}_n^{\mathrm{II}},\ \check{a}_n^{\mathrm{II}},\ \check{d}_n^{\mathrm{I}}}_{\text{2ème description}}\ \}.$$

- **Schéma MD2** - avec :

$$\{\underbrace{\hat{a}_n^{\mathrm{I}},\ \check{a}_n^{\mathrm{I}},\ \hat{d}_n^{\mathrm{I}},}_{\text{1st description}}\ \underbrace{\hat{a}_n^{\mathrm{II}},\ \check{a}_n^{\mathrm{II}},\ \hat{d}_n^{\mathrm{II}}}_{\text{2nd description}}\},$$

- **Schéma MD3** - donné par :

$$\{\underbrace{\hat{a}_n^{\mathrm{I}},\ \check{a}_n^{\mathrm{I}},\ \hat{d}_n^{\mathrm{I}},}_{\text{1st description}}\ \underbrace{\hat{a}_n^{\mathrm{II}},\ \check{a}_n^{\mathrm{II}},\ \check{d}_n^{\mathrm{II}}}_{\text{2nd description}}\}$$

Ces ensembles sont représentés au dernier niveau de résolution (J), ceci étant l'endroit ou la redondance est introduite dans le schéma. Les niveaux de résolution plus fine sont tout simplement crées avec les coefficients de détail issus d'une décomposition en ondelettes usuelle, échantillonnée critique, qui ont été repartis dans les descriptions selon leurs indices pairs et impairs. De plus, ces sous-bandes, regroupées dans un vecteur $\mathbf{c}_n = \{\hat{a}_n^{\mathrm{I}}, \check{a}_n^{\mathrm{I}}, \hat{a}_n^{\mathrm{II}}, \check{a}_n^{\mathrm{II}}, \hat{d}_n^{\mathrm{I}}, \check{d}_n^{\mathrm{I}}, \hat{d}_n^{\mathrm{II}}, \check{d}_n^{\mathrm{II}}\}$ ont été obtenues en passant par un banc de filtres suréchantillonnés ayant les réponses impulsionnelles évoquées plus haut. Pour ce faire, il est nécessaire de définir les composantes polyphase des réponses impulsionnelles des filtres d'analyse comme suit :

$$\forall i \in \{0,1,2,3\}, \qquad h_i(n) = h_{4n-i}, \quad g_i(n) = g_{4n-i}$$

et leur transformées en Z correspondantes, notées par : $H_i(z)$ and $G_i(z)$. De même le signal d'entrée est divisé en quatre composantes polyphase qui s'ecrivent :

$$\forall i \in \{0,1,2,3\}, \qquad x_n^{(i)} = x_{4n+i}.$$

ce qui donne le vecteur polyphase suivant :

$$\mathbf{x}_n = \left(x_n^{(0)}\ x_n^{(1)}\ x_n^{(2)}\ x_n^{(3)}\right)^{\mathsf{T}}.$$

Avec ces considérations, l'equation de transfert correspondante au banc de filtres suréchantillonnés globale sera donnée par :

$$\mathbf{C}(z) = \mathbf{M}(z)\ \mathbf{X}(z),$$

où $\mathbf{C}(z)$ et $\mathbf{X}(z)$ sont les transformées en Z du vecteur de séquences de coefficients, $\mathbf{c}_n$, et du signal d'entrée, $\mathbf{x}_n$ et $\mathbf{M}(z)$ est la matrice polyphase globale de transfert.

Dans les schémas à deux descriptions présentés, un sous-ensemble de 6 sous-bandes à été à chaque fois sélectionné (le schéma **R** est donné comme référence car il est équivalent à un schéma non-redondant en terme d'information).

Un problème important qui est soulevé par notre approche est celui de la reconstruction parfaite au niveau du décodeur central. En effet, l'étape de sous-échantillonnage supplémentaire peut affecter la structure de trame. Nous avons étudié ce problème en détail, en procédant en deux étapes. Dans un premier temps nous avons déduit comme condition nécessaire et suffisante pour la trame, l'inversibilité au dernier niveau de décomposition. Ensuite, grace à la representation polyphase nous avons étudie l'inversibilité pour les deux cas de figure les plus souvent employés en codage vidéo : les filtres de Haar et les bancs de filtres biorthogonaux suréchantillonnés $5/3$. Cela se traduit par l'étude de l'inversibilité des sous-matrices de $\mathbf{M}(z)$ ayant 6 lignes sur les 8. Nous avons donc construit un banc de filtres suréchantillonnés d'un facteur 6×4.

De plus, pour ces bancs de filtres il existe une implementation rapide et efficace du point de vue de la mémoire consommée, appelée schéma "lifting", introduite dans [HP96] et étendue dans [HP98] sous le nom de "décompositions non-linéaires par sous-bande". Le terme "lifting" a été introduit dans le tutoriel [DS98]. Ce schéma garantit la reconstruction parfaite du système.

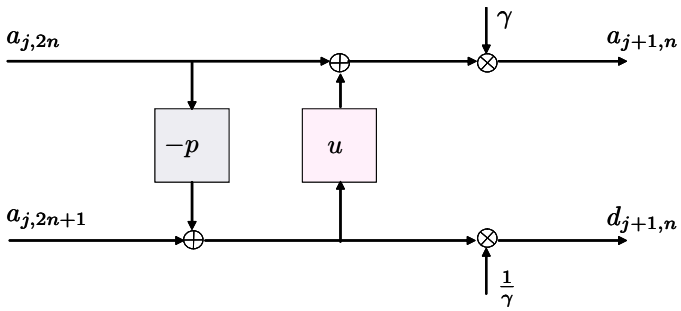


FIGURE 4: Schéma en lifting de base, sur un étage.

Le noyau de lifting que l'on utilise pour l'implementation de notre stratégie par descriptions multiples est présenté dans la Figure 4. Ici, p, respectivement u représentent les opérateurs de prédiction et de mise à jour, tandis que γ est une constante multiplicative réelle, non-nulle.

Ainsi, nous pouvons illustrer une de nos schémas à descriptions multiples temporelles comme dans la Figure 5, dans laquelle nous affichons uniquement deux niveaux de décomposition temporelle pour alléger la lecture.

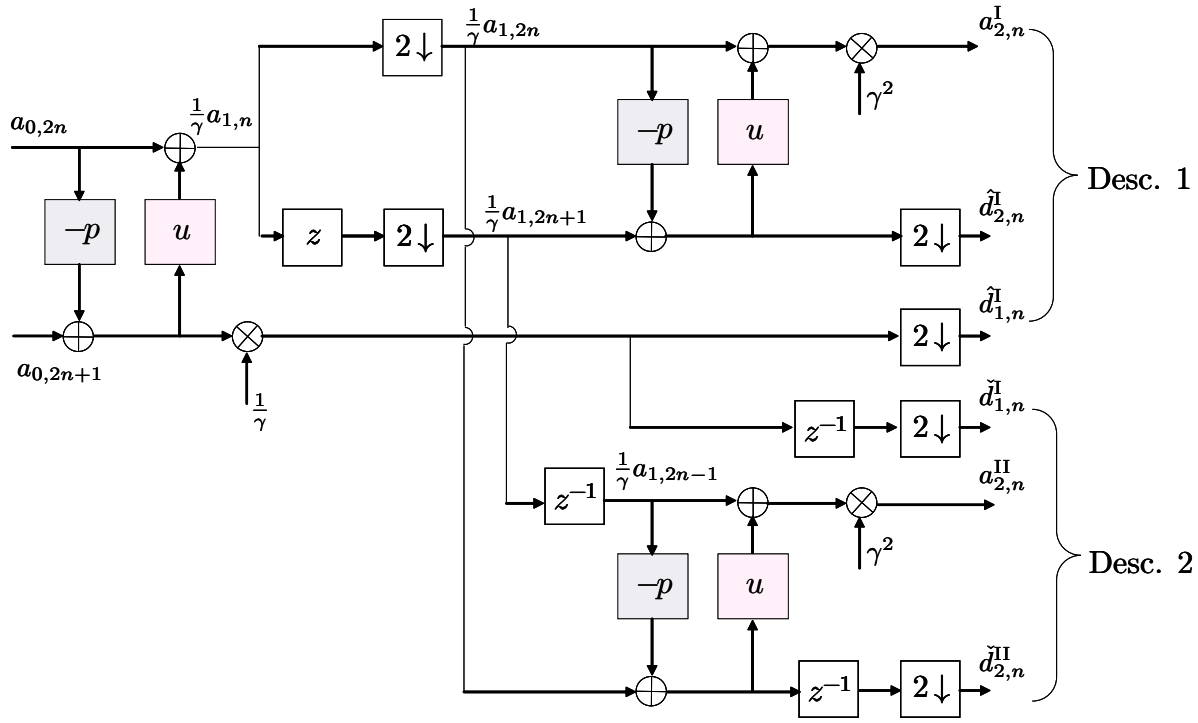


FIGURE 5: Implementation en lifting 2-bandes sur les deux derniers niveaux de décomposition d'un des schémas MDC proposés.

Considérations sur les bancs de filtres utilisés

Pour les deux classes de filtres d'analyse considérées (Haar et biorthogonal $5/3$), nous avons exploré l'inversibilité de chacune des matrices polyphase d'analyse, résultantes. Autrement dit, nous avons cherché une matrice de synthèse, $\mathcal{W}$, pour laquelle nous avons calculé l'erreur quadratique moyenne de la reconstruction pour un bruit de quantification de variance unitaire $\mathrm{EQM_r} = \|\boldsymbol{\mathcal{W}}\|^2$. Il est connu que l'effet du bruit en sortie, dû à la quantification, est minimal si pour tout i, la norme de Frobenius de chaque colonne de $\mathcal{W}$ est minimale. De plus, puisque l'on se place dans un cas suréchantillonné, plusieurs solutions d'inversion sont possibles. Cela est traduit par l'ordre des filtres de synthèse quand l'on exprime la matrices de transfert du système comme des matrices polynomiales de Laurent, $\mathrm{W(z)} = \sum_{p=-p_1}^{p_2} W_p z^{-p}$.

Nous avons mené une étude expérimentale pour le calcul de $\mathrm{EQM_r}$ de nos schémas, en faisant varier l'ordre maximal, P, des filtres de reconstruction. Nous présentons ici les résultats obtenus pour les bancs de filtres biorthogonaux $5/3$, centralisés dans le Tableau 1. Dans ce cas l'ordre des filtres d'analyse est $Q = 3$ ce qui correspond à deux cas de figure : filtres d'analyse causals ($(q_1,\, q_2) = (0, 2)$) ou alors "mixtes" ($(q_1,\, q_2) = (1, 1)$), où la matrice de transfert du système[2] est $\mathrm{M(z)} = \sum_{q=-q_1}^{q_2} M_q z^{-q}$.

Ceci a permis d'en déduire des critères de choix entre les schémas obtenus avec un banc de filtres ainsi que d'établir des conditions d'usage d'un banc de filtre plutôt que l'autre. Il convient pourtant de remarquer que ces résultats portent uniquement sur les décodeurs centraux et que donc les performances de décodage latéral ne sont pas considérées. Ceci explique

[2] en passant par la transformée en Z.

Filtres biorthogonaux $5/3$									
Schémas		Analyse: $(q_1, q_2) = (0, 2)$ Ordre des filtres de synthèse (P)				Analyse: $(q_1, q_2) = (1, 1)$ Ordre des filtres de synthèse (P)			
		2	3	4	5	2	3	4	5
Schéma **R**	$\mathrm{EQM_r}$	4.37	4.37	4.37	4.37	4.37	**4.37**	4.37	4.37
	(p_1, p_2)	$(0,1)$	$(1,1)$	$(1,2)$	$(1,3)$	$(0,1)$	$(1,1)$	$(1,2)$	$(1,3)$
Schéma **MD1**	$\mathrm{EQM_r}$	4.37	3.38	2.58	2.57	95.12	**2.65**	2.59	2.57
	(p_1, p_2)	$(0,1)$	$(1,1)$	$(1,2)$	$(2,2)$	$(1,0)$	$(1,1)$	$(2,1)$	$(2,2)$
Schéma **MD2**	$\mathrm{EQM_r}$	14.60	3.20	2.94	2.93	14.60	**3.20**	2.94	2.93
	(p_1, p_2)	$(0,1)$	$(1,1)$	$(1,2)$	$(2,2)$	$(0,1)$	$(1,1)$	$(1,2)$	$(2,2)$
Schéma **MD3**	$\mathrm{EQM_r}$	9.69	5.70	2.99	2.96	65.44	**3.23**	3.08	2.93
	(p_1, p_2)	$(0,1)$	$(0,2)$	$(1,2)$	$(1,3)$	$(0,1)$	$(1,1)$	$(2,1)$	$(2,2)$

TABLE 1: L'evolution de $\mathrm{EQM_r}$ avec l'ordre maximal de reconstruction des filtres pour les schémas à descriptions multiples (MDC) et le schéma non-redondant (SDC) basé sur une décomposition sur $\mathcal{B}^{\mathrm{I}}$ avec des filtres d'analyse biorthogonaux $5/3$.

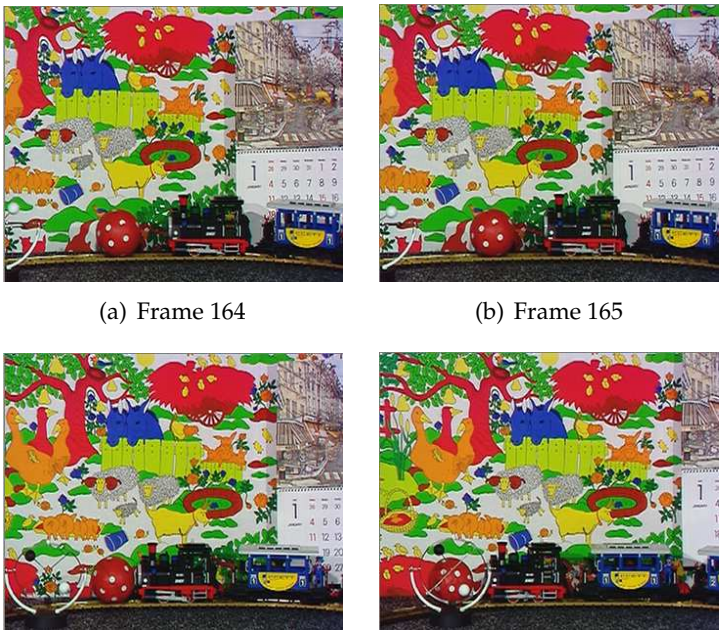
pourquoi le schéma **MD1**, bien que supérieur aux autres en performance centrale n'est pas le meilleur candidat si l'on vise des performances latérales équivalentes. Aussi, les performances en terme d'erreur de reconstruction n'augmentent pas de manière sensible avec l'ordre des filtres de synthèse et donc, compte tenu des difficultés liées au codage vidéo, les filtres ayant la réponse la plus courte sont à favoriser.

Simulations et résultats

Nous avons considéré une application de codage vidéo robuste pour des scénarios à pertes. Pour la séquence de test dont on extrait quelques trames pour exemplification dans la Figure 6, nous présentons dans la Figure 7 une comparaison entre le schéma **MD1** et le schéma **MD3**, quand on se place dans un scénario de fonctionnement tout-ou-rien. Comme nous avions vu précédemment, ces deux schémas sont quasi-équivalentes du point de vue du décodeur central, cependant les performances des décodeurs latéraux sont beaucoup plus proches pour le schéma **MD3**, ce qui le rend plus attractif pour un scénario de transmission dans lequel les canaux ont la même importance.

Nous présentons ici également les performances de codage en fonction de taux de pertes de paquets qui sont illustrées dans la Figure 8 pour la séquence de test "Foreman" en format QCIF à 30 fps et 250Kbs. Nous avons comparé notre approche avec un schéma de MDC basé sur la division temporelle.

La Figure 9 donne les résultats à plusieurs débits pour un des points sur la Figure 8, correspondant à un pourcentage de pertes de paquets de 10%, toujours en comparant à une approche MDC basée sur distribution temporelle des échantillons du signal en deux descriptions (selon les indices pairs et impairs). Ces résultats ont été obtenus en moyennant approximativement 1500 réalisations.



(a) Frame 164 (b) Frame 165

(c) Frame 228 (d) Frame 275

FIGURE 6: Quatre trames extraites de la séquence de test ”Mobile”.

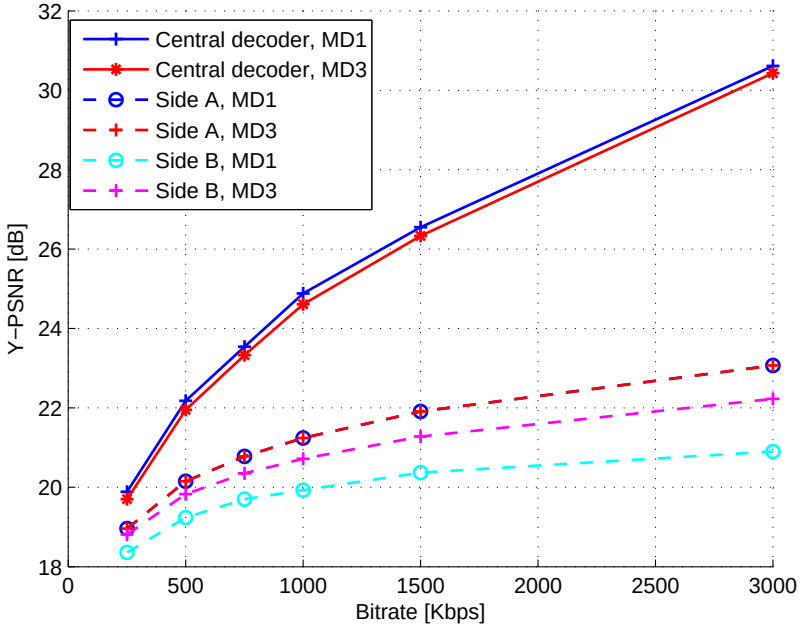


FIGURE 7: Courbes Y-PSNR vs. débit pour deux schémas à deux descriptions temporelles (séquence de test “Mobile”, en format CIF, 30fps).

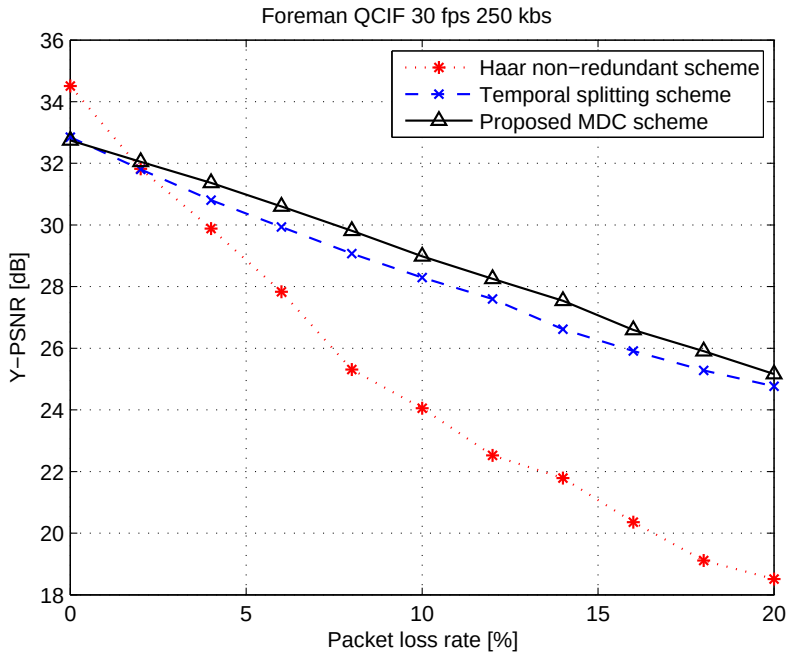


FIGURE 8: Distorsion vs. taux de pertes de paquets (“Foreman” QCIF, 30 fps, 250 Kbs).

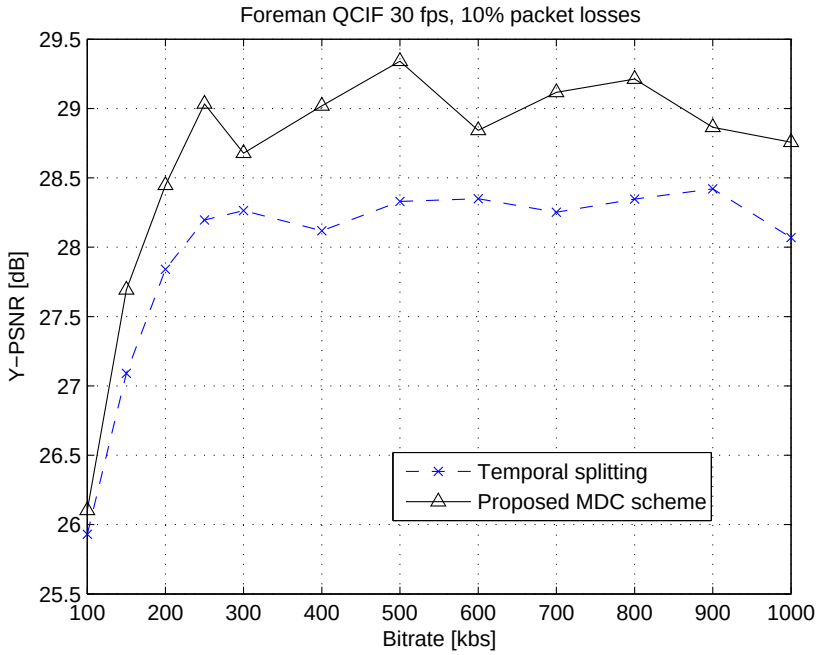


FIGURE 9: Courbes de Y-PSNR vs. débit à un taux de pertes de paquets de 10% (séquence de test “Foreman” en QCIF, 30 fps).

Schémas à descriptions multiples spatiales

Pour réaliser une extension du cas monodimensionnel au cas bidimensionnel pour les schémas proposés, nous allons adapter les notations précédentes comme suit : les sous-bandes d'approximation et de détail au niveau de résolution $j \in \{1, \ldots, J\}$ seront respectivement désignées par a_j, dh_j, dv_j and dd_j, où la deuxième lettre dans les sous-bandes de détail correspond à leur orientation : horizontale, verticale où diagonale, telle que données par une décomposition classique d'image en ondelettes. Comme précédemment, le dernier niveau de résolution est indiquée par la lettre majuscule J.

Ainsi, une décomposition en banc de filtres comme dans le cas temporel utilisera les filtres passe-bas, respectivement passe-haut, à réponses impulsionnelles $(h[n])_{n\in\mathbb{Z}}$ et $(g[n])_{n\in\mathbb{Z}}$.

Alors, sous l'hypothèse de séparabilité les sous-bandes d'ondelettes sont obtenues dans le cas bidimensionnel par une cascade de convolutions et décimations d'un facteur 2 :

$$\begin{aligned} a_j[n,m] &= \sum_{k,l} a_{j-1}[k,l]h[2n-k]h[2m-l] \\ dh_j[n,m] &= \sum_{k,l} a_{j-1}[k,l]h[2n-k]g[2m-l] \\ dv_j[n,m] &= \sum_{k,l} a_{j-1}[k,l]g[2n-k]h[2m-l] \\ dd_j[n,m] &= \sum_{k,l} a_{j-1}[k,l]g[2n-k]g[2m-l], \end{aligned} \tag{7}$$

à chaque niveau de résolution, $j \in \{1, \ldots J\}$, où a_{j-1} représente la sous-bande d'approximation à la résolution immédiatement supérieure (plus fine).

Dans ce cas de figure, nos schémas à deux descriptions utiliseront cette décomposition jusqu'au niveau $J-1$, suivi d'une décomposition non-décimée au dernier niveau (comme dans le cas temporel). Une telle décomposition est obtenue en translatant les reposes impulsionnelles des filtres de 1 dans chacune des directions spatiales ou les deux directions à la fois.

Écrivons donc les équations donnant les sous-bandes d'ondelettes au dernier niveau de résolution pour le cas suréchantillonné :

$$\begin{aligned} a_{J_{(s,s')}}[n,m] &= \sum_{k,l} a_{J-1}[k,l]h[2n+s-k]h[2m+s'-l] \\ dh_{J_{(s,s')}}[n,m] &= \sum_{k,l} a_{J-1}[k,l]h[2n+s-k]g[2m+s'-l] \\ dv_{J_{(s,s')}}[n,m] &= \sum_{k,l} a_{J-1}[k,l]g[2n+s-k]h[2m+s'-l] \\ dd_{J_{(s,s')}}[n,m] &= \sum_{k,l} a_{J-1}[k,l]g[2n+s-k]g[2m+s'-l], \end{aligned} \tag{8}$$

Ici, les paramètres $(s,s') \in \{0,1\}$ introduits dans les réponses impulsionnelles des filtres d'analyse dans chaque direction spatiale représentent le facteur de translation employé.

En faisant le lien avec les notations introduites dans la section précédente, la paire $(s,s') = (0,0)$ correspondra donc aux coefficients issus de $\mathcal{B}^{\mathrm{I}}$ tandis que une paire $(s,s') \neq (0,0)$ indique l'utilisation de $\mathcal{B}^{\mathrm{II}}$ dans au moins une direction spatiale.

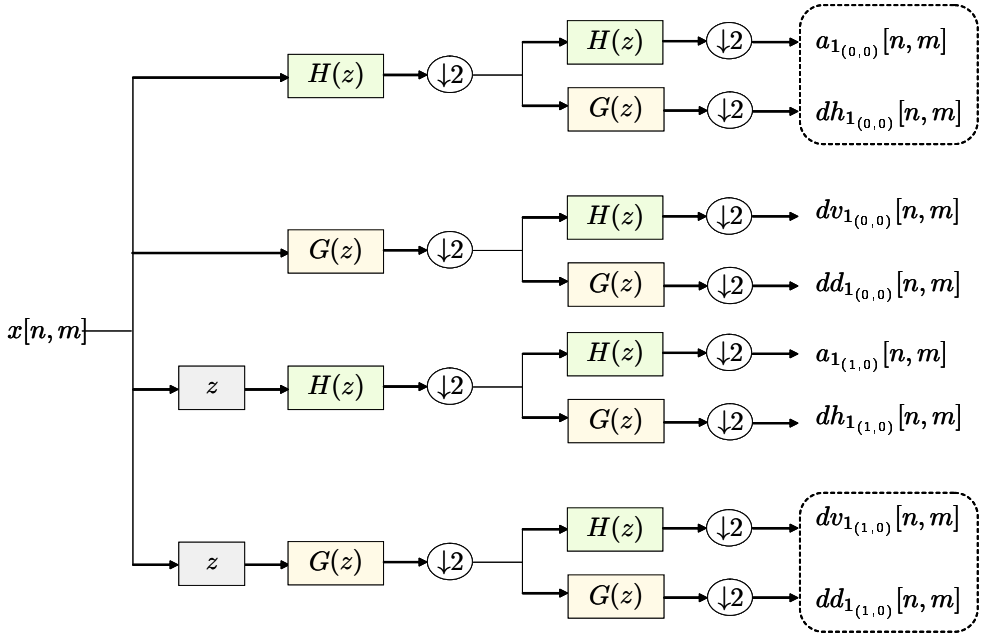


FIGURE 10: Exemple de banc de filtres sur-échantillonné pour un niveau de décomposition separable en ondelettes de l'image x.

La Figure 10 représente une décomposition sur un niveau avec les paramètres $(s, s') = (0, 0)$ pour les deux premières branches et $(s, s') = (1, 0)$ pour les branches suivantes. Donc la deuxième décomposition en ondelettes est obtenue dans ce cas en translatant les réponses impulsionnelles des filtres d'analyse de 1 sur l'horizontale. Les filtres usuels sont employés pour la direction verticale.

L'étape suivante dans la construction de schémas à descriptions multiples est donnée par le sous-échantillonnage supplémentaire des sous-bandes de détail. Dans le cas spatial plusieurs stratégies peuvent être envisagées. Pour des raisons de symétrie des descriptions ainsi que de distribution homogène de l'information à transmettre, nous proposons un sous-échantillonnage en quinconce, que l'on illustre dans la Figure 11.

Pour toutes les combinaisons possibles de sous-bandes d'ondelettes utilisées pour la composition des deux descriptions nous avons également étudié des critères de choix basés sur la reconstruction parfaite et l'erreur quadratique moyenne de la reconstruction pour un pas de quantification de variance unitaire.

Les performances de décodage des schémas à descriptions multiples dans le cas bidimensionnel ont été améliorées par un algorithme itératif d'optimization quadratique, issus des développements récents des méthodes de projection sur des ensembles convexes (POCS). L'idée directrice de ce post-traitement est basée sur la prise en compte des contraintes de quantification qui peuvent être modélisées comme des ensembles convexes et qui représentent de l'information *a priori* sur le système pouvant être prise en compte au décodage. Cet algorithme s'est avéré fort dépendant de l'initialisation et nous avons proposé plusieurs stratégies de combinaison des sous-bandes reçues avant optimisation.

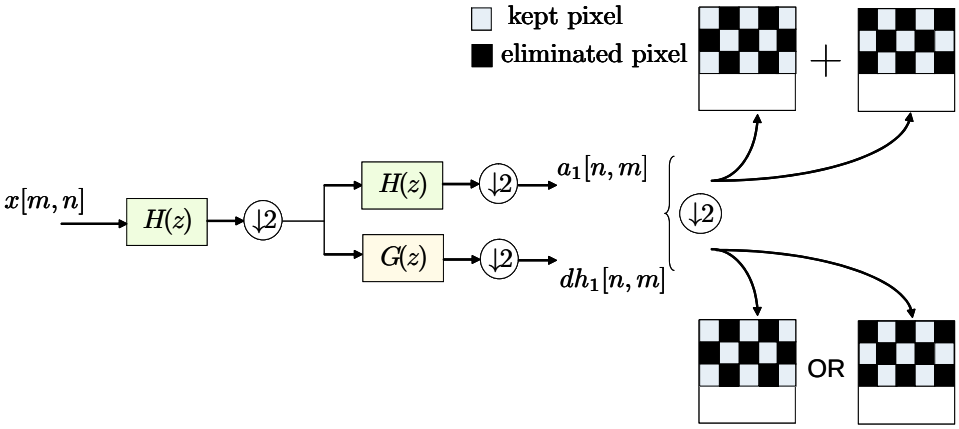


FIGURE 11: Illustration du sous-échantillonnage ultérieur à la transformée en ondelettes classique. Les sous-bandes d'approximation résultantes sont entièrement gardées dans le schéma à descriptions multiples, tandis que les coefficients de détail sont sous-échantillonnés sur une grille en quinconce. Uniquement une des parités du sous-échantillonnage quinconce sera gardé dans la représentation finale.

Résultats

Choix des schémas

Tout d'abord nous avons considéré un fonctionnement du type "marche/arrêt" des deux canaux. Nous avons évalué les performances des schémas ayant présenté la reconstruction parfaite et les performances les plus prometteuses en terme de reduction du bruit de quantification. Dans ce type de fonctionnement basique il n'y a pas d'autres pertes de paquets et on considère donc les descriptions comme reçues ou perdues. Les résultats obtenus au niveau des décodeurs centraux pour deux schémas qui assurent la reconstruction parfaite et diminuent l'erreur quadratique moyenne de la reconstruction[3] sont présentés dans la Figure 12. En outre l'algorithme d'optimisation convexe pour l'amélioration de la reconstruction à été utilisée ici sur les deux schémas. Les performances avant optimisation sont données en traits pointillés. Les traits pleins représentent les performances obtenues après 30 itérations d'optimisation.

Dans cette situation il n'y a pas d'indication sur un critère qui permettrait de choisir un schéma plutôt qu'un autre, en revanche la reconstruction au niveau des décodeurs latéraux donne un meilleur aperçu. Les performance PSNR vs. débit des deux schémas sont indiqués par la Figure 13. Comme précédemment, les traits en pointillés correspondent aux valeurs obtenues avant l'algorithme itératif, avec une initialisation moyenne, tandis que les traits pleins correspondent aux valeurs obtenues après 30 itérations d'optimisation convexe.

Pertes de paquets aléatoires

Le deuxième ensemble d'expérimentations proposé dans cette thèse concerne un scénario de pertes aléatoires selon un modèle de canal à bruit gaussien. Ceci permet d'illustrer l'intérêt des techniques MDC sur les techniques non-redondantes. A titre d'exemple, nous donnons ici les

[3] une analyse d'inversibilité similaire au cas des descriptions temporelles à été réalisée dans le cas spatial en calculant les valeurs singulières minimales de la matrice polyphase

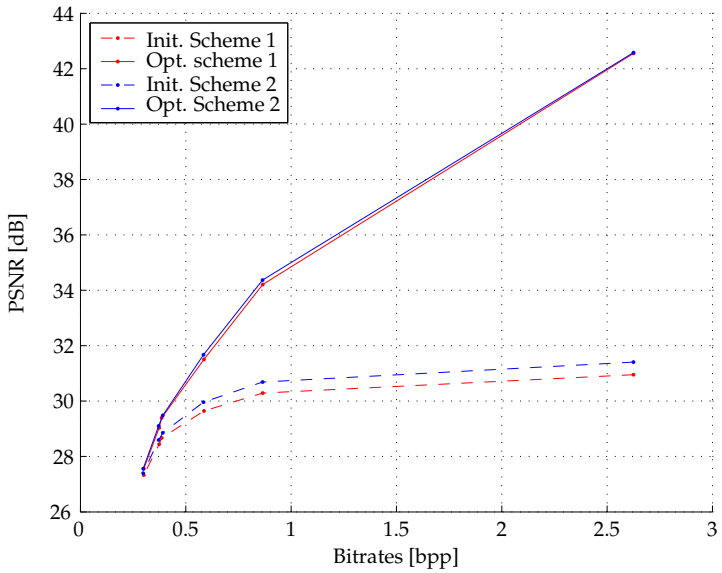


FIGURE 12: Evaluation PSNR-débit de deux schémas MDC au niveau des décodeurs centraux. (image de test : LENA 512x512 pixels).

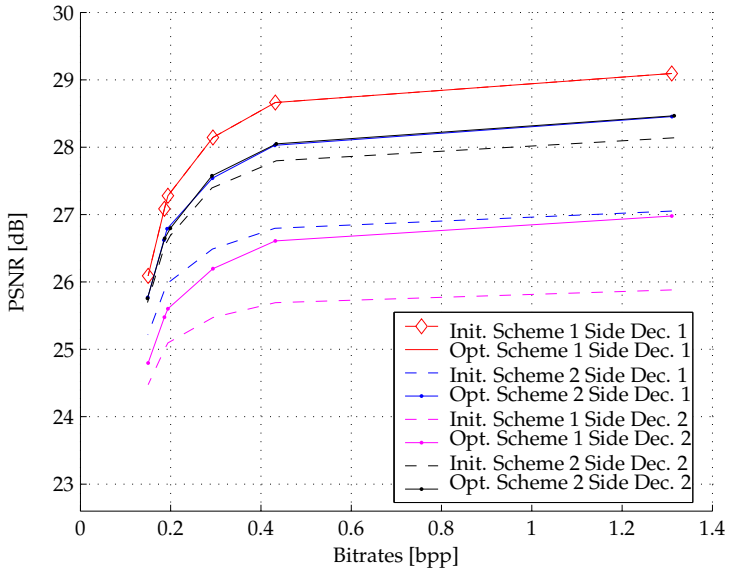


FIGURE 13: Evaluation du PSNR vs. le débit pour les decodeurs lateraux .

résultats pour l'image de test "Man" à 512x512 pixels, obtenus pour 4% de pertes aléatoires de coefficients dans chacune des descriptions, comparés au cas non-rédondant affecté par les mêmes pertes que la première description.

Nous avons prouvé expérimentalement que l'initialisation de l'algorithme itératif avec une combinaison de type moyenne pondérée des sous-bandes individuelles de chaque description donne les meilleures performances. Cependant, le calcul des poids optimaux (dans le sens des moindres carrés et en absence de pertes) et leur utilisation au décodeur nécessiterait un canal de transmission supplémentaire. Une étude empirique sur les valeurs de ces poids obtenues sur une base d'images ayant différentes caractéristiques (images naturelles, images de

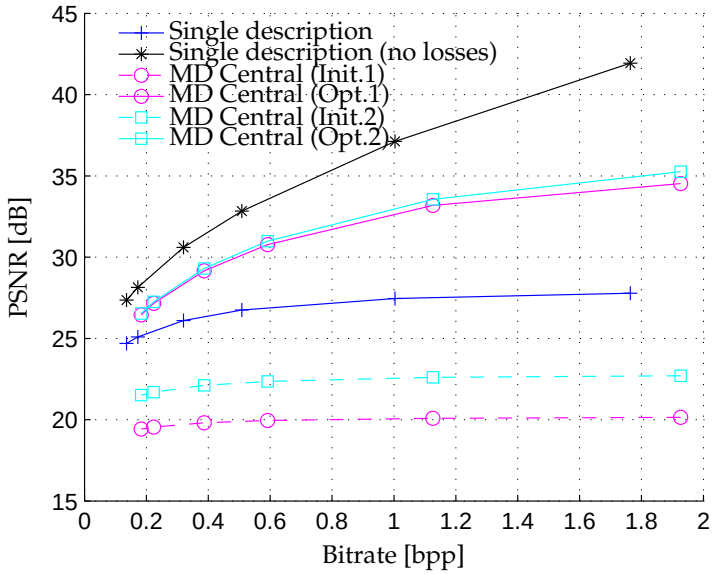


FIGURE 14: PSNR vs. débit. Comparaison entre le décodeur central d'un des schéma MDC bidimensionnels à reconstruction parfaite à 4% de pertes aléatoires dans chaque description et le cas non-redondant affecté par les mêmes pertes que la première description. Pour le schéma MDC nous illustrons les résultats pour deux initialisations differentes, Init 1 et Init 2, ainsi que les résultats après l'algorithme d'optimisation : les courbes Opt. 1 et Opt. 2. Comme réference nous donnons la borne supérieure qui est le cas non-redondant en absence de pertes.

synthèse, images satellitaires, images bio-medicales) et contenant quelque dizaines d'éléments nous a montré qu'il est possible de choisir un ensemble fixe de poids de pondération de chaque sous-bande individuelle qui n'affecte pas de manière trop importante les résultats et qui varie de manière négligeable d'une image à l'autre. Ceci présente l'avantage de pouvoir être predéterminé au niveau du décodeur et donc ne nécessite plus de canal secondaire.

Dans le scénario à pertes aléatoires de coefficients, nous avons également calculé l'ensemble des poids optimaux quand le modèle de pertes est connu dans chaque description. Ceci permet de comparer les résultats obtenus avec les poids fixes par rapport aux performances maximales atteignables. Une telle stratégie peut s'avérer utile si le fonctionnement du réseau de communication peut être considéré comme déterministe.

Approche par trame de synthèse - représentations creuses

Dans la dernière partie de cette thèse, nous avons généralisé le schéma MDC au cas de D descriptions, toujours pour la transmission d'images sur des réseaux à pertes. En outre, nous avons montré comment, à partir d'une représentation choisie au niveau de la synthèse, nous pouvons coder les coefficients de manière à minimiser le débit pour une distorsion maximale fixée. Nous avons ensuite examiné les approximations permettant de reformuler ce problème sous la forme d'un problème d'optimisation convexe.

Trame à l'analyse ou à la synthèse ?

Supposons que le signal à transmettre appartienne à un espace de Hilbert réel, $\mathcal{H}$, de produit scalaire $\langle \cdot, \cdot \rangle$ et de norme associée $\| \cdot \|$. Chaque description $i \in \{1, \dots, D\}$ est obtenue à partir d'une famille de vecteurs $(e_{i,k})_{k \in \mathbb{K}_i}$ de $\mathcal{H}$, où $\mathbb{K}_i \subseteq \mathbb{N}$, l'union de ces D familles étant supposée former une trame de $\mathcal{H}$. Notons les opérateurs d'analyse associés $L_i : \mathcal{H} \to \ell^2(\mathbb{K}_i)$, où $x \mapsto (\langle x, e_{i,k} \rangle)_{k \in \mathbb{K}_i}$. pour tout $i \in \{1, \dots, D\}$, ayant comme opérateurs adjoints $L_i^* : \ell^2(\mathbb{K}_i) \to \mathcal{H}$ avec $(\xi_k)_{k \in \mathbb{K}_i} \mapsto \sum_{k \in \mathbb{K}_i} \xi_k e_{i,k}$

Nous pouvons alors adopter deux points de vue différents pour la construction d'un schéma MDC à partir de ces opérateurs, que nous allons distinguer par les dénominations : paradigme d'analyse et paradigme de synthèse.

Dans le *paradigme d'analyse*, un signal $x \in \mathcal{H}$ est décomposé par les opérateurs linéaires L_i donnés ci-dessus, formant ainsi D descriptions (D ensembles de coefficients) :

$$c_i = L_i x, \qquad i \in \{1, \dots, D\}, \tag{9}$$

qui sont quantifiés et transmis séparément. Le signal reconstruit $\hat{x}$, que l'on cherche à rendre le plus proche possible de x, peut provenir d'un nombre réduit de descriptions ou de toutes (selon le scénario de transmission survenu). Pour améliorer le décodage, on peut considérer que la quantification effectuée au départ se traduit par un ensemble de contraintes convexes imposées au décodeur [CMW99, PPPP05b]. Dans ce cas la reconstruction, bien que pouvant être de bonne qualité, est non linéaire et donc coûteuse en calculs (Figure 15).

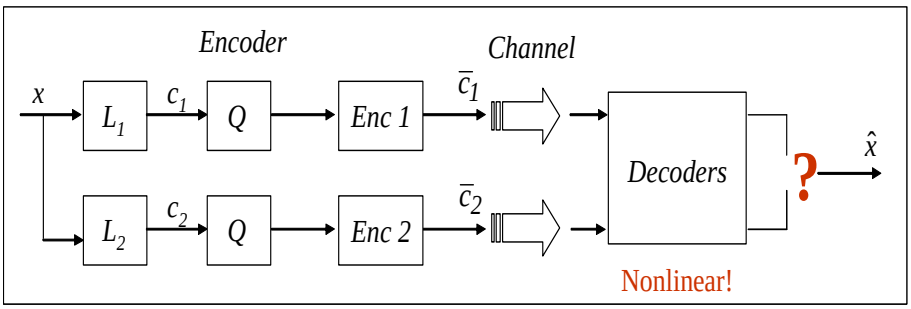


FIGURE 15: Paradigme d'analyse.

La seconde approche est celle du *paradigme de synthèse*, qui consiste à utiliser des opérateurs linéaires au niveau du décodeur. Dans ce cadre, les opérateurs adjoints sont directement appliqués aux coefficients quantifiés $\bar{c}_i$ ou, plus exactement, au sous-ensemble $\mathbb{I} \subset \{1, \dots, D\}$ des descriptions reçues. Les suites de coefficients quantifiés vont être générées par un algorithme d'optimisation débit-distorsion qui prend en compte les caractéristiques du canal. Ceci rend l'encodeur non-linéaire mais en revanche les décodeurs prennent des formes linéaires très simples (Figure 16).

Par exemple, un choix naturel pour le décodeur central (quand $\mathbb{I} = \{1, \dots, D\}$) est : $\hat{x} = \sum_{i=1}^{D} L_i^* \bar{c}_i$. Pour les décodeurs latéraux, on peut de manière similaire adopter une reconstruction du type : $\hat{x} = \sum_{i \in \mathbb{I}} \tilde{L}_{\mathbb{I},i}^* \bar{c}_i$, où $\widetilde{L}_{\mathbb{I},i}^* : \ell^2(\mathbb{K}_i) \to \mathcal{H}$ est un opérateur de reconstruction bien choisi (pas nécessairement le même que celui utilisé au décodeur central).

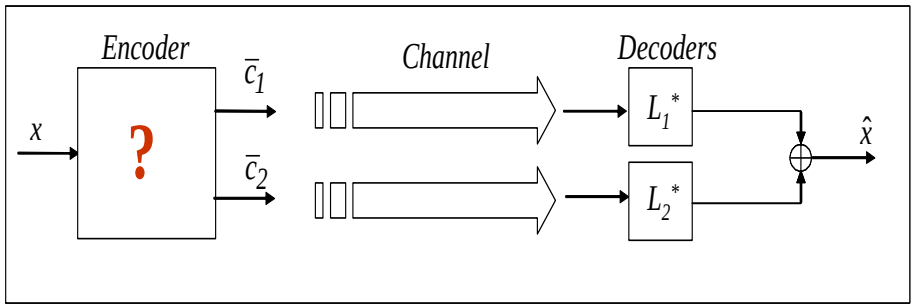


FIGURE 16: Paradigme de synthèse.

Problème d'optimisation débit-distorsion

En introduisant $R(\bar{c}_i)$, $i \in \{1, \ldots, D\}$, le nombre de bits nécessaires à la transmission de la séquence des valeurs quantifiées $\bar{c}_i$, nous nous proposons de minimiser le débit global :

$$R_{\text{global}} = \sum_{i=1}^{D} R(\bar{c}_i) \tag{10}$$

sous une contrainte de distorsion globale donnée par :

$$D = \sum_{\mathbb{I} \in \mathcal{P}} \alpha_{\mathbb{I}} \|x - \sum_{i \in \mathbb{I}} \tilde{L}^*_{\mathbb{I},i} \bar{c}_i\|^2 \leq D_{\max}. \tag{11}$$

où $\mathcal{P}$ est l'ensemble des parties non vides de $\{1, \ldots, D\}$. Les poids positifs $\alpha_{\mathbb{I}}$ représentent, par exemple, les probabilités de réception dans les différents scénarios, mais d'autres considérations peuvent entrer en ligne de compte (qualité perceptuelle etc) dans leur choix.

Cependant, trouver les suites $\bar{c}_i$ qui minimisent le débit sous la contrainte de distorsion globale est un problème d'optimisation non-convexe difficile. En faisant quelques hypothèses sur les coefficients ainsi que sur le bruit de quantification, nous avons ramené ce problème à une optimisation convexe qu'il est possible de résoudre numériquement à l'aide d'algorithmes récents [CCPW06].

Résultats

Pour illustrer le paradigme de trame de synthèse nous avons considéré un exemple de transmission d'images basé sur le standard de compression JPEG2000. Pour un scénario à deux descriptions nous pouvons construire une décomposition en trame utilisant comme opérateur de reconstruction L_1^* associé à une base d'ondelettes biorthogonales $9/7$ et l'opérateur de reconstruction L_2^*, associé à la même base mais translatée d'un pixel dans chaque direction spatiale. Dans ce cas, un choix naturel pour les décodeurs latéraux est donné par : $\widetilde{L}_i^* = 2L_i^*$, $i \in \{1, 2\}$. Nous avons utilisé l'image standard de test Lena, à 512×512 pixels, encodée en descriptions multiples utilisant une décomposition sur trois niveaux de résolution grace à un banc de filtres dyadique. Les poids utilisés dans la contrainte de distorsion ont été pris comme suit : as $\alpha_{1,2} = 0.8$, $\alpha_1 = \alpha_2 = 0.1$.

Les coefficients de trame ont été synthétisés avec l'approche d'optimisation que l'on vient de décrire pour laquelle les paramètres du modèle Gaussien généralisé ont été estimés par une

méthode de maximum de vraisemblance. En outre, le pas de quantification, q, a été optimisé pour chaque débit obtenu avec JPEG2000 pour l'encodage de chaque description quantifiée. Dans la Figure 17 nous présentons l'evolution du PSNR avec le débit global pour les décodeurs central et latéraux. Compte tenu de la sous-optimalité du point de vue de la redondance de la trame choisie pour cet exemple (facteur 2) nous présentons les résultats correspondants à l'application directe de l'encodeur JPEG2000 à moitié du débit. Comme attendu, notre schéma présente des meilleures performances au niveau du décodeur central tout en gardant des performances acceptables au niveau des décodeurs latéraux. Bien entendu, des meilleurs résultats sont à prévoir pour des décompositions plus sophistiquées (notamment plus efficaces en terme de débit).

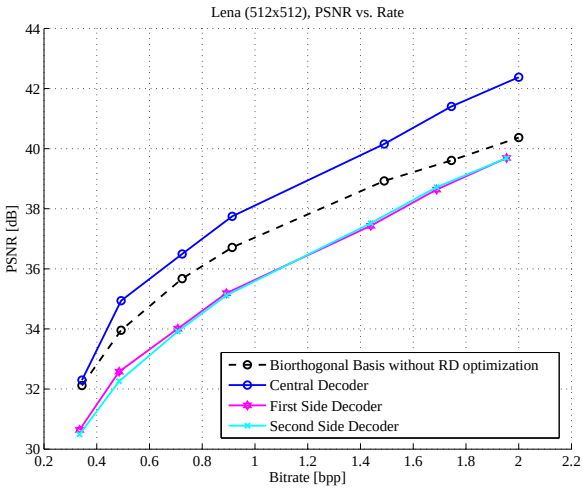


FIGURE 17: Performances débit-distorsion pour Lena, cas de deux descriptions.

Nous avons considéré un scenario similaire avec trois descriptions, en choisissant les opérateurs de synthèse L_1^*, L_2^* and L_3^* correspondant comme avant à une base d'ondelettes biorthogonales $9/7$ et respectivement deux versions translatées : translation de $(1,1)$ i.e. un pixel dans chaque direction spatiale pour L_2^* et de $(1,0)$ pour L_3^*). Nous désignons ces trois descriptions par D_1, D_2, et respectivement D_3, dans la Figure 18.

Dans cette situation, nous avons deux types de décodeurs latéraux : un qui reçoit uniquement une description sur les trois, et l'autre qui reçoit deux descriptions sur trois. Dans ces conditions, un choix naturel pour les opérateurs $\tilde{L}_{\mathbb{I},i}^* q$ est le suivant :

$$\tilde{L}_{\mathbb{I},i}^* = \begin{cases} \frac{3}{2} L_i^*, & \text{if } \mathrm{Card}(\mathbb{I}) = 2 \\ 3 L_i^*, & \text{if } \mathrm{Card}(\mathbb{I}) = 1. \end{cases} \tag{12}$$

Nous avons établi les valeurs de pondération dans la contrainte de distorsion comme étant : $\alpha_{\{1,2,3\}} = 0.8$, $\alpha_{\mathbb{I}} = 0.0618$ quand $\mathrm{Card}(\mathbb{I}) = 2$ et $\alpha_{\mathbb{I}} = 0.0048$, quand $\mathrm{Card}(\mathbb{I}) = 1$.

Les images reconstruites correspondantes à un débit central de 0.8bpp sont présentées dans la Figure 4.5.

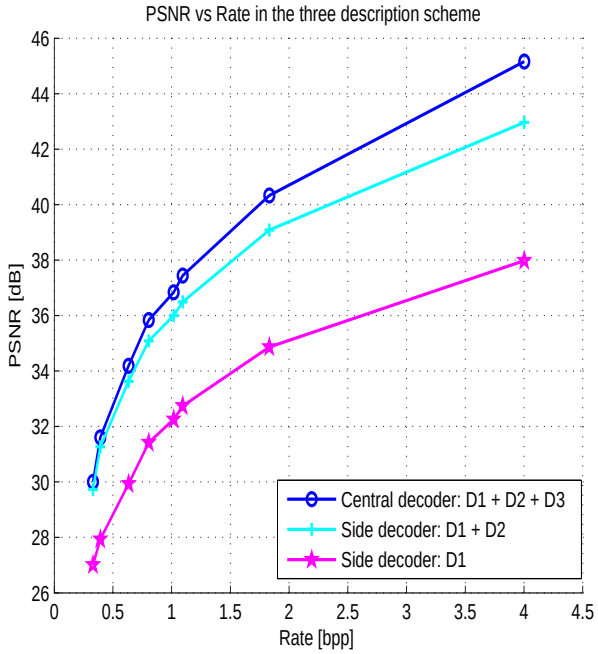


FIGURE 18: Performances débit - distorsion pour le schéma à trois descriptions (Lena).

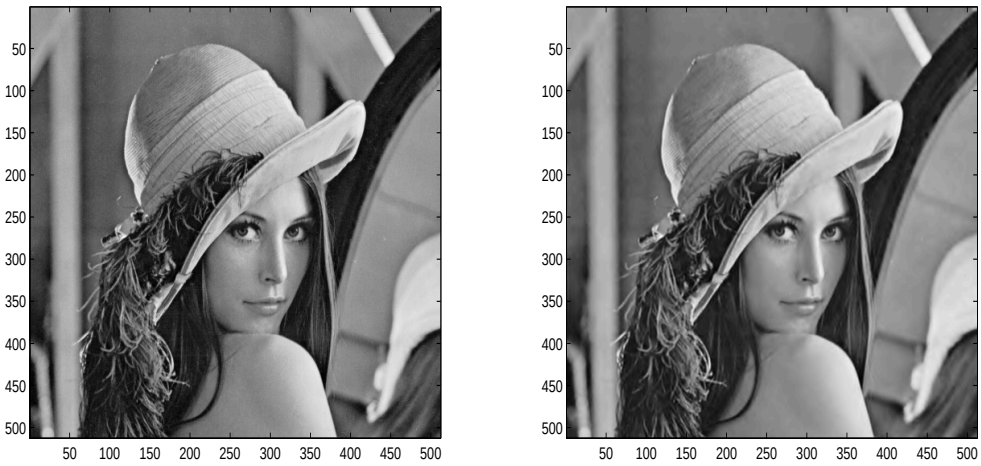


FIGURE 19: Reconstruction: Image original (gauche), Décodeur central $D_1 + D_2 + D_3$ (droite) à 0.8 bpp.

Comparaison entre l'approche MDC classique et l'approche par trame de synthèse

Bien que les deux approches présentées ici sont dans un sens complémentaires il pourrait être utile de comparer leurs performances sans tenir compte qu'elles s'appliquent dans des différents cas de figure. Dans le cas que nous venons de décrire les schémas sont construits tel que la complexité est transférée à l'encodeur, en gardant ainsi la linéarité des décodeurs.

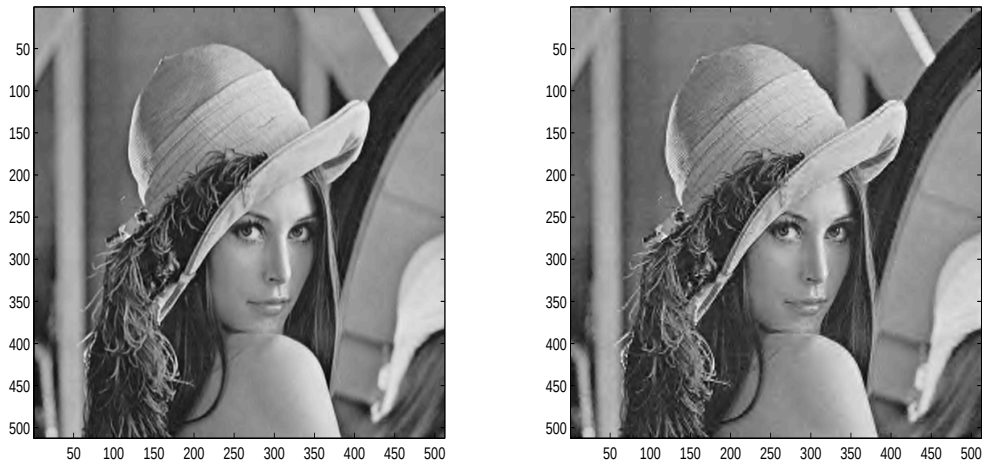


FIGURE 20: Reconstruction: Décodeur latéral $D_1 + D_2$ (gauche), Décodeur latéral D_1 (droite).

Cependant, l'idée de base dans cette approche reside dans la recherche du meilleur pas de quantification devant être appliqué aux coefficients pour obtenir une distorsion maximum acceptable. L'approche précédentes se proposaient surtout de minimiser la redondance en laissant la tâche de reconstruction aux décodeurs quelque soit le pas de quantification utilisé.

Comme résultats préliminaires, nous avons comparé ces deux types de méthodes sous quelques hypothèses simplificatrices portées sur l'approche classique à fin de ramener les experiences au même niveau (des trames très redondantes). Plus concrètement, nous avons comparé l'approche par paradigme de synthèse à une approche par trame classique à deux descriptions dans lesquelles le pas de quantification a été choisi pour obtenir des débits équivalents.

Le résultat de cette comparaison pour l'image de test Barbara est donnée dans la Figure 21. Dans la figure de gauche nous présentons les décodeurs centraux (les deux descriptions reçues) ainsi que une décomposition en ondelettes biorthogonales $9/7$ échantillonnée critique, donnant une borne de performance. Les performances des décodeurs latéraux sont présentées dans la figure de droite. Ces résultats montrent des performances équivalentes pour les deux types d'approches au niveau des décodeurs centraux, avec un léger gain en faveur de la méthode par trame de synthèse à haut débit. Les résultats obtenus pour les décodeurs latéraux contredisent ceux des centraux, ce qui pourrait indiquer que certains raffinements des hypothèses utilisées sont nécessaires.

Perspectives

Plusieurs perspectives s'ouvrent à l'issue de cette thèse, soit sous la forme d'extensions des techniques proposées à fin d'améliorer les performances, soit en les appliquant à d'autres problématiques qui s'y prêtent. Pour conclure ce résumé, nous passons en revue quelques unes de ces perspectives.

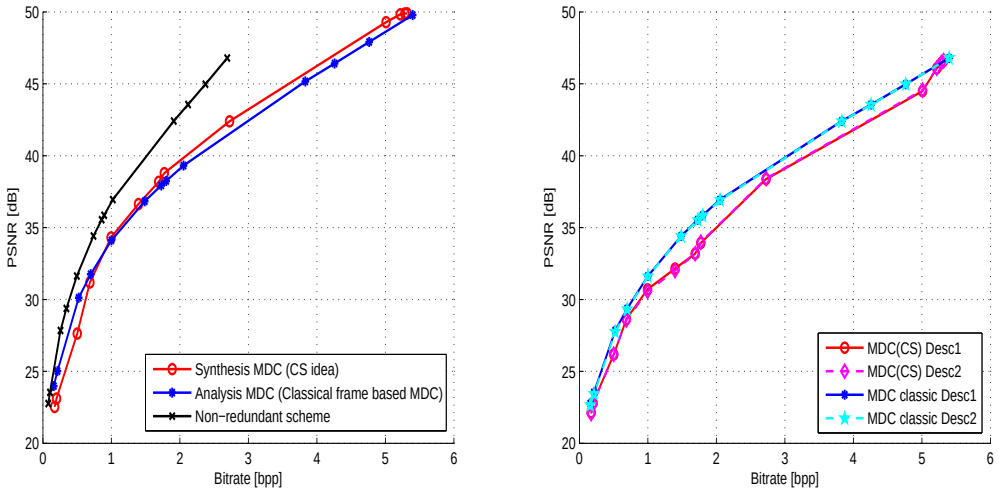


FIGURE 21: Comparaison entre l'approche par paradigme de synthèse et une approche MDC classique avec une décomposition en trame équivalente. L'image de test utilisée est Barbara, 512×512 pixels. Décodeurs centraux (gauche), Décodeurs latéraux (droite). Pour le décodeur central nous présentons également la courbe PSNR en fonction du débit pour une décomposition en ondelettes biorthogonales $9/7$ échantillonnée critique.

Des schémas MDC hybrides pour la vidéo

Les descriptions temporelles (inter-trame) pourraient être par la suite combinées avec d'autres stratégies de rajout de redondance dans le domaine spatial de la vidéo (intra-trame), que nous avons vu proposées dans la littérature de spécialité. Ceci augmenterait la flexibilité des schémas MDC ce qui permettrait de mieux adapter la construction des paquets à des différentes situations de pertes survenues dans les réseaux. Un deuxième effet de cette extension serait une amélioration de la reconstruction à des différents niveaux.

Des schémas spatiaux améliorées

Les performances des schémas à deux descriptions que nous avons proposées dans le cas des images fixes pourraient être améliorées en rajoutant plus de diversité entre les sous-bandes les plus énergétiques - les sous-bandes d'approximation. En effet, à l'issue de la décomposition en trame d'ondelettes, celles ci gardent une très forte corrélation et donc le décodeur central n'exploite pas la redondance introduite (en termes de débit) à son plus haut potentiel. Nous proposons, par exemple, de rajouter une transformation de type codage par descriptions multiples avec quantification scalaire (en anglais Multiple Description Scalar Quantization - MDSQ) entre les deux sous-bandes d'approximation. Ceci serait réalisé en décalant les indices de quantification par $1/2$, de la manière de Vaishampayan [Vai93b]. La reception des deux description fournirait une sous-bande d'approximation non seulement non-décimée mais également quantifiée avec un pas de quantification plus fin qui devrait mener à un meilleur décodage.

Une autre extension que nous pourrions apporter aux schémas spatiaux consisterait dans une stratégie plus adaptative concernant le choix des poids dans l'initialisation de l'algorithme de décodage itératif. En effet, au lieu d'approcher les valeurs du pas de quantification à 1

(moyennant une troncature aux entiers près) pour chaque débit, nous pourrions imaginer un pas de quantification adaptatif.

Des trames moins redondantes pour l'approche qui exploite le caractère creux des données

La dernière direction explorée dans cette thèse est à ses débuts. Plusieurs extensions possibles peuvent être envisagées. La première serait d'employer des trames moins redondantes, telles que celles explorées dans les chapitres précédents. Aussi, nous pourrions considérer des approximations plus fines de l'entropie qui seraient mieux adaptées au régimes bas-débit, en s'appuyant par exemples sur les travaux de Fraysse et al. [FPPP09].

D'autres transformées et critères de performance

Une problématique importante dans toute approche MDC est la qualité de la reconstruction latérale. Dans notre travail, nous avons considéré uniquement des critères d'évaluation de performance "usuels", tels que le PSNR. Deux directions sont possible à partir de ce point.

La première serait d'employer d'autres transformées qui augmentent la qualité subjective perçue à PSNR comparable. Dans ce sens, un effort considérable à été déployé dans la littérature visant par exemple d'amélioration subjective de la qualité d'image perçue avec des transformée préservant la géométrie. Nous pourrions explorer le remplacement des bancs de filtres proposés par des techniques d'ondelettes plus avancées telles que : bandelettes [Pen02], curvelets [CDDY06], contourlets [SW01, Chap. 4], des schéma en lifting adaptatif [PH02], etc. . [SW01] et [Mal09] fournissent d'excellents revues sur les autres transformées qui s'apparentent aux ondelettes (les "x-let"). Nous pourrions envisager de remplacer notre décomposition par trame d'ondelettes construite comme une union de bases par une de ces transformées. Un effet secondaire bénéfique serait une representation plus creuse du signal et, implicitement, des meilleures performances en compression. La deuxième direction qui mérite d'être étudiée pour l'evaluation des performances du décodage latéral serait de changer les critères en s'orientant vers des mesures de qualité perceptuelles, par exemple [vdBLV96], [dFZRS03].

Des applications inter-disciplinaires

Récemment, les méthodes de MDC ont été évaluées pour le développement de la vidéo/télévision stereoscopique 3D [NAB+06], [KHWK08]. Une méthode de MDC simple basée sur la repartition temporelle selon les indices paires et impaires à été proposée dans [KHWK08]. Ces travaux s'appuient sur ceux de Apostolopoulos [Apo99] dans les codecs vidéos à boucle fermée. Nous envisageons une approche similaire en utilisant nos méthodes à redondance réduite. Dans ce contexte, MDC est un très bon candidat pour l'encodage de la couche de base du signal stereoscopique.

Enfin, dans une note un peu plus générale, nous remarquons que la philosophie des méthodes MDC n'est pas restreinte au domaine du traitement du signal (dans la définition du signal communément accepté). L'expansion rapide des architectures de calcul multi- et many-coeurs fait ressortir le besoin de tolérance aux erreurs à différents niveaux du système : materiel, système d'exploitation, middleware, bus de communication etc. Dans ce contexte, il y a une analogie à faire entre un réseau à paquets et une architecture multi-coeur, notamment quand il s'agit d'architectures reconfigurables. Dans ce cas aussi il est commun d'introduire de

la redondance dans le système, que cela soit au niveau logiciel ou materiel, à fin de palier des potentielles erreurs de fonctionnement. De ce fait, le cadre MDC se trouve très en ligne avec cette problématique. En outre, ces architectures disposent d'un degré très élevé de parallélisme, ce qui complète le rapprochement avec la philosophie MDC. Par ailleurs, l'exploitation efficace du parallélisme dans les architectures de calcul est un des problèmes ouverts dans la communauté informatique actuelle. Concernant le traitement d'images, les algorithmes présentés dans cette thèse se prêtent très bien au traitement parallèle. Une prochaine étape pourrait être l'implémentation des schémas MDC proposés dans le chapitre 3 sur une architecture parallèle à fin d'évaluer ses performances de calcul dans des scénarios plus réalistes.

Publications

Revues internationales

1. C. Tillier, T.Petrisor, B. Pesquet-Popescu et J.-C. Pesquet - "A Motion-Compensated Oversampled Temporal Decomposition for Multiple Description Scalable Video Coding" - EURASIP Journal on Image and Video Processing, Volume 2007 (2007), Article ID 31319, 12 pages doi:10.1155/2007/31319

Conférences internationales

2. T. Petrisor, A. Fraysse, B. Pesquet-Popescu et J.-C. Pesquet - "Une nouvelle approche du codage par descriptions multiples, utilisant des representations linéaires creuses", dans les actes de GRETSI, Troye, France, September 2007

3. T. Petrisor, B. Pesquet-Popescu et J.-C. Pesquet - "A Compressed Sensing Approach to Frame-based Multiple Description Coding", dans les actes de l'IEEE International Conference on Speech, Acoustics and Signal Processing (ICASSP), Honolulu, USA, Avril 14-18, 2007

4. T. Petrisor, B. Pesquet-Popescu et J.-C. Pesquet - "Redundant wavelet schemes for Multiple Description coding", poster à WaVE 2006, EPFL, Suisse, juillet 2006

5. T. Petrisor, B. Pesquet-Popescu et J.-C. Pesquet - "Perfect reconstruction in reduced redundancy wavelet-based multiple description coding of images" , dans les actes d'EUSIPCO 2005, Antalya, Turquie

6. T. Petrisor, B. Pesquet-Popescu et J.-C. Pesquet) "Wavelet-based multiple description coding of images with iterative convex optimization techniques", dans les actes d'IEEE ICIP 2005, Gênes, Italie

7. C. Tillier, T. Petrisor, B. Pesquet-Popescu et J.-C. Pesquet - "Codage par descriptions multiples pour la transmission vidéo", dans les actes de GRETSI 2005, Louvain-la-Neuve, Belgique

8. T. Petrisor, C. Tillier, B. Pesquet-Popescu et J.-C. Pesquet - "Comparison of Redundant Wavelet Schemes for Multiple Description Coding of Video Sequences" - dans les actes de l'IEEE ICASSP 2005, Philadelphie, USA

9. T. Petrisor, C. Tillier, B. Pesquet-Popescu et J.-C. Pesquet - ”Redundant Multiresolution Analysis for Multiple Description Video Coding” - IEEE International Workshop on Multimedia Signal Processing, Sienne, Italie, 2004 (article invité)

Introduction and thesis outline

Consider the simplified data transmission scheme in Figure 22. A digital signal, let us say, x, has to be transmitted over some network. Usually, the path between the sender and the receiver contains several processing blocks, for instance: some transform that better prepare the signal for transmission (by better concentrating its energy for instance), a quantizer which only allows a certain set of values and, finally, an encoder which will allocate a binary sequence to each sample to be transmitted. Then, the bits are grouped into packets. In this manner a bitstream is formed and sent over some network channel. This channel is usually subject to perturbations which can alter the bitstream and thus prevent it from being correctly recovered at the receiver, hence the notation $\hat{x}$ for the decoded signal. Obviously, the ultimate goal of the transmission is to minimize the error $\epsilon = \|x - \hat{x}\|$, ideally the two signals should be identical or within a tolerable difference[4].

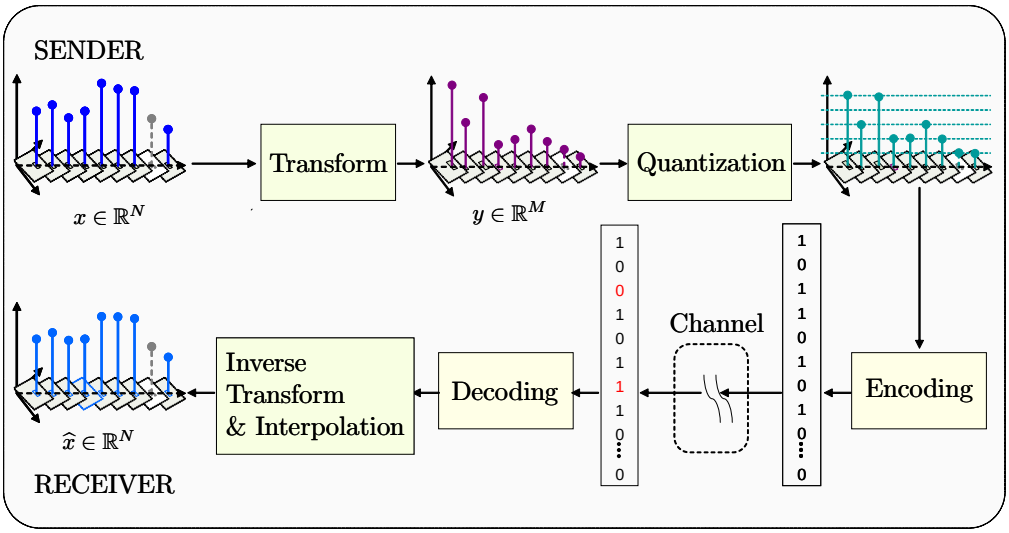


FIGURE 22: Simple data transmission scheme.

There are several strategies available in order to tackle this problem. One is the retransmission of corrupted packets, but this introduces delays which are not always an option. Another one is sending a larger bitstream which contains in addition some error correction code. This is limited to only a few wrong bits in each packet. Then there is the option of the so-called layered coding. This strategy forms several bitstreams which are progressively refinable starting from a base layer. The base layer, however, is essential to data reconstruction with a minimal quality. This means that if the base layer gets corrupted we are back to the initial problem. Then some

[4]Some irreversible transformation might occur for instance if the signal suffers a lossy compression.

protection techniques for this base layer have been employed.

The increasing usage of the Internet and other best-effort networks for diverse multimedia communications, brings with it a stringent need for reliable transmission. For a long time, the research efforts have been concentrated on enhancing the existing error correction techniques, but during the last decades an alternative solution has emerged and gained more and more popularity. This solution mainly answers to the situation in which immediate data retransmission is either impossible (network congestion or broadcast applications) or undesirable (e.g. in conversational applications with very low delay requirements). We are referring to a specific joint source-channel coding technique known as *Multiple Description Coding* (MDC).

Multiple description coding builds several correlated but independently decodable (preferably with equivalent quality) bitstreams, called *descriptions*, that are to be sent over as many independent channels. In an initial scenario these channels are working in a binary manner, in other words, if an error occurs on one channel this is considered entirely damaged and the conveyed bitstream is unusable at the so-called *side decoder* end. As in other robust coding methods, some amount of redundancy has to be added to the source signal, such that an acceptable reconstruction can be achieved from any of the bitstreams. Then, similar to layered coding, the reconstruction quality will be enhanced with every bitstream received, maximal reconstruction quality being attained at the so-called *central decoder*. The major difference with layered coding is that all "layers" have equal importance in MDC.

An ingredient enabling the success of an MDC technique is the path diversity since its usage balances the network load and reduces the congestion probability. In wireless networks, for instance, a mobile receptor can benefit from multiple descriptions if these arrive independently, for example on two neighbour access points; when moving between these access points it might capture one or the other, and in some cases both. Another way to take advantage of MDC in a wireless environment is by splitting in frequency the transmission of the two descriptions: for example, a laptop may be equipped with two wireless cards (e.g., 802.11a and g), each wireless card receiving a different description. Depending on the dynamic changes in the number of clients in each network, one of them may become overloaded and the corresponding description may not be transmitted.

In wired networks, the different descriptions can be routed to a receiver through different paths by incorporating this information into the packet header. In this situation the initial scenario of binary working channels might no longer be of interest, since for a typical CIF format video sequence one frame might be encoded into several packets. Therefore, the system should be designed to take into consideration individual or bursty packet losses rather than a whole description.

Thesis objective

An important issue that concerned the researchers over the years is the amount of introduced redundancy. In order for the transmission to be efficient, one has to consider a trade-off between this redundancy and the resulting distortion.

This thesis focuses on new approaches to Multiple Description Coding in low-redundancy scenarios. We will present their application to the transmission of still images and video sequences. To this end, we have proposed new schemes based on wavelet frame decompositions, which for computational convenience are implemented in a lifting form.

Outline of the work

Background on Multiple Description Coding

This dissertation begins with a short state-of-the-art in Multiple Description Coding, in Chapter 1. The topic of MDC has quite a long life cycle, its theoretical bases being laid out more than almost forty years ago. Then a quiet period followed until about a decade ago when the new transmission scenarios and applications made possible the quest for a wide variety of MDC methods. This chapter does not pretend to be exhaustive, the amount of work on the topic making it almost untraceable. However, it will mention the research guidelines that established the main MDC trends. The purpose of this chapter is to place our proposed strategy among the existent directions in MDC, highlighting the limitations of some of these methods and the motivations behind this work.

Temporal MDC Schemes

The first contribution of this thesis will be presented in Chapter 2. Here, we study new methods of building two descriptions in the temporal axis of a $t + 2D$ video codec. The redundancy of the schemes is inherent to the wavelet frame transform which is equivalent to an oversampled filter bank. Keeping the whole set given by this decomposition would yield a redundancy of a factor of 2 which could be highly inefficient if both paths were error-less. In our schemes we perform an additional subsampling of the detail subbands while keeping the obtained approximation subbands entirely. Thus the redundancy is tuned to the size of an approximation subband in a classical wavelet decomposition. However this raises a new problem which is the perfect reconstruction of such a scheme. In this part we have proven the perfect reconstruction for certain schemes and we have established choice criteria among them based on the minimization of the quantization noise. We have compared the performances of several schemes among the efficient ones in a scalable video coding context provided by the MC-EZBC (Motion Compensated - Embedded Zero-trees Block Coding) codec. Two scenarios (as presented in the previous section), that is: losing a whole descriptions versus losing only packets in each description have been implemented and the results have been compared to the classical critically sampled decomposition.

Spatial MDC Schemes

A second direction that we have explored in this thesis refers to the MDC of still images, which is viewed as an extension of the temporal schemes developed in the first part. This work will be presented in Chapter 3. The same idea of reducing the redundancy by additional subsampling of the wavelet subbands is applied (the subbands which contain the core of the signal energy are again preserved). The problem of structure invertibility is not trivial in these two-dimensional schemes and an exhaustive study has been conducted in order to select the efficient schemes among all possible combinations based on the proposed subsamplings. Moreover, we have explored the possibility of improving the decoding by a post-processing based on a priori information on the system. This information is given by the quantization steps which can be viewed as convex constraints. The reconstruction problem has thus been formulated as the optimization of a quadratic function under convex constraints and the decoded image

gains several dB in terms of Peak Signal to Noise Ratio (PSNR) both when a whole description is lost and when random pixels in each description are destroyed.

A complementary approach exploiting sparsity

Finally, in Chapter 4 we have approached the Multiple Description Problem from a completely different angle, by considering the problem as a rate-distortion optimization in which a certain maximal distortion is allowed, and the best transmission rate (linked to the quantization step) is searched. Some approximations have been formulated in order to be able to solve this problem as a convex optimization as well and the scheme has been generalized to an arbitrary number of descriptions. By doing this the complexity is shifted to the encoder whereas the decoding becomes a simple linear process. We have tested this theory for still images.

For the image applications we have tested two image coders: EZBC and JPEG2000. Most of the presented results will involve the latter though, and this is because it yields slightly better performances, but also mainly because, as a standard, it facilitates referencing.

Chapter 1

Background on Multiple Description Coding

An important issue appearing in heterogeneous packet-switched networks is the handling of packet losses, that might be due either to network congestions or other perturbations generating transmission errors. If a preferential treatment can be performed on the packets allowing some protection for the more important ones, then a very good solution is layered source coding (see for instance [KZC+94], [CHG03]). Otherwise a different technique might be employed, and this has come to be known as Multiple Description Coding (MDC).

Multiple description coding involves the transmission of several correlated representations of the source signal over independent channels. Before discussing the theoretical aspects that have been investigated in this new transmission system, let us consider the simplest case, represented in Figure 1.1. Here, we have to transmit the source X_k and we know that two in-

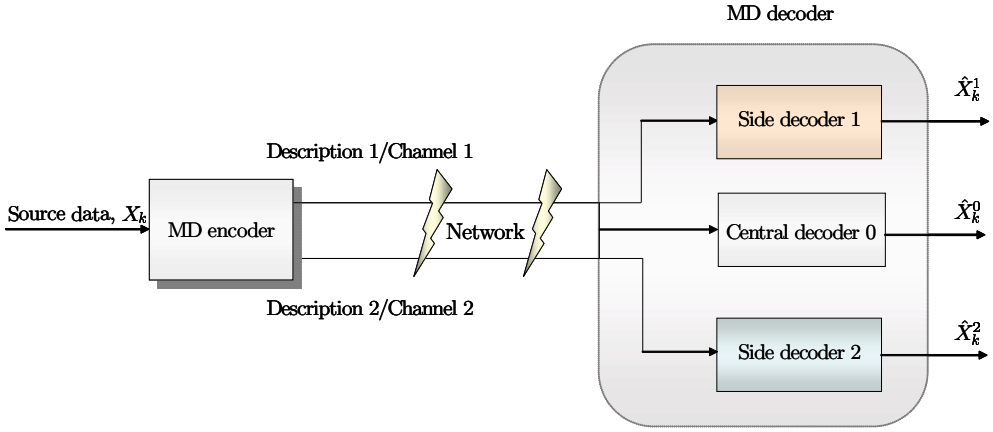


FIGURE 1.1: Two-channel scheme for multiple descriptions.

dependent channels are simultaneously available, in an error-prone network. Therefore two representations (called *descriptions*) are generated by the MD encoder and each of those is sent over its private channel. The MD decoder handles two different situations: in the first one, errors have occurred on one of the channels and the decoder ignores the data coming from it, delivering an approximate version of X, $\widehat{X}_k^j$ with $j \in \{1, 2\}$; in the second situation both channels were unaffected by errors and a central decoder produces a (usually) better version of X,

$\widehat{X}_k^0$.

The question is now: what should these two different representations of X_k be and how can the reconstruction $\widehat{X}_k^i$, $i \in \{0, 1, 2\}$ be best obtained?

An important issue involved in source coding is the use of transforms that minimize the correlation within the signal to be transmitted, thus eliminating the unnecessary redundancy. This is mainly achieved with orthogonal transforms, such as the Discrete Cosine Transform or some wavelet transform. A criterion for the transform's efficiency is the sparse characteristic of the obtained signal. While this idea is very useful in source coding as it leads to highly compressible structures, its end-to-end performance is endangered by channel losses.

Channel coding, on the other hand, focuses on the correct transmission of the data source and thus on the concealment of the eventual errors occurring on the channel. The famous source/channel "separation" principle given by Shannon [Sha48] gives an asymptotic result for the transmission of any source at a maximal bitrate with any desired error probability. This, however, involves maximum length codewords, thus raising a real obstacle to the practical implementation.

In these conditions it has been discovered that building source and channel codes jointly leads to better results. In the joint source-channel coding strategy the source is compressed at a lower ratio, or some particular redundancy is introduced in order to ensure better robustness to transmission. Multiple description coding can be assigned to this class of methods, being especially suitable for erasure channels, in which, even though great amount of data might be lost, one can benefit from the fact that the received data is guaranteed to be accurate.

A note on the terminology that we are going to use from now on (and which has already been introduced in the literature): we shall refer to an encoder which was optimized for coding efficiency, in other words a "non-redundant" coder, as a *Single Description* (SD) coder. Obviously the "redundant" coders that we are studying here are referred to as Multiple Description (MD) coders.

1.1 Information Theory framework

The beginnings of this new coding strategy date as far as 1979 when Gersho, Ozarow, Witenhausen, Wolf, Wyner and Ziv formulated the following question: *What are the achievable distortions for a memoryless source at fixed given transmission bitrates when this source is described by several bitstreams [Oza80],[GC82]?*. This problem was tackled from an information theoretic point of view. In this period the problem was mostly cast in the literature as a source coding technique, probably because of the rate-distortion optimization philosophy. Later on, Goyal, Vaishampayan and others included MDC in the class of joint source-channel coding but consensus has not yet been completely reached among researchers. We are leaning toward the latter classification, since the MDC strategy takes simultaneously into account the possibility of losses and the encoding process.

The main idea was to determine (in the Shannon sense [Sha48]) the set of rates and distortions $\{R_i, D_j\}$ with $i \in \{1, \ldots, N\}$ (N being the maximal number of descriptions) and $j \in \{1, \ldots, 2^N - 1\}$ (the number of decoders corresponding to the scheme), where R stands for the transmission rate and D stands for the obtained decoder distortion. If we consider that the signal to be transmitted is denoted by X_k and the j^{th} reconstruction (at the j^{th} decoder) is

denoted by $\widehat{X}_k^{(j)}$ then the distortion of this decoder is given by:

$$D_j = \frac{1}{N}\sum_{k=1}^{N} E(d_j(X_k, \widehat{X}_k^{(j)})) \tag{1.1}$$

where d_j are some chosen distortion measures.

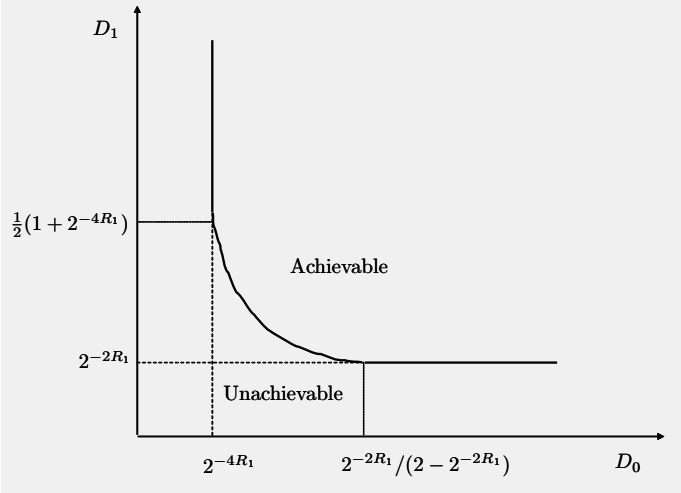


FIGURE 1.2: Achievable central and side distortions for MDC of a memoryless Gaussian source with squared-error distortion and equal individual rates and distortions.

In the two-description case the receiver has three decoders and one has to find the set of reachable values $\{R_1, R_2, D_0, D_1, D_2\}$, where R_1 and R_2 represent the side rates and D_i, $i \in \{0, 1, 2\}$ represent the central and sides distortions, respectively, as before. To this end, the following theorem has been formulated for a set of Gaussian random variables, [GK01].

Theorem 1. *Given the i.i.d Gaussian random variables $\{X_k\}_{k\in\{1,\ldots,N\}}$ (the sequence of source symbols to be transmitted), the set of reachable distortion and rate values under a quadratic norm satisfies:*

$$D_1 \geq 2^{-2R_1} \tag{1.2}$$
$$D_2 \geq 2^{-2R_2} \tag{1.3}$$
$$D_0 \geq 2^{-2(R_1+R_2)}\gamma(D_1, D_2, R_1, R_1) \tag{1.4}$$

with

$$\gamma = \begin{cases} \frac{1}{1-(\sqrt{(1-D_1)(1-D_2)}-\sqrt{D_1D_2-2^{-2(R1+R2)}})^2} & \text{if } D_1 + D_2 \leq 1 + 2^{-2(R_1+R_2)} \\ 1 & \text{otherwise.} \end{cases} \tag{1.5}$$

The case leading to equality in (1.2) and (1.3) has been studied in [ZB95], whereas the equality in (1.4) has been treated in [Ahl85].

In the perfectly balanced case in which the two generated descriptions are equivalent, that is $R_1 = R_2$ and $D_1 = D_2$, the achievable side-central distortion region is given in Figure 1.2. If we

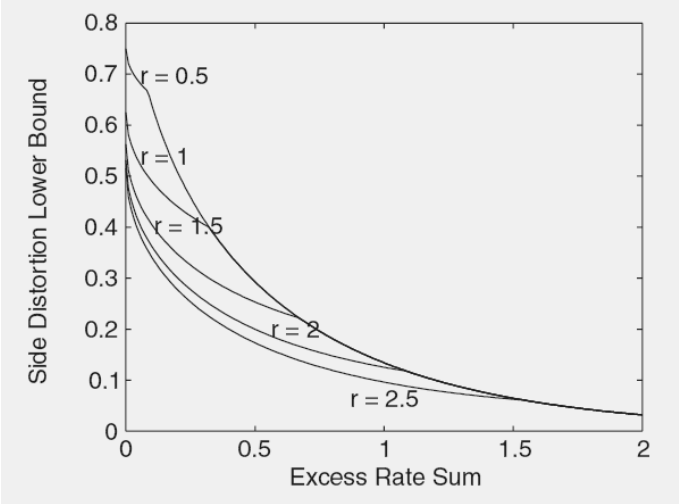


FIGURE 1.3: Side distortion in the balanced descriptions case for different redundancy rates, r (taken from [Goy01]).

express the global rate as the sum of the base rate R_0 (corresponding to the central distortion D_0 in the source-coding sense) and the redundancy r for a Gaussian memoryless source, we have the central distortion $D_0 = 2^{-2R_0}$ and the redundancy is the excess rate introduced in order to get acceptable side distortions: $r = R_1 + R_2 - R_0$. Thus, the side distortion is lower bounded by, [Goy01]:

$$D_1 \geq \begin{cases} \frac{1}{2}[1 + 2^{-2R_0} - (1 + 2^{-2R_0})\sqrt{1 - 2^{-2r}}], & \text{for } r \leq R_0 - 1 + \log_2(1 + 2^{-2R_0}) \\ 1 - \sqrt{1 + 2^{-2R_0}}, & \text{for } r > R_0 - 1 + \log_2(1 + 2^{-2R_0}) \end{cases} \tag{1.6}$$

A representation of this bound for different redundancy rates is given in Figure 1.3.

The M description case with $M > 2$ has been tackled in [VKG03], [WV07] for Gaussian sources.

The information theory community has thus formulated the interest of using a multiple description approach and it has given the theoretical bounds for the so-built systems. However, while we knew what to expect in terms of performances, no practical method had been formulated so far. This has led to almost two decades during which multiple description coding did not find applications in the communication world.

The issue has known a spectacular regain of interest when researchers such as Vaishampayan, [Vai93b], Wang, Orchard, Reibman [WOR97], Kovacevic and Goyal [GK98] proposed viable methods for error resilience via multiple description coding. These works were motivated by the important advances in multimedia communications.

In the next three sections we shall proceed to a brief overview of these practical codes, which highlights the major directions that have been developed in the last decade. Most of the current applications that have emerged in the literature afterward are based on these main directions. After explaining the theoretical framework we shall give a survey of the MDC techniques applied to the transmission of still images and video sequences, data which makes the object of this thesis. A global picture of the MDC directions can be observed in Figure 1.4. The two main directions that we are presenting in the sequel involve either quantization or correlation. They provide the basis for the more complex methods developed later with a more specific

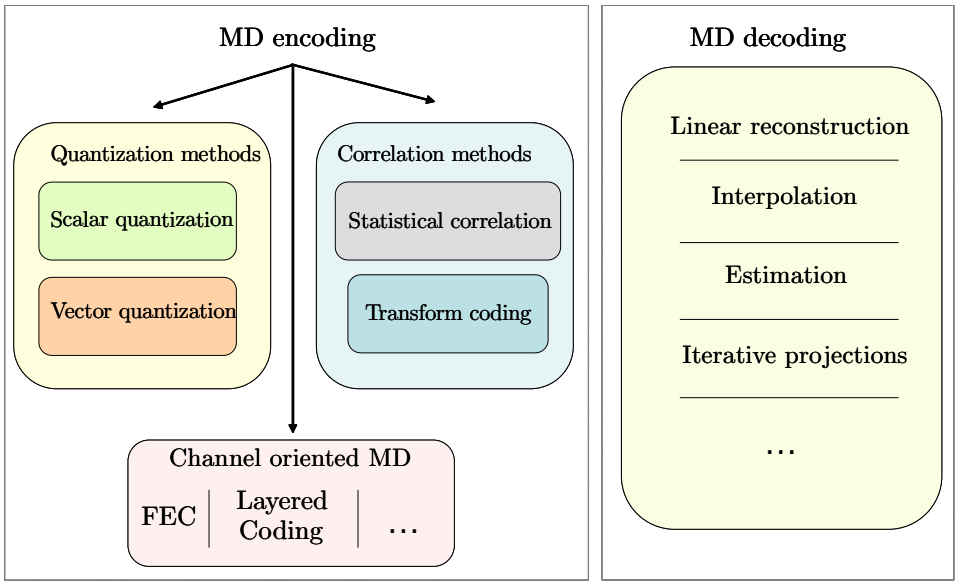


FIGURE 1.4: The major directions in Multiple Description Coding.

application-oriented strategy in mind. Different decoding solutions are also investigated in the literature and they aim at enhancing the reconstruction quality with respect to the rate-distortion trade-off that gives the theoretical bounds.

1.2 MD by quantization

The first practical approach to Multiple Description Coding is proposed by Vaishampayan, [Vai93b]. This technique relies on quantization and the idea is to build two discrete descriptions X_1 and X_2 for a source X, each of them belonging to a certain dictionary of symbols: $\hat{\chi}_1$ or $\hat{\chi}_2$. The imposed criterion is that the resulting quadratic distortion when both channels work correctly is smaller than the individual side distortions.

However, since both side decoders must provide an acceptable distortion, the idea of a hierarchical quantization in which the first description would be a coarse version, X_1, of X and the second one would be the residual $X - X_1$, is not a good strategy. A different solution has been proposed, based on scalar quantization. In this approach two uniform quantizers of step δ are involved and the second one is shifted by half a quantization interval with regards to the first one. Thus, if one description is lost, the source is recovered from a description quantized with a step of δ, whereas if both descriptions are received the resulting quantization step is $\frac{\delta}{2}$.

An important remark concerns the transmission rate. One must notice that in the hypothesis of high resolution such a scheme is very consuming in the sense that the rate will be doubled in order to quantize the source at $\frac{\delta}{2}$, when only an extra bit would suffice.

As in traditional quantization, multiple description quantization follows the same techniques: scalar and vector quantization or entropy-constrained quantization (see, for instance, [GG92],[Pro01]). We shall explain these techniques in more detail in the following.

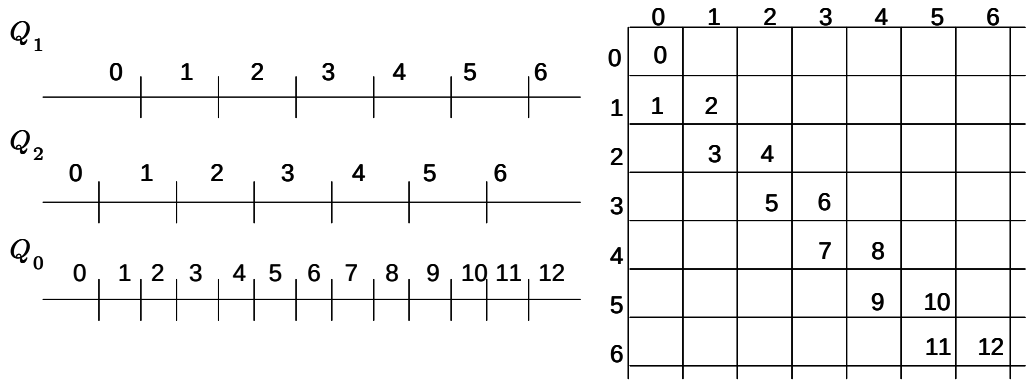


FIGURE 1.5: Nested-cells scalar quantization ("linear" index representation (left), matrix representation (right)).

1.2.1 Scalar quantization

MD Scalar quantization (MDSQ) without entropy coding consists in coding a stationary memoryless source $(X_k)_{k\in\mathbb{Z}}$, having zero-mean and σ_X^2 variance, into two or more quantized descriptions.

A scalar quantizer is given by a dictionary of reconstruction levels $\widehat{\chi} = \{\widehat{X}_1, \widehat{X}_2, ...\widehat{X}_N\}$ of the source and a partition of $\mathbb{R}$, $\mathcal{A} = \{A_1, A_2, ..., A_N\}$. With these notations, the encoder is an application $Q : \mathbb{R} \to \{1, 2, \ldots, N\}$ such that $Q(X) = i, \forall X \in A_i$. The decoder is the inverse application $Q^{-1} : \{1, 2, \ldots, N\} \to \widehat{\chi}$ which associates the index i with a word from the corresponding dictionary: $Q^{-1}(i) = \widehat{X}_i$.

Each encoder is defined by its partition $\mathcal{A}^j$ and an index set $\mathcal{I}_j = \{1, 2, \ldots, M_j\}$, $j \in \{1, 2\}$ and the central partition is given by the intersection of these $\mathcal{A}^j$ whereas the central index set is $\mathcal{I} \subset \mathcal{I}_1 \times \mathcal{I}_1$. If the sizes of the two quantizers are known and the side distortions are fixed, then the problem is to find the optimal quantizer which will minimize the central distortion D_0 under the side distortion constraints $D_j \leq d_j$, with $j \in \{1, 2\}$. The usual measure for the distortion is $D_i = E\left|\widehat{X}_i - X\right|^2, i \in \{1, 2\}$ for the two side distortions and $D_0 = E\left|\widehat{X}_0 - X\right|^2$ for the central distortion.

Finding the optimal quantizer is not a trivial problem, but the conditions for the optimality can be given through a Lagrangian formulation. The optimization can be done in two parts: in the first one the central partition - and implicitly the quantizers Q_j - is optimized for fixed decoding dictionaries and in the second part the dictionaries are optimized for a given central partition. Alternating these two parts can lead to an optimal MDSQ. From classical scalar quantization (e.g. [GG92], [GN98]) it is known that the minimal reachable distortion depends on the quantization cells spread, being proportional to its square, under high-rate assumptions. While the central distortion is mainly related to the number of quantization cells, an additional difficulty has to be overcome in order to minimize the side distortions, because these are highly dependent of the index assignment in each of the side quantizers. Results are given by Vaishmapayan in [Vai93b] who, further on and together with Battlo, also presents an asymptotic analysis of these MDSQ in [VB98]. This work is then further extended to entropy-constrained scalar quantization, [VD94].

A graphical representation of the simplest MDSQ scheme is given in Figure 1.5 in which the

quantization cells from the two descriptions are nested such that the central decoder Q_0 has an index assignment corresponding to a two-times finer quantizer.

The dependencies between the quantized variables have been studied in [GGF02].

Entropy-constrained scalar quantization: In the previous case, equal length code words had been chosen for the encoding of the scalar quantized descriptions. This has been extended [VD94] by using variable length encoding and by changing the rate-distortion optimization problem accordingly. This is achieved by considering the entropies of the random variables $Q_j(X)$, $j \in \{1, 2\}$, denoted by $H_1(\boldsymbol{t})$ and $H_2(\boldsymbol{t})$, respectively, and given by:

$$H_j(\boldsymbol{t}) = -\sum_{m=1}^{M_j} P_{jm} \log_2(1/P_{jm})$$

where:

- $\boldsymbol{t}$ is a vector of thresholds describing the quantization operation as follows: $Q(x) = n$ if $x \in [t_{n-1}, t_n)$, $n \in \{1, \ldots, N\}$, and $\boldsymbol{t} = (t_1, \ldots, t_N)$ with $t_1 \leq \ldots \leq t_N$ (in this case$[t_0, t_N)$ is the support of the probability density function of the source),
- P_{jm} is the probability of the j^{th} decoder for the quantization index $m \in \{1, \ldots M_j\}$.

Then, if we denote the average distortions by $\overline{D}_k$ with $k \in \{0, 1, 2\}$, one has to solve the following problem:

$$\text{Minimize} \quad \overline{D}_0(\boldsymbol{t}, \hat{X}_0)$$

subject to the following constraints:

$$\begin{aligned} \overline{D}_j(\boldsymbol{t}, \hat{X}_j) &\leq D_j, \quad j \in \{1, 2\}, \\ H_j(\boldsymbol{t}) &\leq R_j, \quad j \in \{1, 2\}, \end{aligned}$$

where R_j, $j \in \{1, 2\}$ are the side rates (as before).

1.2.2 Vector quantization

Detailed presentations of these techniques can be found in [Vai93b], [VB98], [VSS01], [KGK00], [DSV02a].

The lattice vector quantizers are quite similar to scalar quantizers, but the source is split into vectors of length L. A first lattice $\Lambda \subset \mathbb{R}^L$ leads to a finely quantized vector $Q(X) = \lambda \in \Lambda$ which is subdivided into two descriptions λ'_1 and λ'_2, each in a coarser sub-lattice Λ' of Λ. The indexing function is given by the injective mapping $\alpha : \Lambda \to \Lambda' \times \Lambda'$, such that $\alpha(\lambda) = (\lambda'_1, \lambda'_2)$.

Such a representation will be decoded simply by inversing the indexation at the central decoder: $\widehat{X}^0 = \alpha^{-1}(\lambda'_1, \lambda'_2)$ and, if only one representation has reached the decoder, then $\widehat{X}^i = \lambda'_i$.

Note that the sub-lattices Λ' of Λ are geometrically similar to Λ or, in other words, $\Lambda' = c\Lambda U$ for an orthogonal matrix U and a scalar c. An example of admissible sub-lattices is given in Figure 1.6, the additional condition being that no elements of Λ should lie on the boundaries of Voronoi regions of Λ'.

As in the scalar case, the index assignment is obtained by solving the rate-distortion optimization of the central distortion D_0 under side rate constraints.

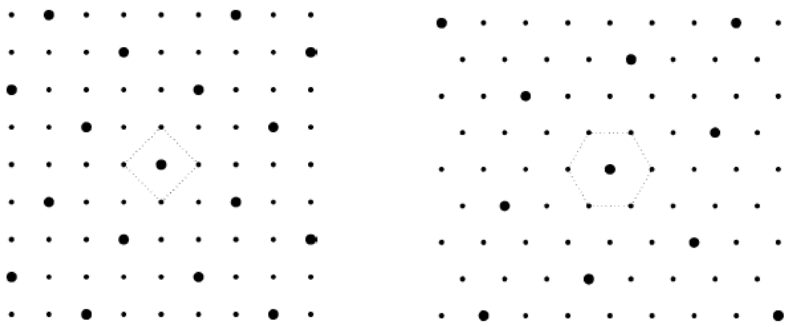


FIGURE 1.6: Examples of trellis quantization lattices. The sub-lattices Λ' are marked by thicker points and the Voronoi cells associated to a point of the sub-lattice are equally represented. The left figure represents a $\mathbb{Z}^2$ lattice while the right figure gives a hexagonal lattice (taken from [GKK02]).

1.3 MD by correlating transforms

In the previous section we have presented the methods which introduce the redundancy after the signal has been decorrelated, by diversifying the quantization. The second direction that has been investigated by researchers aims at introducing the redundancy through the used transform. Most of the strategies investigated in this direction are applied directly to the source signal and prior to the quantization operation. The redundancy is introduced by a correlating transform and the result is further quantized. The place of the quantization block can also be inverted with that of the transform one, as it will be presented shortly, and in this situation one can benefit from the properties of integer-to-integer transforms.

The two main strategies that have been developed in this case are introducing two types of redundancy in the source signal: statistical and deterministic redundancy, respectively. The first method consists in creating a correlated block of n variables out of an i.i.d block of Gaussian variables, and the second one uses redundant transforms such as frame decompositions.

Let us explain these methods into more detail.

1.3.1 Statistical correlation

This technique, known as of Multiple Description Correlating Transform (MDCT) has been introduced in the literature by Wang, Orchard and Reibman in [WOR97] and [WOR98] for two variables. The results have been generalized later on by Goyal and Kovacevic to the case of n variables.

Here, the descriptions are obtained from centered Gaussian variables X_i that are supposed to be independent and have variances σ_i^2 respectively. The method consists in linearly transforming these variables into correlated ones, Y_i, that shall be sent over independent channels with rates R_i. As mentioned above, in [WOR97] the two-description case is handled. Thus, two correlated variables Y_1 and Y_2 are obtained from two independent ones X_1 and X_2.

From Shannon's rate-distortion theory we know that the rate-distortion function of a centered Gaussian variable having σ^2 variance is:

$$R(D) = \frac{1}{2}\log_2 \frac{\sigma^2}{D}. \tag{1.7}$$

On the other hand, it has been proven that the optimal bit rate allocation for a pair of centered Gaussian variables at a given global rate R is:

$$R_1 = R + \frac{1}{2}\log_2 \frac{\sigma_1^2}{\sigma_1\sigma_2} \tag{1.8}$$

$$R_2 = R + \frac{1}{2}\log_2 \frac{\sigma_2^2}{\sigma_1\sigma_2}. \tag{1.9}$$

If we consider the balanced case, in which $D_i = \frac{1}{2}D$ for $i \in \{1, 2\}$, and the global rate-distortion function of the pair (X_1, X_2) as being the sum of individual rate-distortion functions, we deduce that:

$$R(D) = \frac{1}{2}[\log_2 \frac{\sigma_1^2}{D_1} + \log_2 \frac{\sigma_2^2}{D_2}] = \frac{1}{2}\log_2 \frac{\sigma_1^2\sigma_2^2}{D^2/4} = \log_2 \frac{\sigma_1\sigma_2}{D/2}. \tag{1.10}$$

We are then looking for the random vector $Y = [Y_1, Y_2]^t = TX$ obtained with an orthogonal transform, T, so that Y leads to the same distortion as X. If we denote the variances of the new variables $Y_i, i \in \{1, 2\}$ by $\sigma_{y_i}^2$ respectively, and by ϕ the angle defined through the expectation of the variable pair $E\{Y_1Y_2\} = \sigma_{y_1}\sigma_{y_2}\cos\phi$, then we have the following relationship between the variances of the initial and the new variable pairs:

$$\sigma_1^2\sigma_2^2 = \sigma_{y_1}^2\sigma_{y_2}^2\sin^2\phi \tag{1.11}$$

For correlated variables $Y_i, i \in \{1, 2\}$ the angle $\phi \neq \frac{\pi}{2}$ and therefore $\sigma_{y_1}\sigma_{y_2} \geq \sigma_1\sigma_{y_2}$. Thus, at equal distortion, a higher rate would be necessary to transmit the correlated signals.

Goyal and Kovacevic introduced in [GK98] the notion of Multiple Description Transform Coding (MDTC), which is an extension of the Wang, Orchard and Reibman's work presented above. In their method the source vector, x, is first quantized with a uniform scalar quantizer of step Δ, leading to $x_{q_i} = [x_i]_\Delta$, where $[\cdot]_\Delta$ denotes the rounding to the nearest multiple of Δ. The so-obtained vector $x_q = [x_{q_1}, x_{q_2}, \ldots, x_{q_n}]^{\mathrm{T}}$ is then transformed with a discrete invertible transform denoted by $\widehat{T} : \Delta\mathbb{Z}^n \to \Delta\mathbb{Z}^n$, giving the vector $y = \widehat{T}(x_q)$. These coefficients are independently entropy coded after being grouped into $m \leq n$ subsets to be sent over the m channels.

The question is how to obtain $\widehat{T}$ from a linear transform T, and to this the authors propose a factoring of a given T with determinant one into several lifting steps (see [DS98]), or in other words into a product of upper and lower triangular matrices with unit diagonals: $T = T_1T_2 \ldots T_k$. The transform $\widehat{T}(x_q)$ is then taken as the discrete version of this transform. The major advantage of this kind of construction is the immediate invertibility of the transform, obtained by simply inverting the calculations.

It is important to note that the results presented above are valid under a couple of assumptions mostly related to fine quantization. These hypotheses are that the scalar entropy is preserved between the quantized vector y (obtained with $\widehat{T}$) and $[Tx]_\Delta$, that the quantization does not affect the correlation present in y and also that, in the event of component losses in y, the quantization can be ignored with respect to the error caused by erasures.

A gradient-based algorithm for finding optimal correlating transforms used on erasure channels is presented in [RRP05].

1.3.2 Frame expansions

A second method that generates correlation into the transmitted signal was also proposed by Goyal et al. ([Goy98], [KDG02], [GKV99]). In this case the signal is expanded by the means of a frame decomposition $y = Fx$, where F is a frame operator. At this point a recall on frames is useful since the current thesis is built on this kind of decompositions.

Background on frames

The beginnings of these particular transforms are commonly traced back to 1952 when Duffin and Schaeffer [DS52] introduced frames for Hilbert spaces. An article reviewing most of the important theoretical results regarding different types of frames is given by Casazza [Cas00]. Also, the recent tutorial given in [KC07a],[KC07b] presents a structured survey on frames in the literature, mentioning the terminologies involving frames and also the different areas of application.

Let us now review the basic concepts and properties for frame decompositions.

We shall operate in an infinite dimensional Hilbert space denoted by $\mathcal{H}$ which is endowed with the inner product $< ., . >$ and the norm $\|.\|$. Given a family of vectors $\phi = \{\phi_k\}_{k\in\mathbb{Z}}$, and a linear operator $F : \mathcal{H} \to \mathcal{H}$, defined by:

$$(Fx)_k = \langle x, \phi_k \rangle , \quad \forall k \in \mathbb{Z}, \tag{1.12}$$

the theory of frames aims at finding the conditions in which a vector x can be recovered in a stable manner from the projections $(\langle x, \phi_k \rangle)_{k\in\mathbb{Z}}$ or, in other words, the conditions to be imposed to the family ϕ in order to find a left inverse of the frame operator F.

We then have the following definition.

Definition 1.
*The family of functions $(\phi_k)_{k\in\mathbb{Z}}$, with $\phi_k \in l^2(\mathbb{Z})$ is a **frame** of $\mathcal{H}$ for any signal $x \in l^2(\mathbb{Z})$ if there exist two finite positive constants A and B such that:*

$$A\|x\|^2 \leq \sum_{k\in\mathbb{Z}} |\langle x, \phi_k \rangle|^2 \leq B\|x\|^2 \tag{1.13}$$

These constants are known under the name of **frame bounds** and their values determine different types of frames. This condition guarantees in particular the fact that the energy of the transformed signal remains bounded if the signal itself has a finite energy.

Note: In practice, we consider finite dimension signals belonging to $\mathbb{R}^N$ (or $\mathbb{C}^N$) and in this case the vector index, k, belongs to a countable index set $K \subset \mathbb{Z}$ and its cardinality is usually greater than N. The redundancy of the frame is thus given by the ratio $\frac{\text{Card}K}{N}$.

The particular case when $A = B$ leads to a so-called **tight frame** and if $A = B = 1$ then ϕ_k can be found under the name of **normalized tight frame**. The norm of ϕ_k can also be indicated in current types of frames, in particular the unit norm leads to **unit-norm** (found also as uniform) frames. The case $\|\phi_i\| = \|\phi_j\|, \forall(i, j)$ corresponds to **equal-norm** frames.

By using the associated linear frame operator F, Eq. (1.13) can be rewritten as follows:

$$A\mathbf{I} \leq F^*F \leq B\mathbf{I}, \tag{1.14}$$

where $\mathbf{I}$ is the identity matrix, and $F^* : \mathcal{H} \to \mathcal{H}$ is the adjoint operator of F satisfying:

$$\forall (x, y) \in H, \quad \langle Fx, y \rangle = \langle x, F^* y \rangle . \tag{1.15}$$

The operator formulation is useful because it leads to a computational way to find the frame bounds, since the eigenvalues of F^*F lie in $[A, B]$. From this, it immediately follows that F^*F is invertible, by having all eigenvalues to be non-zero, and we have:

$$\frac{1}{B}\mathbf{I} \leq (F^*F)^{-1} \leq \frac{1}{A}\mathbf{I}. \tag{1.16}$$

We can thus define the dual-frame as a family of vectors $\{\tilde{\Phi}_k\}_{k\in\mathbb{Z}}$ having bounds $\frac{1}{B}$ and $\frac{1}{A}$, respectively [GVT98].

Theorem 2. *Given the vector family $\{\tilde{\Phi}_k\}_{k\in\mathbb{Z}}$ defined by:*

$$\tilde{\Phi}_k = (F^*F)^{-1}\Phi_k \tag{1.17}$$

$\tilde{\Phi}$ corresponds to the dual frame and:

$$\forall x \in \mathcal{H}, \quad \frac{1}{B}\|x\|^2 \leq \sum_{k\in\mathbb{Z}} |\left\langle x, \tilde{\Phi}_k \right\rangle|^2 \leq \frac{1}{A}\|x\|^2. \tag{1.18}$$

Moreover the reconstruction of the signal x is given by:

$$\widehat{x} = \tilde{F}^* F x = \sum_{k\in\mathbb{Z}} \langle x, \Phi_k \rangle \, \tilde{\Phi}_k, \tag{1.19}$$

where

$$\tilde{F} = F(F^*F)^{-1}, \tag{1.20}$$

$\tilde{F}^*$ being thus the pseudoinverse of F.

The frame approach to multiple descriptions consists in projecting the source on a redundant frame and each of the resulting coefficient sets will constitute a description after quantization [GKV99]. A block diagram of these operations is given in Figure 1.7.

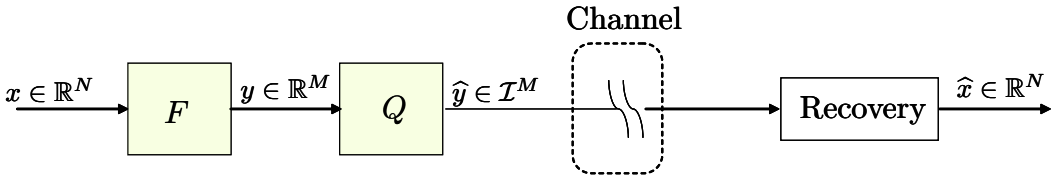


FIGURE 1.7: MD coding by frame decomposition. F is a frame operator and Q stands for a scalar quantization operator ($\mathcal{I}^M$ stands here for the quantization index set).

A note on the usefulness of frames in coding systems

The advantage brought by the use of frame decompositions is an increased robustness to additive noise on the signal, such as the quantization noise at small quantization steps or when using dithering [1] (see for instance [Lar95]). Since the linear operator $P = F\tilde{F}^*$ is the orthogonal projector on the subspace $\Im mF$, applying it to $\hat{y}$ (in Figure 1.7) leads to the elimination of the noise component in $\Im mF^\perp$ without altering the useful signal Fx. Moreover, the size of the subspace $\Im mF^\perp$ increases with the frame redundancy.

If we consider the quantized signal $\hat{y}$ as being

$$\hat{y} = Fx + w$$

where w is an additive zero-mean white noise with components of variance σ^2, and the decoder uses the linear estimator:

$$\widehat{x} = \tilde{F}^* \hat{y} \tag{1.21}$$

in order to recover the original signal, the mean square error, $D = \frac{1}{N} E\|x - \widehat{x}\|^2$ will satisfy:

$$D \leq \frac{1}{A}\sigma^2.$$

The equality is reached for tight frames. Also, the estimator used in Eq. (1.21) is least mean squares optimal but other operators can be used for the reconstruction as, for instance, Projections Onto Convex Sets (POCS), [TV94], [GVT98]. We have adopted a similar strategy for the application of MD coding to still images that we are presenting in detail in Chapter 3.

An important issue when applying a frame decomposition to the data source is the robustness to erasures (channel losses). In [GVT98] the error resilience for harmonic frames is proven.

One thing to note about the frame approach is its similarity with an error-correcting code. Indeed, in the frame approach one can consider that a group of N symbols is transformed into another group of M by adding a certain amount of redundancy. However this approach distinguishes itself from traditional channel coding by the position of the quantization step in the transmission chain. In the frame approach the quantization step occurs after the introduction of redundancy whereas in error-correcting codes the redundancy is added after quantization. In [GKV99] the authors perform a comparison with traditional codes in a simple numerical case and they conclude that the frame approach works better mostly at high rates and gives comparable performances otherwise. However, the redundancy introduced by the frame approach allows to significantly reduce the quantization noise.

Other frame-based methods use the windowed Fourier transform [BDV00].

1.3.3 Filter banks

Another case of MD methods with transform coding, which can be viewed as a particular case of a discrete-time frame decomposition [Cve95], [VK95], [CV98], [BHH98] is based on filter banks. Already very popular for applications like audio and image coding, the new developments linking them to wavelet transforms [Dau88], [Mal89] led to a regain of interest in their regard. The perfect reconstruction for filter bank decompositions has been thoroughly investigated in [Vet86], [Vai93a], [CV98]. We depict an example with two channels in Figure 1.8.

[1] Dither means intentionally adding decorrelated noise to the original signal in order to obtain a uniform distribution for the quantization noise.

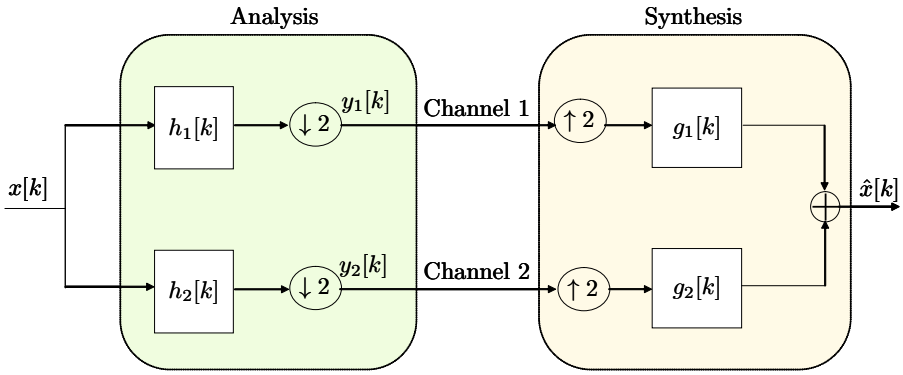


FIGURE 1.8: Two-channel filter bank.

The first application of filter banks to multiple descriptions is proposed by Yang and Ramchandran in [YR98]. Here, the analysis filters H_1 and H_2 are orthonormal and thus satisfy $|H_1(e^{j\omega})|^2 + |H_2(e^{j\omega})|^2 = 2$, $\forall\omega \in [0, 2\pi]$, [VK95, Chap. 3]. At the reconstruction, the associated synthesis filters g_1 and g_2 are used. The filtered signals obtained at the analysis stage are decimated by a factor of two, quantized and entropy coded for transmission over each of the channels. Moreover, in this method a Wiener estimator is used in order to compensate for the loss of one of the descriptions at the side decoders [YR98].

This approach presents in fact a more general case than that of the MD transform coding method proposed by Goyal et al. because no restrictions are imposed on the lengths of the analysis filters. The equivalent transformation from the filter bank is not limited to a 2×2 matrix in this situation. Moreover, the source is viewed as being Gaussian and wide sense stationary and thus not necessarily i.i.d as before.

Filter Optimization: Given the wide sense stationary Gaussian random process with Power Spectral Density (PSD) $S(\omega)$, the rate-distortion function is given by:

$$R(D) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{1}{2}\log_2\frac{S(\omega)}{D}d\omega \tag{1.22}$$

which is a generalization of Eq. (1.10), [Ber71]. By applying this to each of the two descriptions, $y_1[k]$ and $y_2[k]$, which are viewed as Gaussian variables, and by supposing that the theoretical bound is reached for the entropy, one can consider that the two transmission rates are:

$$R_1(D_0) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{1}{2}\log_2\frac{Y_1(\omega)}{D_0}d\omega \tag{1.23}$$

$$R_2(D_0) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{1}{2}\log_2\frac{Y_2(\omega)}{D_0}d\omega \tag{1.24}$$

where $Y_i(\omega)$, $i \in \{1, 2\}$ are the power spectral densities of the two variables and D_0 is the central distortion. In these conditions the redundancy is given by:

$$\rho(D_0) = \frac{1}{2}(R_1(D_0) + R_2(D_0)) - R(D_0). \tag{1.25}$$

The side decoders employ, as said above, a Wiener filter allowing the reconstruction of the lost description. For instance, if only $y_1[k]$ is correctly received, the Wiener filter given by the frequency response $Y_{12}(\omega)/Y_1(\omega)$ is used to estimate $y_2[k]$. Here $Y_{12}(\omega)$ stands for the cross

spectral density[2] of the two random variables. Hence, the side distortions shall be given by the Wiener estimation error, under high rates hypothesis for R_1 and R_2 which leads to a negligible quantization error.

In order to find the optimal analysis filters, one must solve:

$$\min_{H_1(\omega),H_2(\omega)} \frac{1}{2}(D_1 + D_2) + \lambda\rho(D_0) \tag{1.26}$$

under a fixed central distortion (λ is the Lagrangian parameter handling the redundancy constraint).

Several interesting results have been formulated in [YR98] for this trade-off between side and central distortions. Thus,

- the optimal filters when $\lambda \to \infty$ form a filter bank equivalent to a Karhunen-Loeve transform, which is a well known result in source coding [Mor00]. The two descriptions are then completely decorrelated.
- Conversely, if $\lambda \to 0$ then the redundancy is not considered and one only minimizes the side distortions. In this situation it is found that the filter bank corresponds to a polyphase transform and thus the two descriptions are:

$$y_1[k] = x[2k] \tag{1.27}$$
$$y_2[k] = x[2k+1] \tag{1.28}$$

  The correlation between $y_1[k]$ and $y_2[k]$ is then maximal.
- When $x[2k]$ and $x[2k+1]$ are each i.i.d. Gaussian, the optimal analysis filters are FIR filters of length 2, and the proposed filter bank is equivalent to the MDTC method described earlier. This result is also given in [PR00].
- Optimal filters are also given for an AR(1) source model and the rate-distortion results are compared with those obtained with classical orthonormal filters associated to wavelet decompositions, such as Haar, Daubechies, Coifman [Mal98].

The same problem of designing optimal filter banks for MDC has also been approached by Dragotti et al. in [DSV02b]. The difference between the two approaches resides in the place of the quantizer in the transmission chain. Dragotti et al. are building MD systems according to the correlating transforms scenario proposed by Orchard et al. [OWVR97] and Goyal et al. [Goy00] and thus the quantization is performed before the filter bank decomposition. The advantage of this approach is that the quantization cells are not changing shape and the quantization error is not increased by the use of non-orthogonal transforms. In a first stage of this work the authors generalize the information theoretic results given by El Gamal, Cover and Ozarow for the rate-distortion region of stationary Gaussian sources with memory. In the second stage they design, as Yang and Ramchandran, the optimal two-channel filter bank and prove that the optimal redundancy allocation between descriptions is obtained with the reverse "water-filling" strategy.

[2]The cross spectral density of two wide-sense stationary random variables $X[n]$ and $Y[n]$ is given by the Fourier transform of the cross-correlation function $R_{XY}(\tau) = \mathrm{E}X[n]Y[n+\tau]$.

We have mentioned earlier that, for some choice of the Lagrangian parameter, λ, the filters involved in the filter bank are equivalent to a polyphase decomposition. This idea has been exploited in [JO99] where each description contains also a part from the other description. Recall that the M polyphase components of a source $x[k]$ are given by the sequences $(y_i[k])$, $k \in \mathbb{Z}$, $i \in \{1, \ldots, M\}$, where:

$$\forall k \in \mathbb{Z}, \quad y_i[k] = x[Mk + i], \tag{1.29}$$

with $k \in \mathbb{Z}$. The so-obtained components are encoded independently - an idea already used in robust audio coding since the early eighties [Jay81]. In the multiple description scenario, redundancy is added to each description through a coarsely coded version of the other one. An advantage of this strategy is that the two stages: description generation and redundancy allocation are separated - contrary to the MDSQ or the MDTC approaches presented before. In this way, the encoder can have a lower complexity, according to Jiang and Ortega. Their MDC approach is therefore preserving, in some sense, a separation between source and channel coding.

The issue to handle at this point is the optimal rate allocation between the two descriptions (considered in the balanced case). For a centered Gaussian i.i.d. source with variance σ^2 it is found,[JO99], that the rate distortion function for each variable corresponding to a description is, as before (for $i \in \{1, 2\}$):

$$D(R_0) = h\sigma^2 2^{-2R_0}, \tag{1.30}$$

where h is a constant depending on the used quantizer and it is equal to $\sqrt{3}\pi/2$ for a scalar Lloyd-Max quantizer, and to 1 for an optimal quantizer[3]. R_0 is, as above, the rate needed to encode the original source. The rate in each description is given by the sum of R_0 and the introduced redundancy, r, and the rate allocation is obtained by Lagrangian minimization of the central distortion D_0 under a side distortion constraint of type $D_i \leq d_{i_{max}}$ and for a fixed global rate R. In this manner the following expression is found for the redundancy:

$$r = \frac{1}{2}R + \frac{1}{4}\log_2\left(\frac{\lambda}{2+\lambda}\right), \tag{1.31}$$

λ being the Lagrangian parameter tuning the trade-off between central and side distortions. The appropriate value of λ depends on the channel failure probability, in order to know if one has to favor the side or the central decoder. The optimal redundancy according to this probability, p, is given by:

$$r = \begin{cases} \frac{1}{2}R + \frac{1}{4}\log_2(p), & \text{if } p > 2^{-2R} \\ 0, & \text{if } p \leq 2^{-2R}. \end{cases} \tag{1.32}$$

Other methods are designing MD with filter banks which optimize the synthesis filters for a given analysis set in order to achieve minimal average distortion [SM02]. The synthesis filters are no longer necessarily FIR.

1.4 Channel oriented methods

The previous sections gave an overview of the core methods developed for MDC and this, mostly from a source coding point of view. These methods mainly concentrate on the way

[3]The optimal quantizer from an information theory point of view leads to the asymptotic rate-distortion bound reached by an entropic vector quantizer on infinite length blocks.

of building the descriptions without making any assumptions on the transmission network. The design efforts thus focused on the equivalence in reconstruction quality of each individual description (in other words, on the least favorable transmission case which is that of unprioritized networks).

In this section we are recalling a different approach from the literature, which focuses on the possibility of using Unequal Error Protection (UEP) on the transmission channels. To this end researchers proposed to transform a scalable source bitstream into an M-description packet stream in which each packet contains approximately the same amount of information. A strategy for prioritized encoding mainly designed for video conferencing-type applications over lossy packet networks is given in [ABE+96] and serves as a starting point in this new class of MD methods. The scalable representation can be provided by hierarchical progressive coders such as SPIHT [SP96], EZBC [HW00] or EBCOT [Tau98] which were introduced for image compression.

Puri and Ramchandran, [PR99], combine these considerations with Forward Error Correction in order to add redundancy to a given source. They propose to split the information bitstream into several layers in decreasing order of importance and each of those layers is further protected by progressively weaker channel codes (see also [MRL99], [CWP03] for similar MD methods). An optimization algorithm allowing for the maximization of the expected reconstruction quality is proposed in order to decide how much protection needs to be assigned to each layer, given the channel state and the transmission rate constraints.

More precisely, a hierarchical bitstream is first split into M layers of increasing resolution (as presented in the first part of Figure 1.9) for an M-description scheme. Abiding the MDC working philosophy amounts to the condition that the i^{th} layer must be decodable from any subset of i received descriptions. In order to achieve this, the i^{th} layer is further split into i equal parts which are placed in the first i packets. The other packets are filled by channel codes (Reed-Solomon, for instance) of parameters $(M,\, i,\, M-i+1)$[4] applied vertically on each layer.

This mechanism of bitstream partitionning for M descriptions is presented in Figure 1.9.

As we said earlier, an important issue for this MD encoder is the optimal partitioning of the bitstream into layers. A rate-distortion problem needs to be solved for $D(R)$ - the rate-distortion function of the hierarchical bitstream and $R = (R_1,\, R_2,\, \ldots,\, R_M)$ - the vector of rate points in Figure 1.9. If p_i is the probability that i of the M packets reach the decoder and σ^2 is the source variance, then the mean distortion at the decoder is:

$$\mathrm{E}\{D\} = p_0\sigma^2 + \sum_{i=1}^{M} p_i D(R_i) \tag{1.33}$$

and the global bitrate is obtained as:

$$\begin{aligned} R &= R_1 M + \frac{R_2 - R_1}{2} M + \ldots + \frac{R_M - R_{M-1}}{M} M \\ &= \sum_{i=1}^{M} \alpha_i R_i \end{aligned} \tag{1.34}$$

with the notation $\alpha_i = M/(i(i+1))$ for $i \leq M-1$ and $\alpha_M = 1$ otherwise.

[4]Recall that an $(n,\, k,\, d)$ forward error correction code transforms k source bits into n bits obtained by adding $n-k$ redundancy bits. This allows for the correction of maximum $d-1$ erroneous bits, with $d \leq n-k+1$.

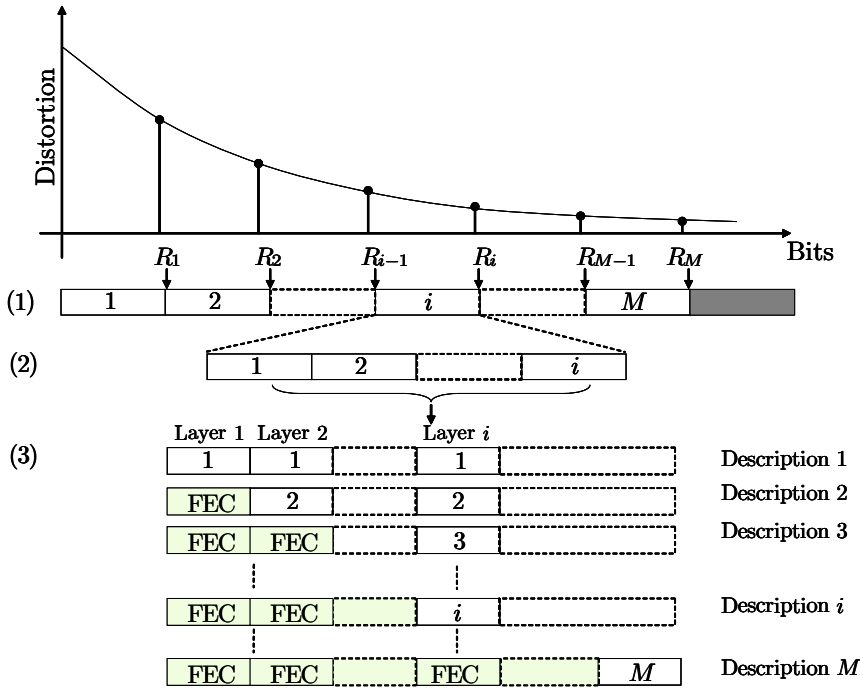


FIGURE 1.9: Bitstream partitioning for unequal error protection (UEP): (1) the bitstream is divided into M successive layers (quality levels) each level i having the bitrate R_i, (2) The i^{th} level is again divided into i equal parts, (3) a block code of redundancy $(M,\ i)$ is applied to the i^{th} layer and the latter is then split into M packets.

For M packets (descriptions) of length L bits, the global needed rate, denoted by R^* can be expressed as $R^* = LM$ and finding the optimal truncation points of the bitstream translates to the minimization of (1.33) under the constraints:

$$\sum_{i=1}^{M} \alpha_i R_i \leq R^* \tag{1.35}$$

$$R_1 \leq R_2 \leq \ldots \leq R_M. \tag{1.36}$$

This rate-distortion optimization is solved using the classical method involving the Lagrangian operator.

A general framework for a variety of transmission scenarios in the packetized media streaming context is proposed in [CM06]. In [KA05] a method of channel balancing for low-delay wireless networks is developed based on a set partitioning algorithm.

The combination of MD with layered coding is also reported for correlating transforms MDC in [WRTJ02]. Other works explore the adequate scenarios for using multiple description (source diversity) coding as opposed to the single description case [CKS01], [LMWA05].

Another interesting direction that has been slightly explored in the literature involves a mixture of turbo codes and multiple description coding [BHG02], [KFR04].

1.5 MDC in the world of multimedia applications

So far we have mainly seen general strategies for building multiple description codes which do not concern a particular application, but instead give an idea of the attainable performances of different methods as well as directions of further improvement. In this section we shall orient our attention on the several practical applications that have appeared in the literature in the ubiquitous multimedia projects. This survey is centred on applications that involve still images and video sequences, since those represent the main interests for this thesis.

1.5.1 Image coding

The first applications of multiple description coding to images are investigated by Wang et al. [WOR97], Goyal et al. [GKAV98], Jiang and Ortega [JO99], Serveto et al. [SRVN00], and they are closely related to the general methods we have enumerated in the previous sections.

The pairwise correlating transform is applied to different blocks of an image, [WOR97], after classifying these blocks into four classes in order to ensure similar statistical properties of the transformed coefficients. The selection is made upon geometrical/image regularity considerations such as smoothness, edge orientation for the first three classes, whereas the fourth is assigned to what is left after this classification. This is due to the fact that real images are not statistically stationary, therefore the correlating transform applied globally could introduce large estimation errors.

The transform coding approach to images of Goyal et al. [GKAV98] uses the generalized multiple description method proposed before and applies it to a four-channel coding scenario similar to JPEG. DCT coefficients are obtained in 8×8 blocks and, after being uniformly quantized, they are split into vectors of length four, based on frequency and space separation. Then the correlating transform is applied on these four-tuples and descriptions are formed by repeating the DC coefficient and splitting the remaining AC coefficients in an approximately balanced manner. A second method building descriptions with deterministic redundancy instead of statistical one is equally proposed. This is based on overcomplete frame expansions designed to match an (n, k) error correcting code, with $n > k$. As in the previous application, a JPEG coding context is considered, and thus descriptions based on 8×8 blocks of DCT coefficients are built. An example using a frame operator of size 10×8 is given as an alternative of an $(10, 8)$ error correcting code. This MD system has thus 10 descriptions and the obtained numerical results are compared to a base line system using DCT coefficients and a systematic $(10, 8)$ block code. It is shown that the MD system outperforms the baseline if more than three descriptions are lost, as it was expected, but it does not outperform the baseline system at the central decoder.

Another technique for MD image transmission was proposed by Jiang and Ortega [JO99] and it uses the polyphase transform for description generation, followed by selective quantization in order to introduce the desired amount of redundancy. An optimal bitrate allocation algorithm for i.i.d scalar and vector sources and independent channel failure probabilities is equally proposed with this method. For the two-description case such a system can be viewed as in Figure 1.10. Two quantizers are involved in each description: a fine quantizer Q_1 and a coarse quantizer Q_2 that serves as redundancy for side reconstruction. The input signal, x, can be here either the spatial or the frequency signal, and there are two proposed decompositions:

- The basic decomposition uses a wavelet transform in order to get a sequence of coeffi-

cients. These are then grouped following their odd index in the first polyphase component and their even index in the second one, and the operation is done in each subband.

- A zero-tree decomposition inspired by Shapiro, [Sha93] according to which the first polyphase component contains the blocks of coefficients corresponding to the same spatial position in each subband.

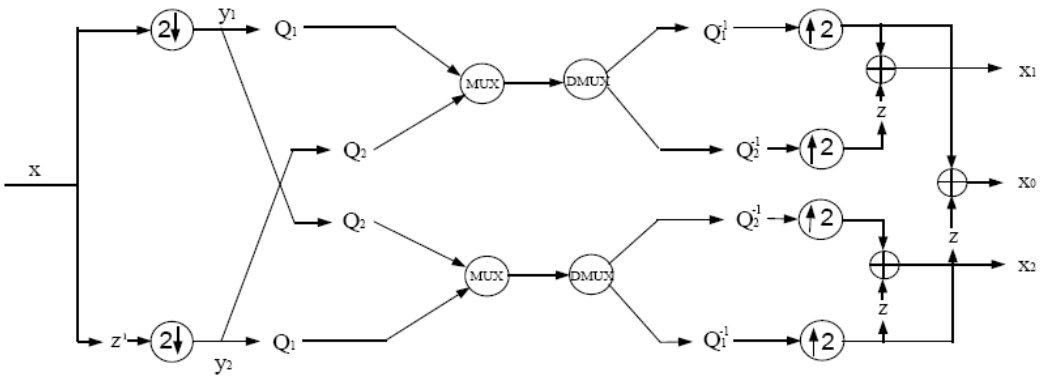


FIGURE 1.10: Polyphase encoding scheme (taken from [JO99]).

A SPIHT encoder [SP96] is used for these polyphase components. The interest of this algorithm is that it exploits the correlation between the subband coefficients and also allows for a fine tunning of the encoding rate. In the polyphase MD method the rate is equally divided between the two descriptions and the central rate is given by a base rate R_0 plus the added redundancy r. Thus, in the first description, the y_1 component will be denoted by $y_1(\frac{1}{2}R_0)$, which means that the encoding rate is $\frac{1}{2}R_0$ and the y_2 one will be encoded at r bits and denoted by $\frac{1}{2}y_2(r)$ and vice versa for the second description. An extension to M descriptions is also possible simply by making M groups of coefficients instead of 2 in the zero-tree structure, but this raises the question of how many secondary polyphase components are needed in each description.

A similar method to Jiang and Ortega's is introduced by Miguel et al. in [MMR99]. This equally uses the SPIHT coder and the descriptions are built by grouping wavelet coefficient trees. The redundancy is obtained by duplicating trees encoded at a lower rate. An algorithm for building packets is proposed here based on the Packetized Zerotree Wavelet (PZW) compression scheme [RC98] and a description might contain several de-interleaved wavelet trees. This method however requires the transmission of some side information containing the position of the first tree and the number of trees in each description which adds up to the already introduced redundancy. Comparisons are provided to Goyal et al. method and also to Jiang and Ortega's but in the first case it is not clear if the performance gain is due to a better encoder or to the MDC and, in the second case, this might be due to the deficiency in the packetization strategy of the latter.

Another recent method based on SPIHT is investigated by Sriraja et al. [SKK05] and it uses overcomplete wavelet decompositions. These decompositions are depicted in Figure 1.11, that we reproduce here since it is closely related (in terms of design starting point) to our proposed two-dimensional application of MDC explained in detail in Chapter 3. A quick overview on wavelets will be presented in the next chapter. For now, we shall only mention the fact that

$H_0(z)$ is the z-transform of a low-pass impulse response filter $h_0[n]$ and $H_1(z)$ represents the z-transform of a high-pass impulse response filter $h_1[n]$, allowing for the filter bank implementation of a discrete wavelet transform (DWT) - the Fast Wavelet Transform [Mal98, Chap. 7]. For images, a total of 4^n equivalent sets of discrete wavelet transformed coefficients is obtained after an n-level wavelet decomposition as in Figure 1.11, adapted for the separable two-dimensional case, and, therefore, as many descriptions can be envisaged.

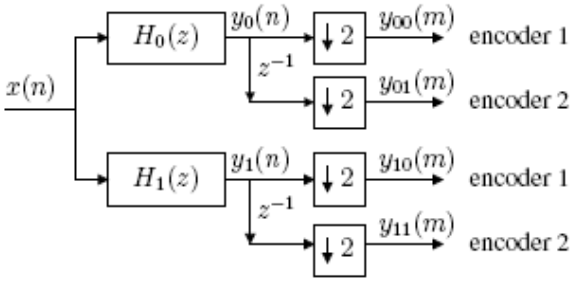


FIGURE 1.11: Overcomplete discrete wavelet transform for a one-dimensional signal (taken from [SKK05]).

The evolution of the PSNR[5] with the description number is studied, aiming to prove that the descriptions are balanced. It is experimentally proven that a subset of two descriptions out of the total possible number yields the best results in terms of performance versus coding gain trade-off. The scheme proposed here has however the disadvantage of being highly redundant and the central decoder cannot fully benefit from the introduced diversity, by that losing much in coding efficiency. The SPIHT algorithm mainly serves here at the packetization in each description as well as at the actual encoding of the coefficients, but not at description forming.

Other methods also use wavelet transforms when building multiple description schemes. Thus Servetto et al. [SRVN00] have introduced MD scalar quantization combined with wavelet transforms. The wavelet transform associated with rate-distortion optimization for bit allocation was equally used in [PAB02a], [PAB02b], [PAB03b], [PAB03a],[Per04], [CLJB05], [TO04], [TGO07], and more generally frame decompositions and oversampled filter banks are reported in [MG04a], [CMW99], [BR05]. Other applications are based on lapped orthogonal transforms in [CW99b], discrete cosine transform (JPEG) [RWG00], [Shi06], phase scrambling in the Fourier domain [SS04], optimal domain partitioning based on lattices [BW03].

Matching-pursuit like applications have also been proposed by Radulovic and Frossard in [RF07]. They are building multiple descriptions based on redundant dictionaries.

1.5.2 Video coding

In this section we are going to review the practical MD schemes which apply to video coding (MDVC). Building MD schemes for the transmission of video sequences offers more degrees of freedom as compared with images since the source has an additional dimension in this case, which is given by the temporal axis (the "frame" direction). However, new problems also arise for these particular signals, and important ones are motion handling, error propagation,

[5] $PSNR(x,\widehat{x}) = 10\log_{10}\frac{\max(x)^2}{MSE(x,\widehat{x})}$, with MSE denoting the mean square error

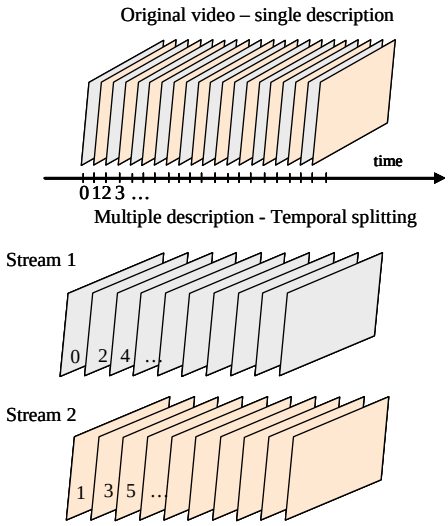


FIGURE 1.12: Example of temporal splitting of a video sequence.

decoding mismatch etc. Naturally, the first ideas that appeared in the field are extensions of the MD methods for still images with the necessary amendments required by video sequences. Thus, a great deal of techniques that have been reported perform the previously explained MD methods mostly intra-frame and we shall refer to these as *spatial MD* techniques. Some works have been done on the temporal part of the signal, but there is less abundance than in the spatial case, even though the approach shows promising results. We are referring to these methods as *temporal MD* ones and, in the next chapter, we shall present our contribution based on transform coding with redundant wavelets of the temporal signal in a $t + 2D$ video coding scenario.

Several directions have been investigated for MDVC. These include protection of the most significant DCT coefficients by MD [LPFA00], modifying the prediction loop [RJW+99] of a closed-loop video encoder, matching pursuit [TZ02], forward error correction [PLRB01], [Pur02], MD performed on motion vectors [KL01] and so on.

Before going into the details of these practical schemes, we present two very simple strategies that are more or less the founding stones to the further considered directions when it comes to dealing with video sequences. In Figure 1.12 is represented a technique known as *temporal splitting*. This involves, in some sense, reducing the frame-rate of the video source by a factor of two in each description. This idea is further investigated in [CGPPT07] for splitting also groups of frames and for triadic schemes, too, in addition to dyadic ones [TCPPG07].

A second simple technique for building two or more descriptions involves the partitioning of each individual frame in the video sequence, and this has come to be known as *spatial splitting*. Figure 1.13 gives an example with two descriptions formed in a balanced manner, as in the temporal case [FFLT05]. Obviously other simple spatial splitting strategies are possible (for instance quincunx splitting).

A very good survey on MD for hybrid video coders is provided by Wang, Reibman and Lin in [WRL05]. Video schemes are classified here according to the solution to drift effect they are proposing (or not) and the introduced redundancy. Thus the existing coders involving prediction loops belong to one of the three classes:

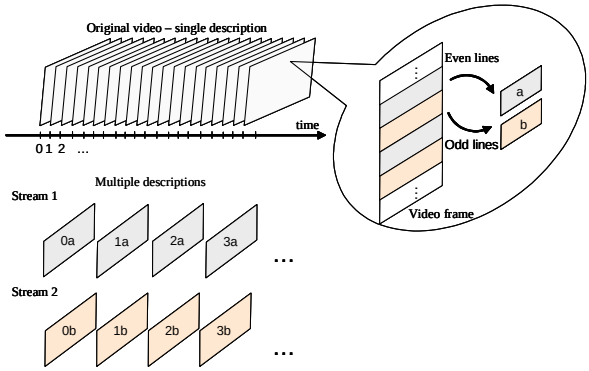


FIGURE 1.13: Example of spatial splitting of a video sequence.

- Class **A** - this includes the predictors without mismatch, where individual predictors are using the information sent in only one the two descriptions (for a two-description scheme);
- Class **B** - these are predictors that do not introduce additional redundancy in the scheme and are those used in the single description scenario. The inconvenience of these predictors is that both descriptions are needed, otherwise they introduce mismatch;
- Class **C** - in this category are given the adaptive schemes capable to trade-off prediction efficiency and the amount of mismatch.

One of the first applications to video coding is given by Reibman et al. in [RJW+99] and detailed in [RJW+02]. The authors propose video coders using motion-compensated predictions without mismatch, thus their coder belongs to the above mentioned Class A. Multiple description transform coding is used in order to generate two descriptions and the encoder is formed of three separate prediction loops, each of them corresponding to a certain type of decoder (two side decoders and a central one). The prediction error is also coded into two descriptions. The redundancy allocation takes into account the importance of the frames in the prediction, thus earlier frames get more redundancy. Three strategies are proposed in order to reduce or eliminate the decoding mismatch. It is worth noting that this work was one of the cornerstones in MD video schemes for prediction loop coders developed afterwards. Boulgouris et al. have also built a drift-free MD coder [BZKS06] in which each description contains a redundant part and a refinement part and the losslessly encoded motion vectors are repeated in each description.

Another application of MD to video coding is presented by Tang and Zakhor in [TZ02]. Here the structure given by the discrete cosine transform is modified in order to allow the use of matching pursuits and the MD system is built upon the three-loop structure proposed in [RJW+02]. Another contribution of this paper is the enhancement of the reconstruction at the central decoder based on Maximum Likelihood. The performances of the scheme are tested over lossy channels using a two-state Markov model and a Rayleigh fading model. Another technique using matching pursuits combined with MDSQ has been reported in [CFH05].

Gallant et al. have developed a standard compliant MDVC scheme based on spatial oversampling of the video signal by the means of an inverse zero-padded DCT, [GSK01] and a polyphase transform that generates two descriptions. Bernardini et al. propose a MD video

scheme with four descriptions [BDR+04] generated by a polyphase subsampling of the spatial components of the video signal. The missing frames at the decoder are recovered by applying different types of linear and non-linear interpolations and a post-processing step is performed in order to eliminate a visual artefact of granularity in the decoded sequence. In these works, H.264/AVC codecs are used. The polyphase transform applied on the spatial component of the video signal has equally been used by Franchi et al [FFLT05] for building a two description scheme based on the MDTC structure from [RJW+02]. Such a scheme is known as drift-compensation multiple description video coding (DC-MDVC) and, by offering the possibility of using the reconstructed reference frame in the side prediction loops (instead of the original reference frame), it provides robustness in error prone networks. A second scheme is also proposed in [FFLT05] under the name of Independent Flow MDVC (IF-MDVC) in order to facilitate the construction of more than two descriptions. An extension to DC-MDVC using Motion-Compensated 3D lifting [PPB01] is presented in [MS05].

In [JZY+04] a redundant wavelet scheme is given in order to encode a video stream into two descriptions, based on temporal odd/even splitting of the input signal. The missing samples in each of the two streams are predicted from the received ones and thus the mismatch is avoided in each description. Each description is encoded with a redundant discrete wavelet transform (RDWT) video codec. In the next chapter we present our application to video coding based on redundant wavelets in the context of $t + 2D$ codecs.

Starting from the idea that the trade-off between error resilience and compression efficiency of most existing MDC methods is dependent on the targeted quality, network capabilities, as well as the characteristics of the video itself, Heng et al. introduced an adaptive multiple description scheme [HAL06]. Different simple MD modes are defined and the system chooses between them based on a rate-distortion optimization. The authors consider the following four modes: single description (SD) coding, temporal splitting (TS), spatial splitting (SS) and repetition coding (RC). Among these modes the most efficient in terms of coding is obviously the SD mode, whereas the most efficient in terms of error resilience is the RC mode. Of course, these two modes represent extreme situations and they do not provide a satisfactory trade-off when applied on the entire sequence. They could prove to be very useful on certain portions, however. The other two modes are particularly efficient for low-motion videos, this being the case of the TS mode (since in this situation reducing the frame rate is less damaging than in high-motion videos) and, in a complementary manner, for high-motion videos. Therefore, if a contextual separation of the original video is properly done, these four modes cover all the important characteristics and good end-to-end performances might be expected. Choosing between the MD modes is done by Lagrangian optimization of a modified rate-distortion function, in which the distortion is split into a known part - which is due to quantization - and an unknown part - due to random packet losses. The distortion due to packet losses is estimated for the particular Gilbert loss model of the channel. The specific application that is developed in this paper uses the H.264 video coding standard and the adaptive mode selection is performed on macroblocks.

Kim et al. approached MDC applied to video streaming in a different manner, [KMA05]. They built unbalanced MD schemes based on the spatio-temporal tree-preserving 3D-SPIHT algorithm. Two independent bitstreams protected by Forward Error Correction (FEC) are generated according to allocated bitrates on each channel after bandwith estimation and the final rate allocation takes into account a Gilbert loss model for the network. In the same direction, Gan et al. [GGM06] proposed a technique that mixes MDC with FEC, ARQ and GOP-level interleaving in order to account for bursty network losses. Their scheme, however, presumes

the existence of a feedback channel which might be a limitation in low-delay applications.

Other methods of unbalanced multiple descriptions are based on optimal rate allocation under a global distortion constraint, as for example in [CSOM03], [FSF06]. In the latter work, the video sequence is viewed as a collection of AR(1) [6] signals. Scalable methods for rate-distortion optimization MD coders can be found in [Kon05] and for P2P networks in [TPL04]. Other existing work in the scalable context, developed for $t + 2D$ video codecs with temporal redundancy addresses 3-band filter banks [vdST03]. Scalable MDVC methods compatible with H.264/AVC are reported in [RSO06].

Lee et al. [LAM05] propose yet another method which uses error concealment by exploiting spatial or temporal smoothness in a video combined with an MDSQ encoding method.

Another approach which is not multiple description coding but follows similar guidelines is proposed by Apostolopoulos in [Apo99], and [Apo01], and it concerns video communication over unreliable networks. In this work multiple states are created at the encoder by temporal splitting. However, there is no explicit redundant coding of video frames, the higher rate resulting when putting together two states encoded individually coming from the fact that the frames are further apart and thus the motion compensation is less effective. The proposed encoding strategy is combined with path diversity which means explicitly sending different packets on different paths. This idea has some important benefits: the burst losses are transformed into individual losses, the outage probability decreases and smaller fluctuations in transmission quality are encountered by averaging the number of paths. However, one must take into account several issues when building such a communication system. Taking advantage of path diversity seems to be a good strategy but it raises the questions of how to choose among different packets and different paths, how many paths should one use, how to balance the load on each of the chosen paths and so forth. Apostolopoulos proposes some architectures enabling this, based on IP routing or relays [Apo01]. The multiple states with path diversity approach also benefits from an additional stage at the decoder which is called state recovery. Related approached applied to Ad-Hoc networks are found in [MLP+03]. Also, an MD technique exploiting multiple states and the redundant pictures option present in $H.264/AVC$ standard is reported in [RWW+07]. A somewhat similar idea based on redundant slices in $H.264/AVC$ is explored for video-over-IP applications in [VDMdW07].

1.6 Conclusion

In this chapter we have given a brief overview of the existing work in the multiple description coding field. We have first outlined the theoretical rate-distortion results developed in the Information Theory community in the late 70's and the 80's, results which determine the bounds for certain types of sources and also for the ideal case in which channels have an on/off functioning regime. However, these results are only stating that such joint source/channel strategies are a preferable solution to separate source and channel coding in the event of erasure channels. While answering the question of "why" one could use this kind of methods they did not answer the "how" question. A couple of decades later, with the explosion of multimedia transmission on packet-based networks, these attractive results have been brought back to light and practical coding methods have been given along with the expected rate-distortion performances.

[6] 1st order Autoregressive model

After briefly discussing the historical background we have presented the two main directions which have been followed in this domain: the quantization methods and the correlation methods. We have recalled the most important results as well as the further developments that have emerged by mixing these techniques with other coding methods in order to better suit the actual network problems or specific application requirements.

We have continued this survey by recalling some of the main applications of MDC in multimedia transmissions, mostly insisting on those which deal with still images and video sequences, since our work is conducted in these two directions.

Finally, it is easily seen that, once the interest for MDC methods was stirred up, the field became highly prolific. In spite of this, comparisons both among different MD techniques and between some of the most competitive ones and other coding solutions are far from being trivial. In the first case, this is mostly due to the fact that a unified benchmark cannot be easily established for different MDC methods, since many of them are problem oriented and answer different problems. In addition to that, the experiments conducted may differ significantly both in the working conditions and the underlying assumptions, this making objective evaluation very difficult. Also, network conditions are a decisive factor when choosing one coding technique over the other and, since MDC is mainly functioning on a trade-off basis, not all conditions are adequate for usage. There is, however, some amount of agreement in the literature, for the case of applications with very low delay requirements on long retransmission time networks or no feedback at all ones, to say that in this situation MDC is preferred to layered coding and FEC techniques.

Chapter 2

Temporal MDC schemes

In this chapter we present a strategy for building two descriptions relying on a temporal multiresolution analysis of finite energy signals, which is associated with a decomposition onto a wavelet frame. In order to set the theoretical framework for our proposed schemes we are first going to specify, in the next section, the notations employed in this chapter and, with slight variations, throughout the whole document. This is followed in Section 2.1.1 by some wavelet frame considerations, allowing to validate the proposed MDC approach. We are discussing the conditions for perfect reconstruction from this angle and also the possible schemes that can be conceived under reduced redundancy constraints. Then, we are presenting the oversampled filter bank (OFB) structures associated with the overcomplete representations described before. Through this approach, practical conditions of invertibility and scheme selection criteria are given.

This chapter is closed by an application of MDC to robust video coding schemes in a lossy transmission scenario. Some of the results in this chapter stem from a joint work with C. Tillier and the have been partially published in [PTPPP04], [PTPPP05], [TPPPP05], [TPPPP07].

2.1 Preliminaries

The MDC schemes proposed in this thesis belong to the class of methods based on transform coding, as described in Chapter 1. In this class, wavelet transforms are a very appealing tool because they offer inherent scalability as well as good compression capabilities [FPP07]. Also, in the multiple description context, one can conceive wavelet transforms that have inherent redundancy.

This enforces the idea that wavelets are a good candidate for multiple description coding, especially for applications related to still images and video sequence transmission. Before going into the design details for our coding schemes, let us give the main notations used in this chapter as well as some preliminary core concepts.

Let $L^2(\mathbb{R})$ denote the space of finite energy real-valued signals. The MDC schemes that we shall build are essentially based on a multiresolution analysis of $L^2(\mathbb{R})$[1] [Mal89],[Mal98], [Mey90]. This consists in projecting a signal onto bases of functions that yield successively coarser approximations of that signal. Mathematically, a multiresolution analysis of a signal is

[1] actually, we shall soon enough refer to signals in $l^2(\mathbb{Z})$, since in practice we handle only signal samples.

defined as a sequence of embedded closed vector subspaces $\{V_j\}_{j\in\mathbb{Z}}$ of $L^2(\mathbb{R})$, which satisfies:

$$\{0\} \subset \ldots \subset V_{j+1} \subset V_j \subset V_{j-1} \subset \ldots \subset L^2(\mathbb{R}), \tag{2.1}$$

$$\forall x \in L^2(\mathbb{R}) \quad \text{we have } x(t) \in V_j \Leftrightarrow x(\frac{t}{2}) \in V_{j+1}, \quad j \in \mathbb{Z}. \tag{2.2}$$

Then, the projection of a signal $x \in L^2(\mathbb{R})$ onto a subspace V_j represents the approximation of x at the resolution level j. Note that the approximation obtained at level j is obviously finer than that obtained at level $j+1$ since V_j is larger than V_{j+1} in the sense given by (2.1).

Also, in a multiresolution analysis, a so-called *scaling function* $\phi \in L^2(\mathbb{R})$ exists such that the countable family of translated functions $\{t \mapsto \phi(t-n)\}_{n\in\mathbb{Z}}$ is an orthonormal basis of V_0. Moreover, dilated and translated versions of $\phi(t)$:

$$\left\{\phi_{j,n}(t), \quad n \in \mathbb{Z}\right\} \quad \text{with } \phi_{j,n}(t) = 2^{-j/2}\phi(2^{-j}t - n)$$

form an orthonormal basis of V_j.

The difference between the approximations of the signal, obtained in two successive subspaces, V_j and V_{j+1}, represents the detail information lost from one resolution level to the next one, and this can be obtained from a subsequent detail space W_{j+1}, orthogonal to V_{j+1}:

$$V_j = V_{j+1} \oplus W_{j+1}.$$

As for the approximation space, it has been proven [Mal98] that an orthonormal basis for W_j can be obtained from translated and scaled versions of a so-called *mother wavelet*, $\psi \in L^2(\mathbb{R})$, and given by the family:

$$\left\{\psi_{j,n}, \quad n \in \mathbb{Z}\right\} \quad \text{with } \psi_{j,n}(t) = 2^{-j/2}\psi(2^{-j}t - n).$$

The approximation and detail coefficients of a signal $x \in L^2(\mathbb{R})$ are given by:

$$a_{j,n} = \langle x, \phi_{j,n}\rangle = \int_{-\infty}^{+\infty} x(t)\phi_{j,n}(t)dt = \int_{-\infty}^{+\infty} x(t)2^{-j/2}\phi(2^{-j}t - n)dt,$$

$$d_{j,n} = \langle x, \psi_{j,n}\rangle = \int_{-\infty}^{+\infty} x(t)\psi_{j,n}(t)dt = \int_{-\infty}^{+\infty} x(t)2^{-j/2}\psi(2^{-j}t - n)dt,$$

respectively.

The set of approximation coefficients at a given resolution is called *approximation subband* and the set of detail coefficients a *detail subband*.

Then, a signal $x \in L^2(\mathbb{R})$ can be represented as:

$$x(t) = \sum_{n\in\mathbb{Z}} a_{J,n}\phi_{J,n}(t) + \sum_{j\le J}\sum_{n\in\mathbb{Z}} d_{j,n}\psi_{j,n}(t). \tag{2.3}$$

Let us denote the basis formed by these families of functions as:

$$\mathcal{B}^{\mathrm{I}} = \{\phi_{J,n},\ n \in \mathbb{Z}\} \cup \bigcup_{j\le J}\{\psi_{j,n},\ n \in \mathbb{Z}\}.$$

The superscript I basically serves in distinguishing between different bases, as it will be best apparent a little further. In this thesis we consider that our wavelet bases are either orthogonal or biorthogonal.

Such a wavelet decomposition is well-known to generate a non-redundant[2] and sparse representation of the signal of interest and this has been exploited in many compression schemes.

With these ideas in mind, we propose an MDC strategy that uses a wavelet frame decomposition obtained with an union of bases built by different translations of $\mathcal{B}^{\mathrm{I}}$. Typically, we shall use a union of two wavelet representations, denoted $\mathcal{B}^{\mathrm{I}}$ and $\mathcal{B}^{\mathrm{II}}$, where the latter is a translated version of the former, as it will be seen in the next section. This is followed by a secondary subsampling operation and the by distribution of the remaining coefficients into two descriptions. The additional subsampling aims at setting the redundancy to the size of an approximation subband, in terms of number of coefficients. Thus, the higher the wavelet decomposition level, the lower the global redundancy of the signal will be.

Before going into the details for these schemes, let us enumerate the main notations employed in this chapter and, with slight variations, throughout the thesis:

- x - *usually represents the input (source) signal*
- a - *denotes the approximation coefficients resulted from a given wavelet decomposition,*
- d - *stands for a corresponding detail subband,*
- j - *is a resolution level associated with the wavelet decomposition,*
- J - *corresponds to the coarsest resolution level in a wavelet decomposition,*
- n - *designates the temporal index of a 1D-signal, in the discrete case,*
- *The superscript symbols* I *and* II *- distinguish between the coefficients in the first and the second wavelet representation, respectively. More information about this second wavelet representation will follow.*

As an example, $d^{\mathrm{I}}_{j,n}$ stands for the detail coefficient obtained from the first basis $\mathcal{B}^{\mathrm{I}}$ at resolution level j and temporal index n.

The additional subsampling operation mentioned above and performed in the detail subbands will be indicated by the characters $\hat{}$ *and* $\check{}$ *. The detail subbands in the first wavelet representation* $\mathcal{B}^{\mathrm{I}}$*, at the coarsest resolution level,* J*, are then written as:*

$$\hat{d}^{\mathrm{I}}_{J,n} = d^{\mathrm{I}}_{J,2n}, \tag{2.4}$$

$$\check{d}^{\mathrm{I}}_{J,n} = d^{\mathrm{I}}_{J,2n-1}. \tag{2.5}$$

We have thus highlighted the even-index ($\hat{d}^{\mathrm{I}}$) and the odd-index ($\check{d}^{\mathrm{I}}$) coefficients, respectively. Similar notations are employed for the coefficients resulting from a second wavelet basis, denoted $\mathcal{B}^{\mathrm{II}}$ and defined in the next section.

Also, in the next chapter we shall present an application of multiple description coding to two-dimensional signals (still images) and the proposed notations will undergo some slight variations that will be best apparent at that moment.

[2] also called *critically sampled* representation, since the number of samples after the wavelet transform is equal to the number of samples in the original signal.

2.1.1 Signal Analysis

Let us now explain the wavelet frame-based signal analysis employed in this thesis. The use of this kind of structures is motivated by the fact that the reconstruction error is confined into an interval depending on the frame bounds, or in other words the quantization operation does not amplify the reconstruction error [MC99], [KDG02], [PTPPP05], [Fow05] as was mentioned in Chapter 1.

A frame decomposition improving the shift invariance property is obtained, for instance, from the union of the previously defined wavelet basis, $\mathcal{B}^{\mathrm{I}}$, with one shifted version of it, given by:

$$\left\{\psi'_{j,n} = \frac{1}{2^{j/2}}\psi(\frac{t}{2^j} + \frac{1}{2} - n), \quad n \in \mathbb{Z}\right\}$$

and the corresponding scaling function which is given in this case by $\phi'_{j,n} = 2^{-j/2}\phi(2^{-j}t + 1/2 - n)$, $\forall n \in \mathbb{Z}$, and forms the family $\{\phi'_{j,n}\}$.

This second wavelet basis for a representation at a resolution $J \in \mathbb{Z}$ is thus given by:

$$\mathcal{B}^{\mathrm{II}} = \{\psi'_{J,n}, n \in \mathbb{Z}\} \cup \bigcup_{j \leq J} \{\phi'_{j,n}, n \in \mathbb{Z}\}.$$

Let us now express the approximation and the wavelet coefficients obtained with these bases, in the one-dimensional case and at each resolution level $j \in \mathbb{Z}$, for a signal $x \in L^2(\mathbb{R})$:

$$a^{\mathrm{I}}_{j,n} = \int_{-\infty}^{\infty} x(t)\ \frac{1}{2^{j/2}}\phi(\frac{t}{2^j} - n)\ dt, \tag{2.6}$$

$$d^{\mathrm{I}}_{j,n} = \int_{-\infty}^{\infty} x(t)\ \frac{1}{2^{j/2}}\psi(\frac{t}{2^j} - n)\ dt, \tag{2.7}$$

and, with the second basis

$$a^{\mathrm{II}}_{j,n} = \int_{-\infty}^{\infty} x(t)\ \frac{1}{2^{j/2}}\phi(\frac{t}{2^j} + \frac{1}{2} - n)\ dt, \tag{2.8}$$

$$d^{\mathrm{II}}_{j,n} = \int_{-\infty}^{\infty} x(t)\ \frac{1}{2^{j/2}}\psi(\frac{t}{2^j} + \frac{1}{2} - n)\ dt. \tag{2.9}$$

Note that if we choose a finite set of resolution levels $\{1, \ldots, J\}$, then the global set of wavelet coefficients, $\{a^{\mathrm{I}}_{J,n}, d^{\mathrm{I}}_{j,n}, a^{\mathrm{II}}_{J,n}, d^{\mathrm{II}}_{j,n}\}$, for all $j \in \{1, \ldots, J\}$, amounts to a redundant scheme in which the number of coefficients as compared to a critically-sampled decomposition is doubled. In this work we aim at finding smaller sets of coefficients that preserve the frame property and yet reduce the overall redundancy. Further in this chapter we are going to discuss practical conditions for the proposed decompositions, and we shall see that these are enabled by the link between filter banks and wavelets [Mal98]. We shall thus see that the coefficients in Eqs. (2.6)-(2.9) can be obtained by cascading operations of filtering and decimation by a factor of 2. However, the whole redundant structure is equivalent to a filter bank decomposition in which the decimation step has been removed [HP98].

Note also that if we associate the decomposition on each family of functions $\mathcal{B}^{\mathrm{I}}$ and $\mathcal{B}^{\mathrm{II}}$ to a description in an MD scheme[3], then the original signal is perfectly and independently

[3]This is only a particular scenario. In our work we are building more general scenarios and the superscript indices I and II will usually **not** stand for a given description.

recoverable from each stream in case of failure of the other channel and no other losses. At the central decoder, however, the coding cost might be too high as compared to the gain in quality that the second wavelet representation might offer.

Our first contribution consists in building a reduced redundancy wavelet coefficient set, by eliminating the decimation only at the last decomposition level (coarsest resolution), and thus finding a better rate-distortion trade-off for a frame-based MD scheme. This will amount to a redundancy of the size of an approximation subband, in terms of number of coefficients, as mentioned before. In this situation though, the coefficients yielded by the two wavelet representations are going to be mixed in the two descriptions. Therefore the issue of perfect reconstruction at the central decoder in the absence of channel losses needs to be investigated. We shall address this problem in a later section.

2.1.2 Decomposition Schemes

A two-description scheme based on the above considerations will thus be built in the following way:

1. Compute $a^{\mathrm{I}}_{j,n}$ and $d^{\mathrm{I}}_{j,n}$ for $j \in \{1, \ldots, J-1\}$, as in Eqs. (2.6) and (2.7),
2. Compute $a^{\mathrm{I}}_{J,n}$, $d^{\mathrm{I}}_{J,n}$, $a^{\mathrm{II}}_{J,n}$ and $d^{\mathrm{II}}_{J,n}$ (at the coarsest resolution level, J),
3. Subsample all the detail subbands by a factor of 2 and then discard half of the resulting coefficients,
4. Form two descriptions with the remaining coefficients.

A schematic representation of our MDC strategy is depicted in Figure 2.1 for a decomposition over one level. The last stage in this figure gives a possible combination of subbands, obtained as explained above, into a two-description scheme. In the next section we are elaborating the possible structures formed in this manner and discuss their properties.

Before going further though, it is important to stress out again the fact that the upper-indices I and II indicate the family of functions used in the wavelet decomposition, meaning $\mathcal{B}^{\mathrm{I}}$ or $\mathcal{B}^{\mathrm{II}}$, respectively, and not the description to which the coefficients belong. The descriptions will be usually formed with "mixed" coefficients, that is, sequences obtained from both wavelet bases.

In the following we present four schemes with two descriptions based on the above strategy, the differences between all of them occurring in the second description at the last level of decomposition, J. In other words, at higher resolution levels, the detail coefficients obtained from $\mathcal{B}^{\mathrm{I}}$ will be split into two sets and each of them will contribute to one description.

Using the wavelet bases introduced above, we construct the following schemes:

1. **R (Repeat)-scheme**: This scheme consists in simply repeating in each description the approximation coefficients from a single wavelet basis and splitting the detail coefficients in half in each description. This corresponds to the decomposition onto the set of functions:

$$\bigcup_{j=1}^{J} \left\{ \frac{1}{2^{j/2}} \psi(\frac{t}{2^j} - n), n \in \mathbb{Z} \right\} \cup \left\{ \frac{1}{2^{J/2}} \phi(\frac{t}{2^J} - n), n \in \mathbb{Z} \right\}. \tag{2.10}$$

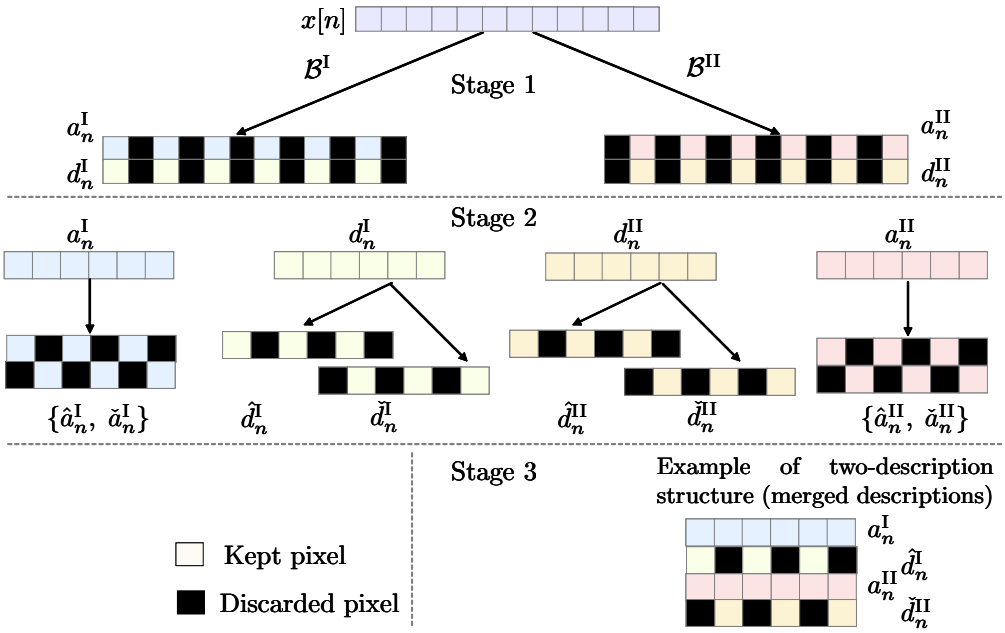


FIGURE 2.1: Building of reduced redundancy frame-based MDC schemes. In a first stage, the wavelet decompositions based on $\mathcal{B}^{\mathrm{I}}$ and $\mathcal{B}^{\mathrm{II}}$ are built. Then an additional subsampling is performed on the resulting subbands and a smaller set of coefficients is selected.

where, if we outline the fact that the detail coefficients are split according to their index parity, we can rewrite:

$$\bigcup_{j=1}^{J}\left\{\frac{1}{2^{j/2}}\psi(\frac{t}{2^j}-n), n\in\mathbb{Z}\right\} =$$

$$= \bigcup_{j=1}^{J}\left\{\frac{1}{2^{j/2}}\psi(\frac{t}{2^j}-2n), n\in\mathbb{Z}\right\} \cup \bigcup_{j=1}^{J}\left\{\frac{1}{2^{j/2}}\psi(\frac{t}{2^j}-2n+1), n\in\mathbb{Z}\right\}.$$

Then, using Eqs. (2.6) - (2.9) the two descriptions are built with the following sets of wavelet coefficients:

- ***description 1*** *contains the sequences:* $(a^{\mathrm{I}}_{J,n})_n$ *and* $(\hat{d}^{\mathrm{I}}_{j,n})_n$*, with* $j\in\{1,\dots,J\}$*, where,* $\hat{d}^{\mathrm{I}}_{j,n}$ *is given by Eq. (2.4),*
- ***description 2*** *is constructed with the same approximation coefficients* $(a^{\mathrm{I}}_{J,n})_n$ *and the remaining detail coefficients:* $(\check{d}^{\mathrm{I}}_{j,n})_n$*, with* $j\in\{1,\dots,J\}$*, where* $\check{d}^{\mathrm{I}}_{j,n} = d^{\mathrm{I}}_{j,2n-1}$.

This scheme is only provided for completeness of the study of possible combinations of subbands leading to the desired amount of redundancy. In what follows, we will be mainly interested in decompositions introducing some diversity between the approximations.

2. **MD1-Scheme**: In this second scheme the considered set of functions is:

$$\bigcup_{j=1}^{J}\left\{\frac{1}{2^{j/2}}\psi(\frac{t}{2^j}-n), n\in\mathbb{Z}\right\} \cup \bigcup_{p\in\{0,1\}}\left\{\frac{1}{2^{J/2}}\phi(\frac{t}{2^J}+\frac{p}{2}-n), n\in\mathbb{Z}\right\}. \tag{2.11}$$

This corresponds to a decomposition onto two wavelet bases, from which we have kept both the approximation subbands and eliminated all the detail subbands from the second decomposition. The detail subbands from the first decomposition are split as before. In other words, this can be expressed as follows:

- *description 1 contains:* $(a^{\mathrm{I}}_{J,n})_n$ *and* $(\hat{d}^{\mathrm{I}}_{j,n})_n$*, with* $j \in \{1, \dots, J\}$*,*
- *description 2 is composed of:* $(a^{\mathrm{II}}_{J,n})_n$ *and* $(\check{d}^{\mathrm{I}}_{j,n})_n$*, with* $j \in \{1, \dots, J\}$*.*

3. **MD2-Scheme**: This scheme stems from the decomposition onto the set of functions:

$$\begin{aligned}&\bigcup_{j=1}^{J-1} \Big\{ \frac{1}{2^{j/2}} \psi(\frac{t}{2^j} - n), n \in \mathbb{Z} \Big\} \cup \bigcup_{p \in \{0,1\}} \Big\{ \frac{1}{2^{J/2}} \psi(\frac{t}{2^J} + \frac{p}{2} - 2n), n \in \mathbb{Z} \Big\} \cup \\ &\cup \bigcup_{p \in \{0,1\}} \Big\{ \frac{1}{2^{J/2}} \phi(\frac{t}{2^J} + \frac{p}{2} - n), n \in \mathbb{Z} \Big\}.\end{aligned} \tag{2.12}$$

In contrast with the previous scheme, we keep some of the detail coefficients from the second decomposition (at the last decomposition level) along with the corresponding approximation coefficients. The two descriptions are thus:

- *description 1 remains unchanged from before:* $(a^{\mathrm{I}}_{J,n})_n$ *and* $(\hat{d}^{\mathrm{I}}_{j,n})_n$*, with* $j \in \{1, \dots, J\}$*,*
- *description 2 has a different subsampling of the detail coefficients at the last decomposition level, which results in the following sequences:* $(a^{\mathrm{II}}_{J,n})_n$*,* $(\check{d}^{\mathrm{I}}_{j,n})_n$*, with* $j \in \{1, \dots, J-1\}$ *and* $(\hat{d}^{\mathrm{II}}_{J,n})_n$*, where* $\hat{d}^{\mathrm{II}}_{J,n} = d^{\mathrm{II}}_{J,2n}$*.*

4. **MD3-Scheme**: In this last scheme we have the following union of functions:

$$\begin{aligned}&\bigcup_{j=1}^{J-1} \Big\{ \frac{1}{2^{j/2}} \psi(\frac{t}{2^j} - n), n \in \mathbb{Z} \Big\} \cup \Big\{ \frac{1}{2^{J/2}} \psi(\frac{t}{2^J} - 2n), n \in \mathbb{Z} \Big\} \cup \\ &\Big\{ \frac{1}{2^{J/2}} \psi(\frac{t}{2^J} + \frac{3}{2} - 2n), n \in \mathbb{Z} \Big\} \cup \bigcup_{p \in \{0,1\}} \Big\{ \frac{1}{2^{J/2}} \phi(\frac{t}{2^J} + \frac{p}{2} - n), n \in \mathbb{Z} \Big\}.\end{aligned} \tag{2.13}$$

The two descriptions thus contain the following sets of wavelet coefficients:

- *description 1 is still the same as in the* **MD2***-Scheme:* $(\hat{d}^{\mathrm{I}}_{j,n})_n$ *for* $j \in \{1, \dots, J\}$*, and* $(a^{\mathrm{I}}_{J,n})_n$*,*
- *description 2 is also built as above until the last level of decomposition, where the index parity for the kept details change. This leads to the sequences:* $(\check{d}^{\mathrm{I}}_{j,n})_n$ *for* $j \in \{1, \dots, J-1\}$*,* $(\check{d}^{\mathrm{II}}_{J,n})_n$*, and* $(a^{\mathrm{II}}_{J,n})_n$*, where* $\check{d}^{\mathrm{II}}_{J,n} = d^{\mathrm{II}}_{J,2n-1}$*.*

Note that at the coarsest resolution other decompositions combining different approximation coefficients can be obtained by exchanging the I and II superscripts but, by symmetry, they are equivalent to one of the **MD1**, **MD2** or **MD3** schemes.

2.2 Wavelet Frame Considerations

The families of functions used in Eqs. (2.10) is obviously a (tight) frame of $V_0 = \text{Vect}\{\phi(t-n), n \in \mathbb{Z}\}$. One important question arising at this point is to know whether the other families (Eqs. (2.11), (2.12) and (2.13)) form frames of the vector space V_0 or not.

In practice, this frame structure guarantees the completeness of the proposed representation, that is the ability to perfectly reconstruct an original signal belonging to V_0 from the two descriptions, in the absence of quantization.

As we have seen in Chapter 1, a well-known condition for this property to hold [Dau92], is that the energy of the wavelet coefficients in the representation is bounded by two positive constants A and B.

We shall now express a necessary and sufficient condition for the frame structure, in the following proposition (in which we have considered the decomposition proposed in the **MD3**-scheme).

Proposition 1. *(Necessary and sufficient frame condition)*
There exists $(A, B) \in (\mathbb{R}_+^*)^2$ *such that, for all* $x \in V_0$,

$$A\|x\|^2 \leq \sum_{j=1}^{J-1}\sum_n |d^{\mathrm{I}}_{j,n}|^2 + \sum_n |\hat{d}^{\mathrm{I}}_{J,n}|^2 + \sum_n |\check{d}^{\mathrm{II}}_{J,n}|^2 + \sum_n |a^{\mathrm{I}}_{J,n}|^2 + \sum_n |a^{\mathrm{II}}_{J,n}|^2 \leq B\|x\|^2, \tag{2.14}$$

is a frame if and only if there exists $(\alpha, \beta) \in (\mathbb{R}_+^*)^2$ *such that* $\forall x \in V_0$ *and* $a^{\mathrm{I}}_{J-1,n} \in l^2(\mathbb{Z})$*:*

$$\alpha \sum_n |a^{\mathrm{I}}_{J-1,n}|^2 \leq \sum_n |\hat{d}^{\mathrm{I}}_{J,n}|^2 + \sum_n |\check{d}^{\mathrm{II}}_{J,n}|^2 + \sum_n |a^{\mathrm{I}}_{J,n}|^2 + \sum_n |a^{\mathrm{II}}_{J,n}|^2 \leq \beta \sum_n |a^{\mathrm{I}}_{J-1,n}|^2, \tag{2.15}$$

where $\|\cdot\|$ *denotes the usual norm of the space* $L^2(\mathbb{R})$.

It should be noted that the other possible subsamplings of the detail coefficients at the last decomposition level, corresponding to the other MDC schemes can by treated similarly, using the same arguments. The above proposition reduces the perfect reconstruction problem to the last level of wavelet decomposition in the redundant schemes and this shall be studied in a later section.

For the moment let us verify that Eq. (2.14)⇔(2.15).

Proof.

Let us first introduce:

$$E = \sum_{j=1}^{J-1}\sum_n |d^{\mathrm{I}}_{j,n}|^2 + \sum_n |\hat{d}^{\mathrm{I}}_{J,n}|^2 + \sum_n |\check{d}^{\mathrm{II}}_{J,n}|^2 + \sum_n |a^{\mathrm{I}}_{J,n}|^2 + \sum_n |a^{\mathrm{II}}_{J,n}|^2.$$

We know that since, $\bigcup_{j=1}^{J-1}\{2^{-j/2}\psi(2^{-j}t-n), n \in \mathbb{Z}\} \cup \{2^{-(J-1)/2}\phi(2^{-J+1}t-n), n \in \mathbb{Z}\}$ has been assumed to be biorthogonal basis of V_0 then $\exists (A', B') \in (\mathbb{R}_+^*)^2$ such that, for all $x \in V_0$,

$$A'\|x\|^2 \leq \sum_{j=1}^{J-1}\sum_n |d^{\mathrm{I}}_{j,n}|^2 + \sum_n |a^{\mathrm{I}}_{J-1,n}|^2 \leq B'\|x\|^2. \tag{2.16}$$

Let us first prove the implication (2.15) $\Longrightarrow$ (2.14).

With the introduced notation and from Eq. (2.16) we have:

$$\begin{aligned} E &\leq \sum_{j=1}^{J-1}\sum_{n}|d^{\mathrm{I}}_{j,n}|^2+\beta\sum_{n}|a^{\mathrm{I}}_{J-1,n}|^2 \\ &\leq \max(1,\beta)(\sum_{j=1}^{J-1}\sum_{n}|d^{\mathrm{I}}_{j,n}|^2+\sum_{n}|a^{\mathrm{I}}_{J-1,n}|^2) \\ &\leq \max(1,\beta)B'\|x\|^2. \end{aligned}$$

Since $B'\max(1,\beta) > 0$ we can choose $B = B'\max(1,\beta)$ as the searched upper bound in Eq. (2.14).

Similarly, we have:

$$\begin{aligned} E &\geq \sum_{j=1}^{J-1}\sum_{n}|d^{\mathrm{I}}_{j,n}|^2+\alpha\sum_{n}|a^{\mathrm{I}}_{J-1,n}|^2 \\ &\geq \min(1,\alpha)(\sum_{j=1}^{J-1}\sum_{n}|d^{\mathrm{I}}_{j,n}|^2+\sum_{n}|a^{\mathrm{I}}_{J-1,n}|^2) \\ &\geq \min(1,\alpha)A'\|x\|^2. \end{aligned}$$

Therefore, we have proven that $\exists A = A'\min(1,\alpha) > 0$ as the lower bound in Eq. (2.14).

Let us now prove the reciprocal statement (2.15) $\Longleftarrow$ (2.14).

For all $x \in V_0$, we know from Eq. (2.14) that

$$A\|x\|^2 \leq E \leq B\|x\|^2.$$

Then, according to Eq. (2.16), we have

$$\|x\|^2 \leq \frac{1}{A'}\left[\sum_{j=1}^{J-1}\sum_{n}|d^{\mathrm{I}}_{j,n}|^2+\sum_{n}|a^{\mathrm{I}}_{J-1,n}|^2\right]$$

which leads to:

$$E \leq \frac{B}{A'}\left[\sum_{j=1}^{J-1}\sum_{n}|d^{\mathrm{I}}_{j,n}|^2+\sum_{n}|a^{\mathrm{I}}_{J-1,n}|^2\right].$$

In other words:

$$\sum_{j=1}^{J-1}\sum_{n}|d^{\mathrm{I}}_{j,n}|^2+\sum_{n}|\hat{d}^{\mathrm{I}}_{J,n}|^2+\sum_{n}|\check{d}^{\mathrm{II}}_{J,n}|^2+\sum_{n}|a^{\mathrm{I}}_{J,n}|^2+\sum_{n}|a^{\mathrm{II}}_{J,n}|^2 \leq \frac{B}{A'}\left[\sum_{n}|d^{\mathrm{I}}_{j,n}|^2+\sum_{n}|a^{\mathrm{I}}_{J-1,n}|^2\right]. \tag{2.17}$$

We want to find an upper bound for $C = \sum_n|\hat{d}^{\mathrm{I}}_{J,n}|^2+\sum_n|\check{d}^{\mathrm{II}}_{J,n}|^2+\sum_n|a^{\mathrm{I}}_{J,n}|^2+\sum_n|a^{\mathrm{II}}_{J,n}|^2$. By considering the approximation space at resolution level $J-1$:

$$V_{J-1} = \mathrm{Vect}\Big\{2^{-\frac{J-1}{2}}\phi\big(\frac{t}{2^{J-1}}-n\big)\Big\},$$

any arbitrary signal x in $V_{J-1} \subset V_0$ can be written as:

$$x(t) = \sum_n a^{\mathrm{I}}_{J-1} 2^{-(J-1)/2} \phi(2^{-J+1}t - n) \text{ and}$$
$$d^{\mathrm{I}}_{j,n} = 0, \quad \forall j < J,$$

since $V_{J-1} = V_J \oplus W_J$. Thus Eq. (2.17) reduces itself to:

$$\sum_n |\hat{d}^{\mathrm{I}}_{J,n}|^2 + \sum_n |\check{d}^{\mathrm{II}}_{J,n}|^2 + \sum_n |a^{\mathrm{I}}_{J,n}|^2 + \sum_n |a^{\mathrm{II}}_{J,n}|^2 \leq \frac{B}{A'} \sum_n |a^{\mathrm{I}}_{J-1,n}|^2.$$

This shows that Eq. (2.15) is satisfied with $\beta = \frac{B}{A'} > 0$.

From Eq. (2.14), we also deduce that

$$E \geq A\|x\|^2 \geq \frac{A}{B'} \left(\sum_{j=1}^{J-1} \sum_n |d^{\mathrm{I}}_{j,n}|^2 + \sum_n |a^{\mathrm{I}}_{J-1,n}|^2 \right)$$

and, by invoking similar arguments, we obtain: $\alpha = \frac{A}{B'} > 0$. □

We have thus found a necessary and sufficient condition for Eq. (2.14).

The perfect reconstruction at the last wavelet decomposition level is tackled by passing to an equivalent filter bank approach which will allow us to express the transfer matrix corresponding to the MDC system. Then, the perfect reconstruction at the coarsest level is ensured by the invertibility of this transfer matrix.

Moreover, by using the polyphase formalism, we can easily express the global polyphase transfer matrix of the system, in which we also highlight the different additional subsampling operations that are possible for the chosen frame structure.

Finding a solution for the system inversion is not only useful in order to guarantee the perfect reconstruction of our schemes, but it equally gives the guidelines for the implementation of the (central) decoders.

Before going to this step let us rewrite Eq. (2.15) in a simpler form. Recall that, the condition (2.15) only involves sequences of coefficients at the last decomposition stage, J. Then, for the sake of simplicity, we can subsequently omit the index J and furthermore denote the approximation coefficients at resolution level $J-1$ by $(x_n)_{n\in\mathbb{Z}}$ so as to avoid scale ambiguities.

With the newly adopted notations, the inequality in (2.15) is then rewritten as

$$\alpha \sum_n |x_n|^2 \leq \sum_n |a^{\mathrm{I}}_n|^2 + \sum_n |a^{\mathrm{II}}_n|^2 + \sum_n |\hat{d}^{\mathrm{I}}_n|^2 + \sum_n |\check{d}^{\mathrm{II}}_n|^2 \leq \beta \sum_n |x_n|^2 \tag{2.18}$$

In this particular case the upper-indices I and II point out the description number in addition to the wavelet basis.

Let us now express our MDC schemes in terms of oversampled filter banks.

2.3 Filter Bank Representations for Discrete Frames

It is now well-known [Mal98], [VK95] that there exists an equivalent filter bank structure leading to these sequences of wavelet coefficients and that this allows for fast implementation of the

transform as well as for simple reconstruction. Therefore, in this section we are investigating the filter banks related to our proposed schemes. The filter bank approach to frames has also been recently used for Joint-Source Channel Coding in [MG04b].

Let us now consider the discrete one-dimensional input signal, $(x_n)_{n\in\mathbb{Z}}$. This could be, for instance, the signal given by the temporal variation of the a pixel in the frames in a video sequence[4].

In the following we are expressing the previously defined wavelet subbands in terms of filter banks and for discrete input signals.

Let $(h_n)_{n\in\mathbb{Z}}$ and $(g_n)_{n\in\mathbb{Z}}$ be the impulse responses of the analysis low-pass and high-pass filters, respectively, corresponding to the considered multiresolution decomposition.

Then in our MDC schemes we perform a standard wavelet decomposition for the first $J-1$ resolution levels, which leads to the equivalent writing of Eqs. (2.6) and (2.7), as follows:

$$a^{\mathrm{I}}_{j,n} = \sum_k h_{2n-k}\; a^{\mathrm{I}}_{j-1,k},$$

for the temporal approximation subband, and

$$d^{\mathrm{I}}_{j,n} = \sum_k g_{2n-k}\; a^{\mathrm{I}}_{j-1,k},$$

for the detail one, where $j \in \{1,\ldots,J-1\}$. The four schemes discussed in section 2.1.2 are different at the last level of wavelet decomposition, J.

At this point let us recall that we perform an additional subsampling on the detail subbands, which we have indicated by the symbols $\hat{\cdot}$ and $\check{\cdot}$ for the odd and even-sampled sequences, respectively.

Then, let $\mathbf{c}_n$ be the vector which contains all the possible subsampled sequences involved in any of the schemes, that is

$$\mathbf{c}_n = \left(\hat{a}^{\mathrm{I}}_n\; \check{a}^{\mathrm{I}}_n\; \hat{a}^{\mathrm{II}}_n\; \check{a}^{\mathrm{II}}_n\; \hat{d}^{\mathrm{I}}_n\; \check{d}^{\mathrm{I}}_n\; \hat{d}^{\mathrm{II}}_n\; \check{d}^{\mathrm{II}}_n\right)^{\mathsf{T}}.$$

We shall now express each component of $\mathbf{c}_n$ in terms of filter bank decomposition. For the coefficients generated by the "classical" wavelet scheme and further subsampled by a factor of 2, we have:

$$\begin{cases} \hat{a}^{\mathrm{I}}_n = a^{\mathrm{I}}_{2n} = \displaystyle\sum_k x_k h_{4n-k} \\ \check{a}^{\mathrm{I}}_n = a^{\mathrm{I}}_{2n-1} = \displaystyle\sum_k x_k h_{4n-2-k} \\ \hat{d}^{\mathrm{I}}_n = \displaystyle\sum_k x_k g_{4n-k}. \\ \check{d}^{\mathrm{I}}_n = \displaystyle\sum_k x_k g_{4n-2-k}. \end{cases} \tag{2.19}$$

As we have previously seen, the coefficient sets a^{II}_n and d^{II}_n result from the decomposition onto the translated basis $\mathcal{B}^{\mathrm{II}}$. This is equivalent in terms of filter banks to the convolution with the

[4] We have chosen this particular example because it suits the application proposed in the second half of this chapter, but the theoretical results stand for any general 1D signal.

translated filters having the following impulse responses:

$$\tilde{h}_n = h_{n-1}, \qquad \tilde{g}_n = g_{n-1}.$$

This leads to the sequences:

$$\begin{cases} \hat{a}_n^{\mathrm{II}} = a_{2n}^{\mathrm{II}} = \sum_k x_k h_{4n-1-k} \\ \check{a}_n^{\mathrm{II}} = a_{2n-1}^{\mathrm{II}} = \sum_k x_k h_{4n-3-k} \\ \hat{d}_n^{\mathrm{II}} = \sum_k x_k g_{4n-1-k} \\ \check{d}_n^{\mathrm{II}} = \sum_k x_k g_{4n-3-k}. \end{cases} \tag{2.20}$$

The four sets of coefficients generated by the R-scheme are obviously obtained from a critically sampled filter bank. We will now see that this scheme as well as the other redundant schemes can be viewed as an oversampled 6-band filter bank structure with a decimation factor of 4.

In order to express these equations in terms of the desired 6×4 oversampled filter banks, in each of the four schemes discussed earlier, we now define the polyphase components of the impulse responses of the filters as follows:

$$\forall i \in \{0,1,2,3\}, \qquad h_i(n) = h_{4n-i}, \quad g_i(n) = g_{4n-i}$$

and introduce the corresponding z-transforms[5]: $H_i(z)$ and $G_i(z)$. Similarly, the input signal split into four polyphase components can be written as:

$$\forall i \in \{0,1,2,3\}, \qquad x_n^{(i)} = x_{4n+i}.$$

With this notation, the corresponding polyphase component vector for the input signal is

$$\mathbf{x}_n = \left(x_n^{(0)}\ x_n^{(1)}\ x_n^{(2)}\ x_n^{(3)}\right)^{\mathsf{T}}.$$

In order to describe the polyphase transfer matrices corresponding to the oversampled filter banks of the different schemes, we use the following global representation:

$$\mathbf{C}(z) = \mathbf{M}(z)\ \mathbf{X}(z),$$

where $\mathbf{C}(z)$ and $\mathbf{X}(z)$ are the z-transforms of the coefficient vector sequence, $\mathbf{c}_n$, and of the input signal, $\mathbf{x}_n$ and $\mathbf{M}(z)$ is the global polyphase transfer matrix (of the fully redundant system containing both complete wavelet representations for a one-level decomposition).

2.3.1 Expression of the polyphase transfer function matrix for the MD schemes

We are now aiming at finding the expression of the global polyphase transfer matrix as well as those of the transfer matrices corresponding to each of the MD schemes.

[5]Recall that the z-transform of a discrete-time signal, $(x_n)_{n\in\mathbb{Z}}$, is given by $X(z) = \sum_{n\in\mathbb{Z}} x_n z^{-n}$, with $z \in \mathbb{C}$.

By highlighting the four polyphase components of the input signal and according to Eq. (2.19), the approximation coefficients in the first wavelet representation can be expressed as

$$\hat{a}_n^{\mathrm{I}} = \sum_k x_{4k} h_{4(n-k)} + \sum_k x_{4k+1} h_{4(n-k)-1} + \sum_k x_{4k+2} h_{4(n-k)-2} + \sum_k x_{4k+3} h_{4(n-k)-3} \tag{2.21}$$

$$\check{a}_n^{\mathrm{I}} = \sum_k x_{4k} h_{4(n-k)-2} + \sum_k x_{4k+1} h_{4(n-k)-3} + \sum_k x_{4k+2} h_{4(n-k-1)} + \sum_k x_{4k+3} h_{4(n-k-1)-1}. \tag{2.22}$$

Similarly, from Eq. (2.20), the approximation coefficients in the second wavelet representation are given by:

$$\hat{a}_n^{\mathrm{II}} = \sum_k x_{4k} h_{4(n-k)-1} + \sum_k x_{4k+1} h_{4(n-k)-2} + \sum_k x_{4k+2} h_{4(n-k)-3} + \sum_k x_{4k+3} h_{4(n-k-1)} \tag{2.23}$$

$$\check{a}_n^{\mathrm{II}} = \sum_k x_{4k} h_{4(n-k)-3} + \sum_k x_{4k+1} h_{4(n-k-1)} + \sum_k x_{4k+2} h_{4(n-k-1)-1} + \sum_k x_{4k+3} h_{4(n-k-1)-2}. \tag{2.24}$$

Now, taking the z-transform of the expressions in Eqs. (2.21) - (2.24), gives:

$$\begin{cases} \hat{A}^{\mathrm{I}}(z) = X_0(z)H_0(z) + X_1(z)H_1(z) + X_2(z)H_2(z) + X_3(z)H_3(z) \\ \check{A}^{\mathrm{I}}(z) = X_0(z)H_2(z) + X_1(z)H_3(z) + z^{-1}X_2(z)H_0(z) + z^{-1}X_3(z)H_1(z) \\ \hat{A}^{\mathrm{II}}(z) = X_0(z)H_1(z) + X_1(z)H_2(z) + X_2(z)H_3(z) + z^{-1}X_3(z)H_0(z) \\ \check{A}^{\mathrm{II}}(z) = X_0(z)H_3(z) + z^{-1}X_1(z)H_0(z) + z^{-1}X_2(z)H_1(z) + z^{-1}X_3(z)H_2(z) \end{cases}$$

where capital characters are used to denote the z-transforms of the different sequences of coefficients.

Similar expressions are obtained for the z-transforms of $\hat{d}_n^{\mathrm{I}}$, $\check{d}_n^{\mathrm{I}}$, $\hat{d}_n^{\mathrm{II}}$ and $\check{d}_n^{\mathrm{II}}$ by replacing the polyphase components of h_n by those of g_n.

Thus, the expression of the global transfer matrix uniting the two critically-sampled wavelet decompositions into a factor 2-redundant structure is given by:

$$\mathbf{M}(z) = \begin{bmatrix} H_0(z) & H_1(z) & H_2(z) & H_3(z) \\ H_2(z) & H_3(z) & H_0(z)z^{-1} & H_1(z)z^{-1} \\ H_1(z) & H_2(z) & H_3(z) & H_0(z)z^{-1} \\ H_3(z) & H_0(z)z^{-1} & H_1(z)z^{-1} & H_2(z)z^{-1} \\ G_0(z) & G_1(z) & G_2(z) & G_3(z) \\ G_2(z) & G_3(z) & G_0(z)z^{-1} & G_1(z)z^{-1} \\ G_1(z) & G_2(z) & G_3(z) & G_0(z)z^{-1} \\ G_3(z) & G_0(z)z^{-1} & G_1(z)z^{-1} & G_2(z)z^{-1} \end{bmatrix}.$$

An MD-scheme following the proposed reduced-redundancy strategy is obtained by taking a subset of $\mathbf{c}_n$ having 6 components (3 for each description). Thus we build a new vector, denoted by $\overline{\mathbf{c}}_n$, for any of the specific schemes previously described and resulting from the equation:

$$\overline{\mathbf{C}}(z) = \overline{\mathbf{M}}(z)\ \mathbf{X}(z).$$

Here, $\overline{\mathbf{C}}(z)$ is the z-transform of $(\overline{\mathbf{c}}_n)_{n\in\mathbb{Z}}$ and $\overline{\mathbf{M}}(z)$ is the polyphase transfer function of the considered MD scheme. The corresponding transfer matrix for each scheme is obtained by extracting some of the lines of $\mathbf{M}(z)$ as we will show in the following.

Let us review the four schemes discussed earlier, this time from the filter bank perspective, in order to identify the polyphase transfer matrix in each of these cases. We shall also present some considerations on the possibility to invert these schemes.

R-Scheme

As we said earlier, this scheme consists in splitting the detail coefficients of the decomposition onto $\mathcal{B}^{\mathrm{I}}$ into two groups: even-index coefficients and odd-index coefficients, each group belonging to one of the descriptions. The approximation coefficients are simply duplicated. This corresponds to the vector:

$$\{\underbrace{\hat{a}_n^{\mathrm{I}},\ \check{a}_n^{\mathrm{I}},\ \hat{d}_n^{\mathrm{I}},}_{\text{1st description}}\ \underbrace{\hat{a}_n^{\mathrm{I}},\ \check{a}_n^{\mathrm{I}},\ \check{d}_n^{\mathrm{I}}}_{\text{2nd description}}\}.$$

Therefore, the multiple description coefficient subset of $\mathbf{c}_n$, denoted above by $\overline{\mathbf{c}}_n$, corresponding to this scheme is given by $\overline{\mathbf{c}}_n = (\hat{a}_n^{\mathrm{I}}\ \check{a}_n^{\mathrm{I}}\ \hat{d}_n^{\mathrm{I}}\ \check{d}_n^{\mathrm{I}})^{\mathsf{T}}$. In this case, $\overline{\mathbf{M}}(z)$ is formed with the $1^{\mathrm{st}}, 2^{\mathrm{nd}}, 5^{\mathrm{th}},$ and the 6^{th} line of $\mathbf{M}(z)$ (lines 1 and 2 are used twice). More precisely this corresponds to the following transfer matrix:

$$\overline{\mathbf{M}}(z) = \begin{bmatrix} H_0(z) & H_1(z) & H_2(z) & H_3(z) \\ H_2(z) & H_3(z) & H_0(z)z^{-1} & H_1(z)z^{-1} \\ G_0(z) & G_1(z) & G_2(z) & G_3(z) \\ G_2(z) & G_3(z) & G_0(z)z^{-1} & G_1(z)z^{-1} \end{bmatrix}.$$

In the absence of quantization, the invertibility of this decomposition is immediate since it is derived from a classical critically-sampled filter bank with perfect reconstruction.

MD1-Scheme

In this scheme we have split the detail subbands according to the same scheme as above, but in the second description, instead of repeating the approximation coefficients from the first decomposition, we have used the approximation coefficients from the second one. This corresponds to

$$\{\underbrace{\hat{a}_n^{\mathrm{I}},\ \check{a}_n^{\mathrm{I}},\ \hat{d}_n^{\mathrm{I}},}_{\text{1st description}}\ \underbrace{\hat{a}_n^{\mathrm{II}},\ \check{a}_n^{\mathrm{II}},\ \check{d}_n^{\mathrm{I}}}_{\text{2nd description}}\}.$$

and $\overline{\mathbf{M}}(z)$ is formed with the $1^{\mathrm{st}}, 2^{\mathrm{nd}}, 5^{\mathrm{th}}, 3^{\mathrm{rd}}, 4^{\mathrm{th}},$ and the 6^{th} lines of $\mathbf{M}(z)$. Thus,

$$\overline{\mathbf{M}}(z) = \begin{bmatrix} H_0(z) & H_1(z) & H_2(z) & H_3(z) \\ H_2(z) & H_3(z) & H_0(z)z^{-1} & H_1(z)z^{-1} \\ H_1(z) & H_2(z) & H_3(z) & H_0(z)z^{-1} \\ H_3(z) & H_0(z)z^{-1} & H_1(z)z^{-1} & H_2(z)z^{-1} \\ G_0(z) & G_1(z) & G_2(z) & G_3(z) \\ G_2(z) & G_3(z) & G_0(z)z^{-1} & G_1(z)z^{-1} \end{bmatrix}.$$

The coefficient vector is in this case: $\overline{\mathbf{c}}_n = (\hat{a}_n^{\mathrm{I}}\ \check{a}_n^{\mathrm{I}}\ \hat{a}_n^{\mathrm{II}}\ \check{a}_n^{\mathrm{II}}\ \hat{d}_n^{\mathrm{I}}\ \check{d}_n^{\mathrm{I}})^{\mathsf{T}}$. As this scheme includes a complete decomposition onto a wavelet basis, its invertibility does not raise any problem as well.

MD2-Scheme

Here, the two wavelet decompositions are combined so that

$$\{\underbrace{\hat{a}_n^{\mathrm{I}},\ \check{a}_n^{\mathrm{I}},\ \hat{d}_n^{\mathrm{I}},}_{\text{1st description}}\ \underbrace{\hat{a}_n^{\mathrm{II}},\ \check{a}_n^{\mathrm{II}},\ \hat{d}_n^{\mathrm{II}}}_{\text{2nd description}}\},$$

which, with the chosen ordering convention, gives $\overline{\mathbf{c}}_n = (\hat{a}_n^{\mathrm{I}}\ \check{a}_n^{\mathrm{I}}\ \hat{a}_n^{\mathrm{II}}\ \check{a}_n^{\mathrm{II}}\ \hat{d}_n^{\mathrm{I}}\ \hat{d}_n^{\mathrm{II}})^{\mathsf{T}}$. The transfer matrix for this scheme, $\overline{\mathbf{M}}(z)$, is formed with the $1^{\mathrm{st}}, 2^{\mathrm{nd}}, 5^{\mathrm{th}}, 3^{\mathrm{rd}}, 4^{\mathrm{th}}$ and the 7^{th} lines of $\mathbf{M}(z)$. Or, explicitly:

$$\overline{\mathbf{M}}(z) = \begin{bmatrix} H_0(z) & H_1(z) & H_2(z) & H_3(z) \\ H_2(z) & H_3(z) & H_0(z)z^{-1} & H_1(z)z^{-1} \\ H_1(z) & H_2(z) & H_3(z) & H_0(z)z^{-1} \\ H_3(z) & H_0(z)z^{-1} & H_1(z)z^{-1} & H_2(z)z^{-1} \\ G_0(z) & G_1(z) & G_2(z) & G_3(z) \\ G_1(z) & G_2(z) & G_3(z) & G_0(z)z^{-1} \end{bmatrix}.$$

MD3-Scheme

By selecting the odd-subsampled detail coefficients in the second wavelet decomposition rather than the even-subsampled one, we get:

$$\{\underbrace{\hat{a}_n^{\mathrm{I}},\ \check{a}_n^{\mathrm{I}},\ \hat{d}_n^{\mathrm{I}},}_{\text{1st description}}\ \underbrace{\hat{a}_n^{\mathrm{II}},\ \check{a}_n^{\mathrm{II}},\ \check{d}_n^{\mathrm{II}}}_{\text{2nd description}}\}$$

and $\overline{\mathbf{c}}_n = (\hat{a}_n^{\mathrm{I}}\ \check{a}_n^{\mathrm{I}}\ \hat{a}_n^{\mathrm{II}}\ \check{a}_n^{\mathrm{II}}\ \hat{d}_n^{\mathrm{I}}\ \check{d}_n^{\mathrm{II}})^{\mathsf{T}}$. Now, $\overline{\mathbf{M}}(z)$ is formed with the $1^{\mathrm{st}}, 2^{\mathrm{nd}}, 5^{\mathrm{th}}, 3^{\mathrm{rd}}, 4^{\mathrm{th}},$ and the 8^{th} lines of $\mathbf{M}(z)$. We then get,

$$\overline{\mathbf{M}}(z) = \begin{bmatrix} H_0(z) & H_1(z) & H_2(z) & H_3(z) \\ H_2(z) & H_3(z) & H_0(z)z^{-1} & H_1(z)z^{-1} \\ H_1(z) & H_2(z) & H_3(z) & H_0(z)z^{-1} \\ H_3(z) & H_0(z)z^{-1} & H_1(z)z^{-1} & H_2(z)z^{-1} \\ G_0(z) & G_1(z) & G_2(z) & G_3(z) \\ G_3(z) & G_0(z)z^{-1} & G_1(z)z^{-1} & G_2(z)z^{-1} \end{bmatrix}.$$

In the case of **MD2** and **MD3** schemes, their invertibility is not immediate. This has determined the study presented in the following section which looks for invertibility criteria for the polyphase transfer matrix. However, since we are designing oversampled systems, the inverses, if they exist, are not unique. Therefore we have also studied the influence of the chosen inverse on the quantization noise occurring during transmission.

2.4 Invertibility Using the Polyphase Transfer Matrix

Assuming that the analysis filters are stable (i.e. $(h_n)_{n\in\mathbb{Z}}$ and $(g_n)_{n\in\mathbb{Z}}$ belong to the space $\ell^1(\mathbb{Z}) \subset \ell^2(\mathbb{Z})$ of summable sequences), this means that the function $\omega \mapsto \overline{\mathbf{M}}(e^{i\omega})$ is contin-

uous and bounded on $[0, 2\pi]$. By expressing the energy of the coefficients as:

$$\int_0^{2\pi} \frac{1}{2\pi}|C(e^{i\omega})|^2 \, d\omega = \int_0^{2\pi} \frac{1}{2\pi}|\overline{\mathbf{M}}(e^{i\omega})X(e^{i\omega})|^2 \, d\omega$$
$$\leq \int_0^{2\pi} \frac{1}{2\pi}|\overline{\mathbf{M}}(e^{i\omega})|^2|X(e^{i\omega})|^2 \, d\omega,$$

and considering that $\overline{\mathbf{M}}(e^{i\omega})$ bounded implies $\exists \beta \in \mathbb{R}$, such that:

$$|\overline{\mathbf{M}}(e^{i\omega})| \leq \beta,$$

this leads to:

$$\int_0^{2\pi} \frac{1}{2\pi}|C(e^{i\omega})|^2 \, d\omega \leq \beta \int_0^{2^\pi} \frac{1}{2\pi}|X(e^{i\omega})|.$$

But, from the Parseval relation, $\int_0^{2\pi} \frac{1}{2\pi}|X(e^{i\omega})|^2 d\omega$ equals $\int_0^T |x(t)|^2 dt$. Passing into discrete time again, we conclude that the energy of the wavelet coefficients corresponding to the chosen redundant representation is bounded by $\beta \sum_n |x_n|^2$. This is equivalent to saying that the condition (2.18) is satisfied if and only if the oversampled filter bank leading to these wavelet sequences is with perfect reconstruction [BHH98]. Or, more precisely, (2.18) holds if and only if the rank of $\overline{\mathbf{M}}(e^{i\omega})$ is equal to the number of polyphase components (4 in our case) for all $\omega \in [0, 2\pi)$ [BHH98], [CV98]. Note that, by having less columns than rows in $\overline{\mathbf{M}}(z)$, the corresponding MIMO (Multiple-Input Multiple-Output) linear system is likely to admit several inverses and one of the problems to be addressed is choosing the most appropriate one.

In the particular case when the analysis filters are with Finite Impulse Response (FIR), sharper results concerning the invertibility of Laurent polynomial matrices can be applied, [FV97], [GL99]. These results are especially useful for the schemes **MD2** and **MD3**, whose invertibility is not a priori guaranteed. A polynomial matrix having more lines than columns is left-invertible by another polynomial matrix provided that all its maximal order minors (determinants of the square submatrices of maximal size) are coprime, i.e. these determinants do not share a common non-null zero. This is also equivalent to saying that the matrix is full rank for all $z \in \mathbb{C}^*$. By studying the minors of the polyphase transfer function matrix, three cases may occur:

1. All the maximal order minors have a common zero on the unit circle and the system is not invertible.

2. There are non-null zeros common to the maximal order minors but none of these zeros lies on the unit circle. Then, the system is left-invertible by a rational stable MIMO filter having an Infinite Impulse Response (IIR).

3. The maximal order minors have no common non-null zero and consequently the system is left-invertible by an FIR MIMO filter.

Note that, if one of the maximal order minors reduces to a monomial, we are necessarily in the third situation. Then, the sub-system corresponding to the lines of the minor constitutes an oversampled filter bank which is invertible by an FIR MIMO filter. This provides a solution to the inversion problem, which may however be suboptimal as only four coefficient sequences are used to reconstruct $(x_n)_{n\in\mathbb{Z}}$.

We shall later on illustrate these facts through the examples of Haar and biorthogonal 5/3 decompositions.

In the case where a causal FIR MIMO inverse is needed the following Lemma [Kai80], [GL99] has been proposed in the literature:

Lemma 1. *Given* $\mathbf{M}(z) \in \mathbb{C}[z]^{K\times N}$ *a polynomial matrix with* $K > N$, *if* $rank(\mathbf{M}(z)) = n$, $\forall z$ *and* $\mathbf{M}(\mathbf{z})$ *is also column reduced, i.e.* $rank([M_1(l_1), \ldots M_N(l_N)]) = n$, *where* l_i, $i \in 1, \ldots, N$ *are the maximum degrees of each column* $M_i(z)$, $i \in 1, \ldots, N$ *of the matrix* $M(z)$, *then* $\exists Q \geq 0$ *and a polynomial matrix* $\mathbf{W}(z) = \sum_{q=0}^{Q} W_q \cdot z^{-q}$, *such that:*

$$\mathbf{W}(z)\mathbf{M}(z) = \mathbf{I}_N.$$

When $z \in \mathbb{C}^*$ these conditions do not necessarily hold, therefore we shall not impose the causality property for the synthesis FIR filter-banks. Moreover, when applying our schemes to video coding, the causality might even be a drawback since the reconstruction might involve coefficients lying in a too far away group of pictures (GOP) as referred to the current GOP.

2.4.1 Solution of the System Inversion

Recall that, for the considered filter banks, we have found a necessary and sufficient condition for perfect reconstruction from the two descriptions, in the absence of quantization. This amounts to proving the invertibility of the MDC scheme at the last level of wavelet decomposition. The wavelet sequences involved at this level were obtained from the temporal signal via the polyphase formulation of the transfer function. Let us now further investigate the inversion of the polyphase matrix.

With the notations used in the previous section, we can formulate the problem as follows: we want to find a MIMO $N \times K$ transfer function $\mathbf{W}(z)$, such that

$$\mathbf{W}(z)\overline{\mathbf{M}}(z) = \mathbf{I}_{N\times N} \tag{2.25}$$

where $\mathbf{W}(z) = [W_{i,j}(z)]_{1\leq i\leq N, 1\leq j\leq K}$, $\overline{\mathbf{M}}(z) = [M_{i,j}(z)]_{1\leq i\leq K, 1\leq j\leq N}$ and $\mathbf{I}_{N\times N}$ is the identity matrix of order N. We remind that $K = 6$ is the number of wavelet coefficient sequences and $N = 4 < K$ is the number of polyphase components of the input signal, needed by the redundant transform. The maximum length of the scalar filters with transfer function $W_{i,j}(z)$ (resp. $M_{i,j}(z)$) is assumed to be equal to $P \in \mathbb{N}^*$ (resp. $Q \in \mathbb{N}^*$). We thus have to solve N^2 scalar polynomial equations in order to satisfy Eq. (2.25).

Let $\mathbf{W}(z)$ and $\overline{\mathbf{M}}(z)$ in Eq. (2.25) be explicitly written as Laurent polynomial matrices of the form:

$$\mathbf{W}(z) = \sum_{p=-p_1}^{p_2} \mathbf{W}_p \, z^{-p} \tag{2.26}$$

$$\overline{\mathbf{M}}(z) = \sum_{q=-q_1}^{q_2} \overline{\mathbf{M}}_q \, z^{-q} \tag{2.27}$$

where, for all p (resp. q) $\mathbf{W}_p$ (resp. $\overline{\mathbf{M}}_q$) is an $N \times K$ (resp. $K \times N$) matrix. In this case the maximal orders of the analysis and synthesis filters are $Q = q_1 + q_2 + 1$ and $P = p_1 + p_2 + 1$, respectively, where the constants p_1, p_2, q_1, q_2 belong to $\mathbb{N}$.

Then, the global $N \times N$ transfer function in the left-hand side of Eq. (2.25) reads

$$\mathbf{G}(z) = \mathbf{W}(z)\overline{\mathbf{M}}(z) = \sum_{l=-(p_1+q_1)}^{q_1+q_2} \mathbf{G}_l \, z^{-l}$$

where, for all l,

$$\mathbf{G}_s = \sum_{l=\max(-p_1, s-q_1)}^{\min(p_2, s+q_2)} \mathbf{W}_l \, \overline{\mathbf{M}}_{s-l}. \tag{2.28}$$

This shows that the solution $\mathbf{W}(z)$ of Eq. (2.25) is obtained by solving a system of $N^2(Q+P-1)$ linear equations. On the other hand, the number of unknown variables in $\mathbf{W}(z)$ is NKP.

Our goal is to find an inverse $\mathbf{W}(z)$ which satisfies Eq. (2.25) and is optimal in a sense that will be made more precise in Section 2.4.2. Using Eq. (2.28), Eq. (2.25) may be rewritten in the following matrix form:

$$\overline{\boldsymbol{\mathcal{M}}}\boldsymbol{\mathcal{W}} = \boldsymbol{\mathcal{U}} \tag{2.29}$$

where $\boldsymbol{\mathcal{W}}$ and $\boldsymbol{\mathcal{U}}$ are real matrices of sizes $KP \times N$ and $N(Q+P-1) \times N$ respectively, which are given by

$$\begin{aligned}
\boldsymbol{\mathcal{W}}^{\mathsf{T}} &= [\mathbf{W}_{-P+1} \ \dots \ \mathbf{W}_0] \\
\boldsymbol{\mathcal{U}}^{\mathsf{T}} &= [\underbrace{\mathbf{0}_{N\times N} \dots \mathbf{0}_{N\times N}}_{p_1+q_1 \text{ times}} \ \mathbf{I}_{N\times N} \ \underbrace{\mathbf{0}_{N\times N} \dots \mathbf{0}_{N\times N}}_{p_2+q_2 \text{ times}}]
\end{aligned}$$

whereas $\overline{\boldsymbol{\mathcal{M}}}^{\mathsf{T}}$ is the $KP \times N(Q+P-1)$ generalized Sylvester matrix:

$$\overline{\boldsymbol{\mathcal{M}}}^{\mathsf{T}} = \begin{bmatrix}
\boldsymbol{M}_0 & \boldsymbol{M}_1 & \dots & \boldsymbol{M}_{Q-1} & \mathbf{0}_{K\times N} & \dots & \dots & \mathbf{0}_{K\times N} \\
\mathbf{0}_{K\times N} & \boldsymbol{M}_0 & \boldsymbol{M}_1 & \dots & \boldsymbol{M}_{Q-1} & \mathbf{0}_{K\times N} & \dots & \mathbf{0}_{K\times N} \\
\vdots & \ddots & \ddots & \ddots & & \ddots & \ddots & \vdots \\
\vdots & & \ddots & \ddots & \ddots & & \ddots & \mathbf{0}_{K\times N} \\
\mathbf{0}_{K\times N} & \dots & \dots & \mathbf{0}_{K\times N} & \boldsymbol{M}_0 & \boldsymbol{M}_1 & \dots & \boldsymbol{M}_{Q-1}
\end{bmatrix}. \tag{2.30}$$

Finding $\boldsymbol{\mathcal{W}}$ obviously amounts to finding its column vectors. In order to identify these vectors, we can rewrite:

$$\boldsymbol{\mathcal{W}} = [\mathbf{w}_0 \dots \mathbf{w}_{N-1}] \tag{2.31}$$

$$\boldsymbol{\mathcal{U}} = [\mathbf{u}_0 \dots \mathbf{u}_{N-1}] \tag{2.32}$$

where, for all $i \in \{0, \dots, N-1\}$, $\mathbf{w}_i \in \mathbb{R}^{KP}$ and $\mathbf{u}_i \in \mathbb{R}^{N(Q+P-1)}$. Solving Eq. (2.29) is therefore equivalent to solving

$$\forall i \in \{0, \dots, N-1\}, \qquad \overline{\boldsymbol{\mathcal{M}}} \, \mathbf{w}_i = \mathbf{u}_i. \tag{2.33}$$

Let $\|\cdot\|_2$ denote the usual Euclidean norm. We further know that there exists a unique minimum norm vector $\mathbf{w}_i$ minimizing $\|\overline{\boldsymbol{\mathcal{M}}} \, \mathbf{w}_i - \mathbf{u}_i\|_2$, which is expressed as

$$\mathbf{w}_i = \overline{\boldsymbol{\mathcal{M}}}^{\sharp} \, \mathbf{u}_i \tag{2.34}$$

where $\mathbf{A}^{\sharp}$ denotes the pseudo-inverse of a matrix $\mathbf{A}$. The vector given by Eq. (2.34) is also a solution of Eq. (2.33), provided that, for all i, $\mathbf{u}_i$ belongs to the column space of $\overline{\boldsymbol{\mathcal{M}}}$. The latter

condition is fulfilled when $\mathrm{rank}([\overline{\boldsymbol{\mathcal{M}}}\ \boldsymbol{\mathcal{U}}]) = \mathrm{rank}(\overline{\boldsymbol{\mathcal{M}}})$. If $\overline{\mathbf{M}}(z)$ can be inverted by a Laurent polynomial matrix, there exists a minimal value of P for which this equality is reached. By combining Eqs. (2.34), (2.31) and (2.32), we then obtain the following whole system inversion solution:

$$\boldsymbol{\mathcal{W}} = \overline{\boldsymbol{\mathcal{M}}}^{\sharp}\, \boldsymbol{\mathcal{U}}. \tag{2.35}$$

2.4.2 Optimality criteria for system inverse

In terms of coding, it is of main importance to study the influence of the choice of $\boldsymbol{\mathcal{W}}$ on the effect of the quantization noise. By modeling the quantization as the addition of a noise on each sequence at the output of $\overline{\mathbf{M}}(z)$, we aim at reducing as much as possible the influence of this noise on the reconstruction $(\mathbf{x}'_n)_{n\in\mathbb{Z}}$ of $(\mathbf{x}_n)_{n\in\mathbb{Z}}$. We assume in the sequel that the noise vector sequence $(\mathbf{b}_n)_{n\in\mathbb{Z}}$ is zero-mean, independent and identically distributed with non singular covariance matrix $\boldsymbol{\Lambda}$. We obviously have

$$\mathbf{x}'_n = \mathbf{x}_n + \mathbf{v}_n$$

where $(\mathbf{v}_n)_{n\in\mathbb{Z}}$ is the multivariate moving average process defined by

$$\mathbf{v}_n = \sum_p \mathbf{W}_p \mathbf{b}_{n-p}\ .$$

The autocovariance matrix for $\mathbf{v}_n$ is

$$\mathsf{E}\{\mathbf{v}_n {\mathbf{v}_n}^{\mathsf{T}}\} = \sum_{p,q} \mathbf{W}_p \mathsf{E}\{\mathbf{b}_{n-p} {\mathbf{b}_{n-q}}^{\mathsf{T}}\} {\mathbf{W}_q}^{\mathsf{T}} = \sum_{p,q} \mathbf{W}_p \boldsymbol{\Lambda} \delta_{p-q} {\mathbf{W}_q}^{\mathsf{T}} = \sum_p \mathbf{W}_p \boldsymbol{\Lambda} {\mathbf{W}_p}^{\mathsf{T}}.$$

The global noise power on the components of $\mathbf{x}'_n$ is

$$\mathsf{E}\{\|\mathbf{v}_n\|_2^2\} = \sum_p \mathrm{tr}(\mathbf{W}_p \boldsymbol{\Lambda} {\mathbf{W}_p}^{\mathsf{T}}) = \mathrm{tr}\left(\begin{bmatrix} \mathbf{W}_{-P+1} & \dots & \mathbf{W}_0 \end{bmatrix} \boldsymbol{\Lambda}' \begin{bmatrix} {\mathbf{W}_{-P+1}}^{\mathsf{T}} \\ \vdots \\ {\mathbf{W}_0}^{\mathsf{T}} \end{bmatrix} \right)$$

where $\boldsymbol{\Lambda}'$ is the block-diagonal matrix of size $(PK) \times (PK)$ given by

$$\boldsymbol{\Lambda}' = \begin{bmatrix} \boldsymbol{\Lambda} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{0} & \boldsymbol{\Lambda} \end{bmatrix}.$$

By defining

$$\forall i \in \{0, \dots, N-1\}, \qquad \mathbf{w}'_i = (\boldsymbol{\Lambda}')^{1/2} \mathbf{w}_i$$

we obtain

$$\begin{aligned} \mathsf{E}\{\|\mathbf{v}_n\|_2^2\} &= \mathrm{tr}\left(\begin{bmatrix} {\mathbf{w}'_0}^{\mathsf{T}} \\ \vdots \\ {\mathbf{w}'_{N-1}}^{\mathsf{T}} \end{bmatrix} \begin{bmatrix} \mathbf{w}'_0 & \dots & \mathbf{w}'_{N-1} \end{bmatrix} \right) = \mathrm{tr}\begin{pmatrix} {\mathbf{w}'_0}^{\mathsf{T}}\mathbf{w}'_0 & \dots & {\mathbf{w}'_0}^{\mathsf{T}}\mathbf{w}'_{N-1} \\ \vdots & & \vdots \\ {\mathbf{w}'_{N-1}}^{\mathsf{T}}\mathbf{w}'_0 & \dots & {\mathbf{w}'_{N-1}}^{\mathsf{T}}\mathbf{w}'_{N-1} \end{pmatrix} \\ &= \sum_{i=0}^{N-1} \|\mathbf{w}'_i\|_2^2\ . \end{aligned}$$

This shows that the power of the output noise is minimum if, for all i, $\|\mathbf{w}'_i\|_2$ is minimum. By noting that Eq. (2.33) can be rewritten as

$$\forall i \in \{0, \ldots, N-1\}, \qquad \overline{\boldsymbol{\mathcal{M}}}(\boldsymbol{\Lambda}')^{-1/2}\mathbf{w}'_i = \mathbf{u}_i$$

the optimal solution is

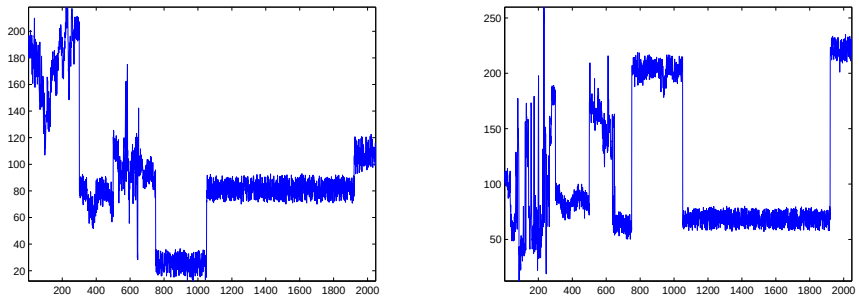
$$\begin{aligned}\mathbf{w}'_i &= (\overline{\boldsymbol{\mathcal{M}}}(\boldsymbol{\Lambda}')^{-1/2})^{\sharp}\,\mathbf{u}_i \\ &= (\overline{\boldsymbol{\mathcal{M}}}(\boldsymbol{\Lambda}')^{-1/2})^{\mathsf{T}}(\overline{\boldsymbol{\mathcal{M}}}(\boldsymbol{\Lambda}')^{-1/2}(\overline{\boldsymbol{\mathcal{M}}}(\boldsymbol{\Lambda}')^{-1/2})^{\mathsf{T}})^{-1},\end{aligned}$$

$\forall i \in \{0, \ldots, N-1\}$. We deduce that the inverse system minimizing the effect of the quantization error is

$$\boldsymbol{\mathcal{W}} = (\boldsymbol{\Lambda}')^{-1/2}\,\overline{\boldsymbol{\mathcal{M}}}^{\mathsf{T}}(\overline{\boldsymbol{\mathcal{M}}}(\boldsymbol{\Lambda}')^{-1}\overline{\boldsymbol{\mathcal{M}}}^{\mathsf{T}})^{-1}\,\boldsymbol{\mathcal{U}},$$

which is equivalent to Eq. (2.35) if $\boldsymbol{\Lambda}' = \mathbf{I}_{KP\times KP}$.

Thus, a particular case of interest for us is the quantization of all the coefficient sequences with the same precision, in which case the covariance matrix of the quantization noise can be written as $\boldsymbol{\Lambda} = \sigma^2\mathbf{I}_{K\times K}$, even though the polyphase coefficients are correlated. When the quantization noise variance equals one, the above mentioned condition for $\boldsymbol{\Lambda}'$ is satisfied.



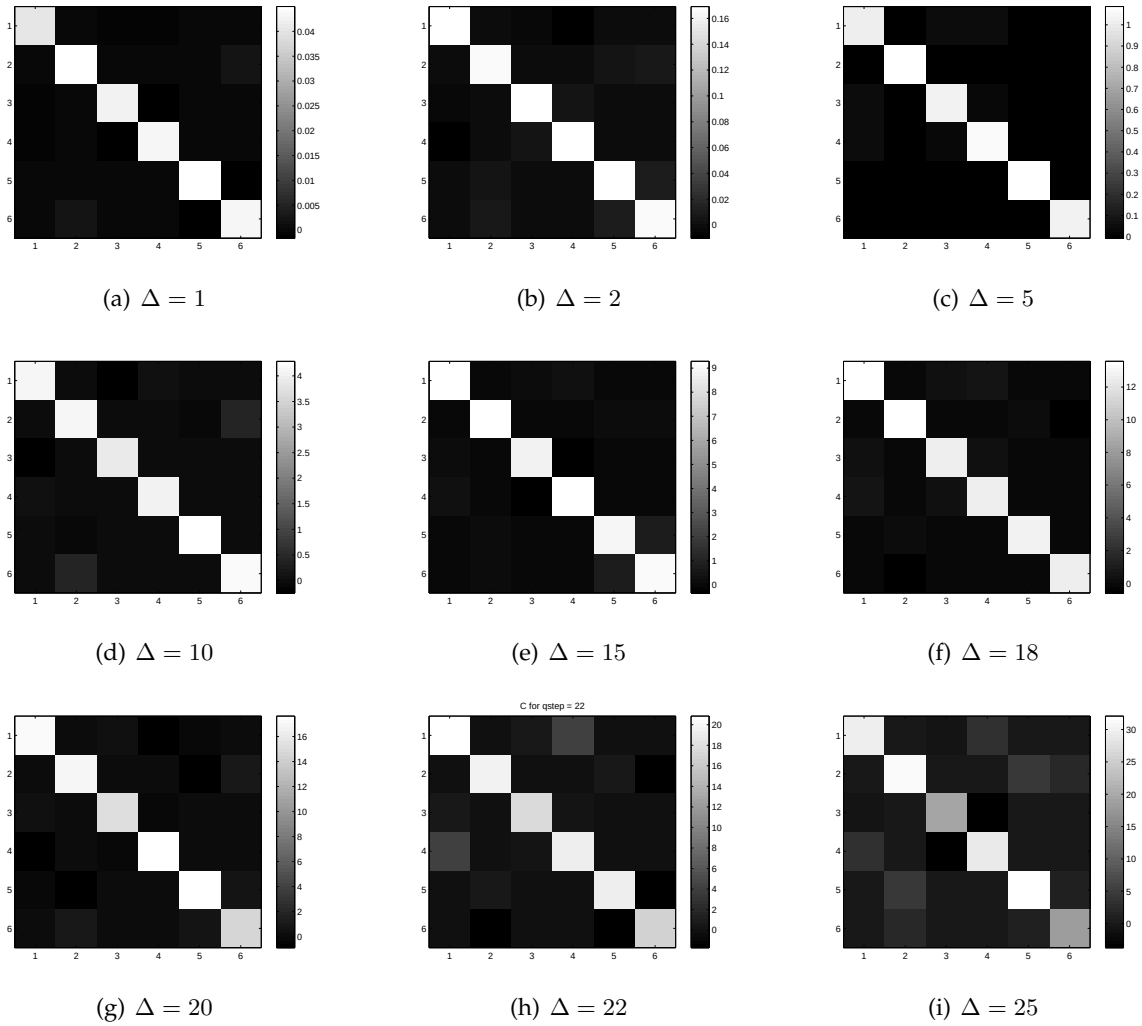
(a) Spatial pixel 60 (b) Spatial pixel 100

FIGURE 2.2: Test temporal signals formed by concatenation of 7 classical QCIF test videos, given here for two spatial positions in the video frame

Let us then verify the diagonality of the covariance matrix, $\boldsymbol{\Lambda}$ on a sample of 100 temporal signals obtained in as many different spatial positions of a concatenated sequence of $QCIF$ classical test videos (such as Foreman, Crew, Hall-Monitor, Mobile, etc..., in YUV color format). We have chosen such signals in order to comply with the application framework proposed at the end of this chapter aiming the multiple description transmission of a video sequence. This corresponds to the case without motion estimation, which, as it will be explained later, presents an interest in the MDC scenario for computation and complexity reasons. Two of those signals are illustrated in Figure 2.2.

We have computed the covariance matrix of the resulting quantization noise in each of the six coefficient sets forming the scheme **MD3**, for instance[6]. In Figure 2.3 we present the covariance matrix corresponding to the whole set of considered temporal signals at different quantization steps denoted here by Δ. We have obtained the wavelet coefficients with the Haar

[6]Recall that the scheme **MD3** has the following wavelet sequences at the coarsest resolution: $\{\hat{a}_n^{\mathrm{I}},\ \check{a}_n^{\mathrm{I}},\ \hat{d}_n^{\mathrm{I}},\ \hat{a}_n^{\mathrm{II}},\ \check{a}_n^{\mathrm{II}},\ \check{d}_n^{\mathrm{II}}\}$.



(a) $\Delta = 1$ (b) $\Delta = 2$ (c) $\Delta = 5$

(d) $\Delta = 10$ (e) $\Delta = 15$ (f) $\Delta = 18$

(g) $\Delta = 20$ (h) $\Delta = 22$ (i) $\Delta = 25$

FIGURE 2.3: Covariance matrix for the quantization noise at different quantization steps Δ.

filter bank. Similar results were observed for the $5/3$ filter bank. We can thus observe that our hypothesis stands at sufficiently low quantization steps.

It is obvious that this result does not hold when the redundancy comes from the mere repetition of some coefficients (**R**-scheme, for instance). In this case, the optimal choice reduces to Eq. (2.35).

2.4.3 Practical examples

In the following we shall present two practical examples that use Haar and biorthogonal $5/3$ filter banks, which are the common choice for wavelet video codecs - the application aimed in this chapter. Let us study the invertibility of the MD3-scheme, for instance. We have chosen this scheme here mainly for its additional subsampling diversity, but the following results can be easily adapted to the other proposed MD schemes.

We begin by evaluating the perfect reconstruction of the corresponding filter bank and after-

wards we shall compute a left inverse for the analysis polyphase matrix, following the previous theoretical considerations.

After computing the pseudoinverses for Haar and biorthogonal 5/3 filter banks, in this chosen example, we shall draw some conclusions about our frame-based schemes with respect to the robustness to quantization noise (as before) when the maximal degree of the reconstruction polynomials in the synthesis matrix varies.

2.4.3.1 Haar filter banks

Perfect Reconstruction

Let us now investigate the perfect reconstruction of the $MD3$-scheme, based on oversampled Haar filter banks. Recall that the z-transforms of the Haar low-pass and high-pass filters are given by

$$H(z) = \frac{1+z}{\sqrt{2}}, \qquad G(z) = \frac{-1+z}{\sqrt{2}}.$$

The polyphase components of $H(z)$ and $G(z)$ are thus

$$\begin{cases} H_0(z) = \dfrac{1}{\sqrt{2}} \\ H_1(z) = \dfrac{1}{\sqrt{2}} \\ H_2(z) = 0 \\ H_3(z) = 0 \end{cases}, \qquad \begin{cases} G_0(z) = -\dfrac{1}{\sqrt{2}} \\ G_1(z) = \dfrac{1}{\sqrt{2}} \\ G_2(z) = 0 \\ G_3(z) = 0 \end{cases}.$$

By substituting these values in

$$\overline{\mathbf{M}}(z) = \begin{bmatrix} H_0(z) & H_1(z) & H_2(z) & H_3(z) \\ H_2(z) & H_3(z) & H_0(z)z^{-1} & H_1(z)z^{-1} \\ H_1(z) & H_2(z) & H_3(z) & H_0(z)z^{-1} \\ H_3(z) & H_0(z)z^{-1} & H_1(z)z^{-1} & H_2(z)z^{-1} \\ G_0(z) & G_1(z) & G_2(z) & G_3(z) \\ G_3(z) & G_0(z)z^{-1} & G_1(z)z^{-1} & G_2(z)z^{-1} \end{bmatrix}$$

we obtain the following polyphase matrix:

$$\overline{\mathbf{M}}(z) = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & z^{-1} & z^{-1} \\ 1 & 0 & 0 & z^{-1} \\ 0 & z^{-1} & z^{-1} & 0 \\ -1 & 1 & 0 & 0 \\ 0 & -z^{-1} & z^{-1} & 0 \end{bmatrix}. \tag{2.36}$$

As explained in the previous subsection, in order to study the perfect reconstruction properties of the corresponding oversampled filter bank, we have to examine the maximal order minors of $\overline{\mathbf{M}}(z)$. Table 2.1 presents all the 4^{th} order minors for the different possible 4×4 sub-matrices of $\sqrt{2}\,\overline{\mathbf{M}}(z)$. The first column of the table indicates the lines of $\overline{\mathbf{M}}(z)$ which have been selected to build each sub-matrix and the third one gives the moduli of the zeros of the associated determinant. As several minors are monomial, we conclude that $\overline{\mathbf{M}}(z)$ is left-invertible by

Lines	Minor	Moduli of the zeros
1 2 3 4	$z^{-2}(z^{-1}-1)$	1
1 2 3 5	$2z^{-2}$	None
1 2 3 6	$-z^{-2}(1+z^{-1})$	1
1 2 4 5	$-2z^{-2}$	None
1 2 4 6	$2z^{-3}$	None
1 2 5 6	$2z^{-2}$	None
1 3 4 5	$-2z^{-2}$	None
1 3 4 6	$2z^{-3}$	None
1 3 5 6	$2z^{-2}$	None
1 4 5 6	0	-
2 3 4 5	$z^{-2}(z^{-1}+1)$	1
2 3 4 6	$-2z^{-3}$	None
2 3 5 6	$z^{-2}(z^{-1}-1)$	1
2 4 5 6	$-2z^{-3}$	None
3 4 5 6	$-2z^{-3}$	None

TABLE 2.1: Maximal order minors of $\sqrt{2}\,\overline{\mathbf{M}}(z)$ for Haar filters.

a polynomial matrix. Each monomial minor corresponds to sequences of coefficients from which the original signal can be reconstructed by an FIR filter bank. For example, the second line of the table shows that an FIR MIMO filter allows us to compute $(x_n)_{n\in\mathbb{Z}}$ from $(\hat{a}^{\mathrm{I}}_n)_{n\in\mathbb{Z}}$, $(\check{a}^{\mathrm{I}}_n)_{n\in\mathbb{Z}}$, $(\hat{a}^{\mathrm{II}}_n)_{n\in\mathbb{Z}}$ and $(\hat{d}^{\mathrm{I}}_n)_{n\in\mathbb{Z}}$. (The corresponding synthesis filter bank is easily determined by inverting the corresponding polynomial submatrix.) On the contrary, some minors have a zero on the unit circle showing that the corresponding sub-matrices are not invertible. Thus, the first line of the table shows that it is not possible to reconstruct $(x_n)_{n\in\mathbb{Z}}$ from the only knowledge of $(\hat{a}^{\mathrm{I}}_n)_{n\in\mathbb{Z}}$, $(\check{a}^{\mathrm{I}}_n)_{n\in\mathbb{Z}}$, $(\hat{a}^{\mathrm{II}}_n)_{n\in\mathbb{Z}}$ and $(\check{a}^{\mathrm{II}})_{n\in\mathbb{Z}}$. This is consistent with the fact that the detail coefficients cannot be completely discarded since they carry useful high-frequency information.

Haar Polyphase Matrix Inversion

Based on the polyphase matrix determined in Eq. (2.36), and the above considerations, we can find a polynomial left inverse for $\overline{\mathbf{M}}(z)$. In this case, we have $Q = 2$. Thus, Eq. (2.27) becomes:

$$\overline{\mathbf{M}}(z) = \overline{\mathbf{M}}_0 + \overline{\mathbf{M}}_1 z^{-1}$$

where

$$\overline{\mathbf{M}}_0 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

and

$$\overline{\mathbf{M}}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 \end{bmatrix}.$$

It can be checked that the minimal value of P for which $\mathrm{rank}([\overline{\boldsymbol{\mathcal{M}}}\,\boldsymbol{\mathcal{U}}]) = \mathrm{rank}(\overline{\boldsymbol{\mathcal{M}}})$ is equal to 2. Then, the block matrix $\boldsymbol{\mathcal{M}}$ in Eq. (2.30) reduces to

$$\overline{\boldsymbol{\mathcal{M}}} = \begin{bmatrix} \overline{\mathbf{M}}_0{}^{\mathsf{T}} & \mathbf{0}_{4\times 6} \\ \overline{\mathbf{M}}_1{}^{\mathsf{T}} & \overline{\mathbf{M}}_0{}^{\mathsf{T}} \\ \mathbf{0}_{4\times 6} & \overline{\mathbf{M}}_1{}^{\mathsf{T}} \end{bmatrix}.$$

Computing the pseudo-inverse of $\overline{\boldsymbol{\mathcal{M}}}$ and using Eq. 2.26 yields:

$$\mathbf{W}(z) = \sqrt{2} \begin{bmatrix} \frac{5}{12} & -\frac{1}{6} & \frac{1}{6} & \frac{1}{12} & -\frac{5}{12} & \frac{1}{12} \\ \frac{1}{4} & 0 & 0 & \frac{1}{4}z & \frac{1}{4} & -\frac{1}{4}z \\ \frac{1}{12}z & \frac{1}{6}z & -\frac{1}{6}z & \frac{5}{12}z & -\frac{1}{12}z & \frac{5}{12}z \\ -\frac{1}{4}z & \frac{1}{2}z & \frac{1}{2}z & -\frac{1}{4}z & \frac{1}{4}z & -\frac{1}{4}z \end{bmatrix}. \tag{2.37}$$

We shall now present the same study for biorthogonal 5/3 filter banks.

2.4.3.2 Biorthogonal 5/3 Filters

Perfect reconstruction

The z-transform of the biorthogonal $5/3$ filter pair is the following:

$$\begin{aligned} H(z) &= \gamma(-p_1u_1z^{-2} + u_1z^{-1} + 1 - 2p_1u_1 + u_1z - p_1u_1z^2) \\ \bar{H}(z) &= \frac{1}{\gamma}(p_1z^{-1} + 1 + p_1z), \end{aligned}$$

where $H(z)$ and $\bar{H}(z)$ are the transfer functions of the low-pass analysis and synthesis filters and γ, p_1 and u_1 are some nonzero real constants. The usual values for these constants are: $\gamma = \sqrt{2}$, $p_1 = 1/2$ and $u_1 = 1/4$.

Let $(\bar{h}_n)_{n\in\mathbb{Z}}$ be the impulse response of the low-pass synthesis filter. From the relations between the analysis and synthesis filters [VK95]:

$$g_n = (-1)^{n+1}\bar{h}_{n+1},$$

we obtain

$$G(z) = \sum_n (-1)^{n+1}\bar{h}_{n+1}z^{-n} = z\bar{H}(-z) = \frac{1}{\gamma}(-p_1 + z - p_1z^2).$$

The polyphase components of $H(z)$ and $G(z)$ are then given by:

$$\begin{cases} H_0(z) = \gamma(1 - 2p_1u_1) \\ H_1(z) = \gamma u_1 \\ H_2(z) = -\gamma p_1 u_1 (1 + z^{-1}) \\ H_3(z) = \gamma u_1 z^{-1} \end{cases}, \qquad \begin{cases} G_0(z) = -\gamma^{-1} p_1 \\ G_1(z) = \gamma^{-1} \\ G_2(z) = -\gamma^{-1} p_1 \\ G_3(z) = 0, \end{cases}$$

and they lead to the following polyphase matrix:

$$\overline{\mathbf{M}}(z) = \begin{bmatrix} \gamma(1-2p_1u_1) & \gamma u_1 & -\gamma p_1 u_1(1+z^{-1}) & \gamma u_1 z^{-1} \\ -\gamma p_1 u_1(1+z^{-1}) & \gamma u_1 z^{-1} & \gamma(1-2p_1u_1)z^{-1} & \gamma u_1 z^{-1} \\ \gamma u_1 & -\gamma p_1 u_1(1+z^{-1}) & \gamma u_1 z^{-1} & \gamma(1-2p_1u_1)z^{-1} \\ \gamma u_1 z^{-1} & \gamma(1-2p_1u_1)z^{-1} & \gamma u_1 z^{-1} & -\gamma p_1 u_1 z^{-1}(1+z^{-1}) \\ -\gamma^{-1}p_1 & \gamma^{-1} & -\gamma^{-1}p_1 & 0 \\ 0 & -\gamma^{-1}p_1 z^{-1} & \gamma^{-1}z^{-1} & -\gamma^{-1}p_1 z^{-1} \end{bmatrix}. \tag{2.38}$$

Note that for the calculation of the zeros of the minors, the constant γ or γ^{-1} weighting each line of the matrix has no effect. Therefore, we can focus our attention on the matrix:

$$\begin{aligned} \widetilde{\overline{\mathbf{M}}}(z) &= \mathrm{Diag}(\gamma^{-1}, \ldots, \gamma^{-1}, \gamma, \gamma)\overline{\mathbf{M}}(z) \\ &= \begin{bmatrix} 1-2p_1u_1 & u_1 & -p_1u_1(1+z^{-1}) & u_1 z^{-1} \\ -p_1u_1(1+z^{-1}) & u_1 z^{-1} & (1-2p_1u_1)z^{-1} & u_1 z^{-1} \\ u_1 & -p_1u_1(1+z^{-1}) & u_1 z^{-1} & (1-2p_1u_1)z^{-1} \\ u_1 z^{-1} & (1-2p_1u_1)z^{-1} & u_1 z^{-1} & -p_1u_1 z^{-1}(1+z^{-1}) \\ -p_1 & 1 & -p_1 & 0 \\ 0 & -p_1 z^{-1} & z^{-1} & -p_1 z^{-1} \end{bmatrix}. \end{aligned} \tag{2.39}$$

By considering the values of the constants p_1 and u_1, we obtain the results presented in Table 2.2. As for Haar filters, we specify the lines of matrix $\overline{\mathbf{M}}(z)$ which have been used to calculate each maximal order minor as well as the moduli of the zeros of the minor.

For the lines $1, 4, 10, 12$ and 13 of the table, a zero on the unit circle is observed. Therefore, we cannot invert the respective sub-systems. Each other minor corresponds to a critically sampled filter bank which is invertible, but with an IIR synthesis filter bank (since the minor does not reduce to a monomial). As there is no common zero between all the maximal minors, we can claim that a MIMO FIR synthesis filter bank allowing us to reconstruct $(x_n)_{n\in\mathbb{Z}}$ from the complete set of coefficients.

Biorthogonal 5/3 Polyphase Matrix Inversion

Similarly to the case of Haar filters, we want to calculate a polynomial left inverse of the polyphase matrix $\overline{\mathbf{M}}(z)$ for biorthogonal $5/3$ filters. Now, we have $Q = 3$ and, for symmetry reasons that will be best apparent when presenting the design of the $5/3$-based decoders, $\overline{\mathbf{M}}(z)$ can be written as:

$$\widetilde{\overline{\mathbf{M}}}(z) = \sum_{q\in\mathbb{Q}} \overline{\mathbf{M}}_q \, z^{-q}.$$

where $\mathbb{Q} = \{-1, 0, 1\}$. This is possible by taking $\widetilde{\overline{\mathbf{M}}}(z) = \mathrm{Diag}(1, z, z, z, 1, z) \cdot \widetilde{\overline{\mathbf{M}}}(z)$ in Eq. (2.39). The real matrices $\overline{\mathbf{M}}_q$, $q \in \{-1, 0, 1\}$ involved in $\widetilde{\overline{\mathbf{M}}}(z)$ are the explicitly the following:

Lines	Minors	Moduli of the zeros
1 2 3 4	$\frac{z^{-3}}{4096}(z-1)^2(1-194\,z^{-1}+z^{-2})$	5.155×10^{-3}, 1, 1, 194.0
1 2 3 5	$\frac{z^{-1}}{64}(1+30\,z^{-1}+z^{-2})$	3.337×10^{-2}, 29.97
1 2 3 6	$\frac{z^{-1}}{1024}(1-123\,z^{-1}-397\,z^{-2}+7z^{-3})$	1.753×10^{-2}, 3.16, 126.15
1 2 4 5	$-\frac{z^{-2}}{4}(1+z^{-1})$	1
1 2 4 6	$-\frac{z^{-2}}{1024}(7-397\,z^{-1}-123\,z^{-2}+z^{-3})$	7.927×10^{-3}, 3.160×10^{-1}, 57.02
1 2 5 6	$\frac{z^{-2}}{16}(9-z^{-1})$	1.111×10^{-1}
1 3 4 5	$\frac{z^{-1}}{1024}(7-397\,z^{-1}-123\,z^{-2}+z^{-3})$	7.927×10^{-3}, 3.160×10^{-1}, 57.02
1 3 4 6	$\frac{z^{-2}}{64}(1+30z^{-1}+z^{-2})$	3.337×10^{-2}, 29.97
1 3 5 6	$-\frac{z^{-1}}{256}(7-142z^{-1}+7z^{-2})$	4.941×10^{-2}, 20.24
1 4 5 6	$\frac{z^{-3}}{256}(z-1)(49-z^{-1})$	2.041×10^{-2}, 1
2 3 4 5	$-\frac{z^{-1}}{1024}(1-123\,z^{-1}-397\,z^{-2}+7\,z^{-3})$	1.754×10^{-2}, 3.164, 126.15
2 3 4 6	$-\frac{z^{-3}}{4}(1+z^{-1})$	1
2 3 5 6	$\frac{z^{-2}}{256}(1-49z^{-1})$	1, 49
2 4 5 6	$-\frac{z^{-2}}{256}(7+114z^{-1}+7z^{-2})$	6.164×10^{-2}, 16.22
3 4 5 6	$\frac{z^{-2}}{16}(1-9z^{-1})$	9

TABLE 2.2: Maximal order minors of $\widetilde{\overline{\mathbf{M}}}(z)$ for biorthogonal $5/3$ filters.

$$\overline{\mathbf{M}}_{-1}=\begin{bmatrix}0&0&-p_1u_1&u_1\\0&0&0&0\\0&0&0&0\\0&0&0&-p_1u_1\\0&0&0&0\\0&0&0&0\end{bmatrix}$$

$$\overline{\mathbf{M}}_{0}=\begin{bmatrix}1-2p_1u_1&u_1&-p_1u_1&0\\-p_1u_1&u_1&1-2p_1u_1&u_1\\0&-p_1u_1&u_1&1-2p_1u_1\\u_1&1-2p_1u_1&u_1&-p_1u_1\\-p_1&1&-p1&0\\0&-p_1&1&-p_1\end{bmatrix}$$

and

$$\overline{\mathbf{M}}_{1}=\begin{bmatrix}0&0&0&0\\-p_1u_1&0&0&0\\u_1&-p_1u_1&0&0\\0&0&0&0\\0&0&0&0\\0&0&0&0\end{bmatrix}.$$

The pseudoinverse of $\overline{\boldsymbol{\mathcal{M}}}$, computed for $P=3$, gives $\boldsymbol{\mathcal{W}}$ which contains the blocks $\boldsymbol{W}_{-1},\ \boldsymbol{W}_0,\ \boldsymbol{W}_1$ such that:

$$\boldsymbol{W}(z)=\boldsymbol{W}_{-1}\,z+\boldsymbol{W}_0+\boldsymbol{W}_1\,z^{-1}$$

We thus obtain the inverse transfer matrix:

$$\boldsymbol{W}(z) = \begin{bmatrix} 0.7567 - 0.07465\,z & -0.2034 + 0.1081\,z^{-1} & 0.1396 - 0.2211\,z^{-1} & 0.1819 + 0.01995\,z & \dots \\ 0.02789 + 0.05653\,z & 0.2015 + 0.003985\,z^{-1} & -0.08500 + 0.05633\,z^{-1} & 0.4580 - 0.01214\,z & \dots \\ -0.08500 + 0.05633\,z & 0.4580 - 0.01214\,z^{-1} & 0.02789 + 0.05653\,z^{-1} & 0.2015 + 0.003985\,z & \dots \\ 0.1396 - 0.2211\,z & 0.1819 + 0.01995\,z^{-1} & 0.7567 - 0.07465\,z^{-1} & -0.2034 + 0.1081\,z & \dots \end{bmatrix}$$

$$\begin{bmatrix} \dots & -0.5499 + 0.03732\,z - 0.05405\,z^{-1} & 0.03988 - 0.009973\,z + 0.1105\,z^{-1} \\ \dots & 0.4953 - 0.02827\,z - 0.001992\,z^{-1} & -0.2199 + 0.006071\,z - 0.02816\,z^{-1} \\ \dots & -0.2199 - 0.02816\,z + 0.006071\,z^{-1} & 0.4953 - 0.001992\,z - 0.02827\,z^{-1} \\ \dots & 0.03988 + 0.1105\,z - 0.009973\,z^{-1} & -0.5499 - 0.05405\,z + 0.03732\,z^{-1} \end{bmatrix}$$

2.4.4 Observations on the proposed MDC schemes

At this point some observations are necessary in order to assess the reconstruction performances of the different proposed MDC schemes. For the two categories of analysis filters (Haar and biorthogonal $5/3$), we have investigated the invertibility of the resulting polyphase analysis matrix $\overline{\boldsymbol{M}}(z)$ in each of the MDC situations previously presented. In other words, we have computed $\mathcal{W}$ (as in Eq. (2.35)) leading to the synthesis matrix $\boldsymbol{W}(z)$. For each $\mathcal{W}$ we have evaluated the mean squared error of the reconstruction, supposing a quantization noise of variance one:

$$\mathrm{MSE_r} = \|\boldsymbol{\mathcal{W}}\|^2 . \tag{2.40}$$

As we have seen in section 2.4.2, the output noise due to quantization has a minimum effect on the reconstruction if, for all i, the Frobenius norm of the i^{th} column of $\mathcal{W}$ is the smallest, and thus the reconstruction error is given by Eq. (2.40)[7].

We have carried out some numerical tests computing the $\mathrm{MSE_r}$ for our MDC schemes, when varying the maximal order of the reconstruction filters, P and choosing the minimum norm $\mathcal{W}$ obtained with the couple of bounds (p_1, p_2), such that $P = p_1 + p_2 + 1$ (as before). These results are summarized in Table 2.3 for the case of Haar filter banks. Note that, from the transfer matrix point of view, the MSE_r computed for the **R**-scheme is the same as in the Single Description case, since this scheme merely repeats coefficients from the same wavelet decomposition. Therefore, these results are the reference for comparison between multiple and single description coding modes.

Scheme	$\mathrm{MSE_r}$ for Haar FB Synthesis order (P)		
	2	3	5
R-scheme	**4**	4	4
MD1-scheme	**3**	3	3
MD2-scheme	**3.66**	3.66	3.66
MD3-scheme	**3.66**	3.66	3.66

TABLE 2.3: Evolution of the $\mathrm{MSE_r}$ with the order of the reconstruction filters for the MDC schemes.

[7] The "r" in $\mathrm{MSE_r}$ stands for "reconstruction".

Biorthogonal $5/3$ filters									
Schemes		Analysis: $(q_1, q_2) = (0, 2)$ Synthesis order (P)				Analysis: $(q_1, q_2) = (1, 1)$ Synthesis order (P)			
		2	3	4	5	2	3	4	5
R-scheme	$\mathrm{MSE_r}$	4.37	4.37	4.37	4.37	4.37	**4.37**	4.37	4.37
	$(p_1,\, p_2)$	$(0,1)$	$(1,1)$	$(1,2)$	$(1,3)$	$(0,1)$	$(1,1)$	$(1,2)$	$(1,3)$
MD1-scheme	$\mathrm{MSE_r}$	4.37	3.38	2.58	2.57	95.12	**2.65**	2.59	2.57
	$(p_1,\, p_2)$	$(0,1)$	$(1,1)$	$(1,2)$	$(2,2)$	$(1,0)$	$(1,1)$	$(2,1)$	$(2,2)$
MD2-scheme	$\mathrm{MSE_r}$	14.60	3.20	2.94	2.93	14.60	**3.20**	2.94	2.93
	$(p_1,\, p_2)$	$(0,1)$	$(1,1)$	$(1,2)$	$(2,2)$	$(0,1)$	$(1,1)$	$(1,2)$	$(2,2)$
MD3-scheme	$\mathrm{MSE_r}$	9.69	5.70	2.99	2.96	65.44	**3.23**	3.08	2.93
	$(p_1,\, p_2)$	$(0,1)$	$(0,2)$	$(1,2)$	$(1,3)$	$(0,1)$	$(1,1)$	$(2,1)$	$(2,2)$

TABLE 2.4: Evolution of the $\mathrm{MSE_r}$ with the maximal order of the reconstruction filters for the MDC schemes and the SDC scheme given by a non-redundant decomposition based on $\mathcal{B}^{\mathrm{I}}$, using biorthogonal $5/3$ analysis filters.

Another remark concerning the Haar filter banks is that if the analysis filters are supposed to be causal, then the inverse filters will be anti-causal and conversely. In other words, the results given in Table 2.3 stand for $(q_1,\, q_2) = (0,\, 1)$ and $(p_1,\, p_2) = (1,\, 0)$ in Eqs. (2.27) and (2.26), respectively (as presented in Section 2.4.3.1). From this table it can be seen that increasing the order P of the synthesis filters does not influence the reconstruction performances of this filter bank.

When considering biorthogonal $5/3$ filter banks, however, the maximal order of the analysis filters is $Q = 3$ which leaves more degrees of freedom in choosing the pair $(q_1,\, q_2)$ summing up to Q. We have considered two cases of interest: $(q_1,\, q_2) = (0, 2)$ (in other words causal analysis filters) and $(q_1,\, q_2) = (1, 1)$ (as in the biorthogonal example given in Section 2.4.3.2). After setting the analysis pair of bounds $(q_1,\, q_2)$ in this scenario, we have searched for the pair $(p_1,\, p_2)$ leading to the minimal norm of $\mathcal{W}$. Table 2.4 summarizes the results for the evolution of $\mathrm{MSE_r}$ with the variation of P, for the optimal pair $(p_1,\, p_2)$ such that $p_1 + p_2 + 1 = P$.

These tables (2.3 and 2.4) show that the best performances at the central decoder are to be expected from the scheme **MD1** whereas the schemes **MD2** and **MD3** exhibit comparable performances for a good choice of (p_1, p_2) bounds in the synthesis filter maximum order. If one wants to judge the performances of Haar versus biorthogonal $5/3$ filter banks, the **R**-scheme seems to be worse in the latter case, whereas the others are generally better than the Haar filter-bank schemes. This first (counter-intuitive) result concerning the **R**-scheme could be explained by the fact that here we did not take into account the energy compaction properties of the different bases. Indeed, if one of the filter banks leads to a more compact representation of the signal, there will be less significant coefficients, thus one could expect the possibility of using a coarser quantization step, leading to better compression performances. This reasoning however does not necessarily apply when comparing configurations using the same filters.

Another remark about the two considered filter banks is that these results concern solely the central decoders, therefore they do not infer anything about the side decoding performances. This is why, even though the **MD1**-scheme performs better at the central decoder, it is not a good choice in a balanced channels scenario. Indeed, if comparable quality is expected from both side decoders, a better choice would be one of the schemes **MD2** or **MD3** (any of the two

being a good choice).

Note also that the performances of these filter banks in terms of reconstruction error do not improve significantly with the increasing of the synthesis order P. If we take into account the implementation difficulties in a video coding system, than the shortest are the filters the faster and less complex is the implementation. This is the reason behind the choices made in the last section of this chapter - an application to robust video coding. Moreover, Table 2.4 clearly shows that for the **MD3**-scheme (which is our final implementation choice) imposing causality to our filter is too much of a restriction, significantly altering the end-to-end performances.

In the following we shall detail the encoding and decoding issues both for Haar and biorthogonal $5/3$-based schemes, and this chapter will be closed by the mentioned application to video coding.

2.5 Lifting-based design of the Haar MD encoder

The aim of this section is to find equivalent lifting implementations for the previously introduced MDC schemes, allowing for an easier and more flexible implementation of the one-dimensional signal decompositions. This will be further exploited for non-linear motion-compensated temporal decompositions of video sequences.

Let us consider in this section only the implementation of the **MD3**-scheme, knowing that with the corresponding modification in the shifting factor all these considerations stand for the other schemes as well. We also choose this particular scheme because it is the most balanced in terms of the positions of the kept samples in the detail subbands resulting from any of the two wavelet bases, at the coarsest level. This configuration is especially useful when it comes to the side decoding.

2.5.1 2-band lifting approach

Since the first $J-1$ levels in our decompositions are obtained from a usual wavelet analysis, in the following we shall only be interested in the last resolution level.

The corresponding coefficients in the two descriptions are computed as follows:

$$\begin{cases} a_n^{\mathrm{I}} = \sum_k h_{2n-k} x_k & \text{(2.41a)} \\ \hat{d}_n^{\mathrm{I}} = \sum_k g_{4n-k} x_k & \text{(2.41b)} \\ a_n^{\mathrm{II}} = \sum_k h_{2n-1-k} x_k & \text{(2.41c)} \\ \check{d}_n^{\mathrm{II}} = \sum_k g_{4n-3-k} x_k, & \text{(2.41d)} \end{cases}$$

where, for simplicity, we have denoted by x_k the approximation coefficients at the $(J-1)$-th level and we have omitted the subscript J.

A quick and memory efficient implementation for filter banks is given by the so-called lifting scheme, introduced in [HP96] and extended in [HP98] under the name of nonlinear subband

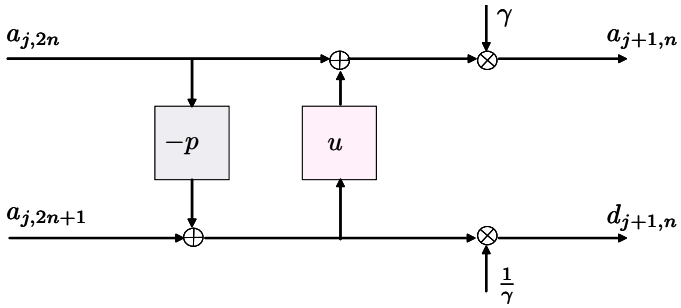


FIGURE 2.4: Basic single-stage lifting scheme.

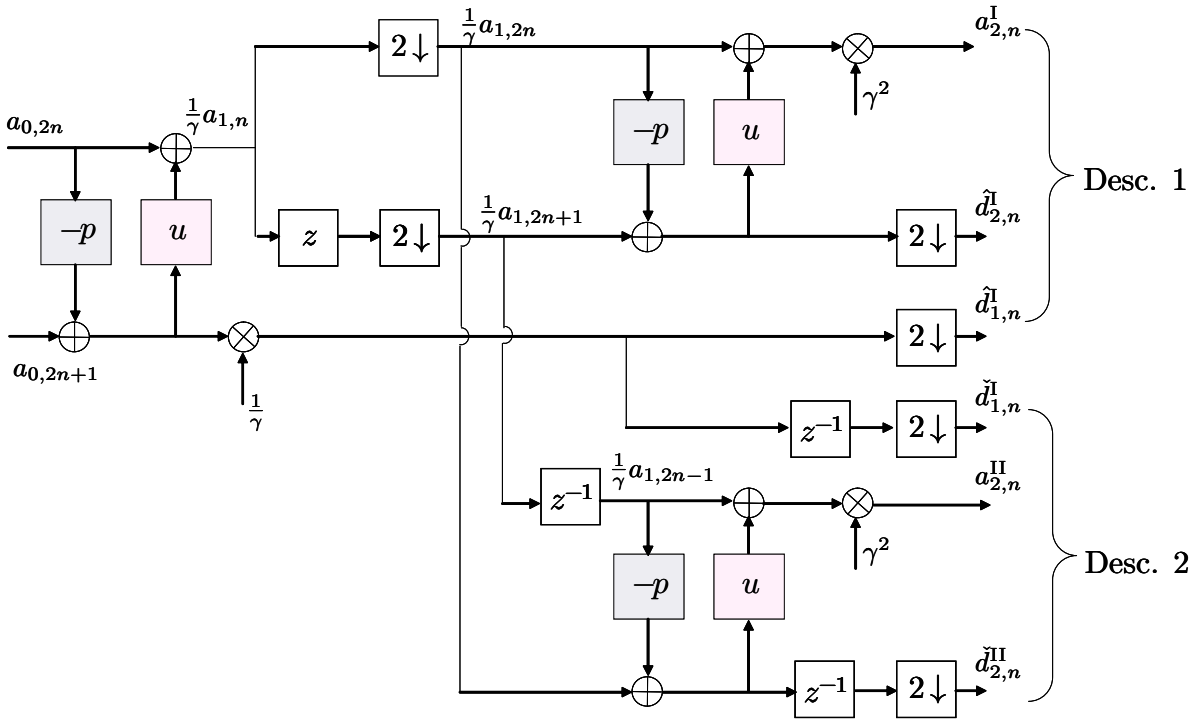


FIGURE 2.5: 2-band lifting implementation of the proposed multiple description coder for the last two resolution levels.

decompositions. The "lifting" term was introduced in the tutorial paper [DS98]. This scheme guarantees the perfect reconstruction of the system. The basic lifting core that we are using when implementing our MDC strategy is given in Fig. 2.4. The p and u operators in the scheme stand for the predict and update, respectively, and γ is a real non-zero multiplicative constant.

Using a one-stage lifting implementation of the filter bank as above, we illustrate our scheme in Fig. 2.5. For readability we shall only display two levels of resolution.

2.5.2 Equivalent 4-band lifting implementation for the Haar filter bank

The 2-band lifting approach presented above does not yield an immediate inversion scheme, in particular when using nonlinear operators, such as those involving motion estima-

tion/compensation in the temporal decomposition of a video sequence. This is the motivation behind searching an equivalent scheme for which global inversion would be easier to prove. In the following we build a simpler equivalent lifting scheme for the Haar filter bank, by using directly the four-band polyphase components of the input signal, instead of the two-band ones. For the first description, the approximation coefficients can be rewritten from Eq. (2.41a), while the detail coefficients are still obtained with Eq. (2.41b), leading to:

$$\begin{cases} \hat{a}_n^{\mathrm{I}} = a_{2n}^{\mathrm{I}} = \sum_k h_{4n-k} x_k \\ \check{a}_n^{\mathrm{I}} = a_{2n-1}^{\mathrm{I}} = \sum_k h_{4n-2-k} x_k \\ \hat{d}_n^{\mathrm{I}} = \sum_k g_{4n-k} x_k. \end{cases} \tag{2.42}$$

Similarly, for the second description, we express the approximation subband from Eq. (2.41c) and keep the details from Eq. (2.41d):

$$\begin{cases} \hat{a}_n^{\mathrm{II}} = \sum_k h_{4n-1-k} x_k \\ \check{a}_n^{\mathrm{II}} = \sum_k h_{4n-3-k} x_k \\ \check{d}_n^{\mathrm{II}} = \sum_k g_{4n-3-k} x_k. \end{cases} \tag{2.43}$$

To go further and find an equivalent scheme for the Haar filter bank, note that the two-band polyphase components of the input signal, $x_{2n} = a_{J-1,2n}$ and $x_{2n+1} = a_{J-1,2n+1}$ are first filtered and then subsampled (see Fig. 2.5). However, for the Haar filter bank, recall that the predict and update operators are respectively $p = \mathbf{I}$ and $u = \frac{1}{2}\mathbf{I}$ (and the constant $\gamma = \sqrt{2}$). Since these are both instantaneous operators, one can reverse the order of the filtering and downsampling operations. This yields the following very simple expressions for the coefficients in the first description:

$$\begin{cases} \hat{a}_n^{\mathrm{I}} = \dfrac{x_{4n} + x_{4n+1}}{\sqrt{2}} = \dfrac{x_n^{(0)} + x_n^{(1)}}{\sqrt{2}} & \text{(2.44a)} \\ \check{a}_n^{\mathrm{I}} = \dfrac{x_{4n-2} + x_{4n-1}}{\sqrt{2}} = \dfrac{x_{n-1}^{(2)} + x_{n-1}^{(3)}}{\sqrt{2}} & \text{(2.44b)} \\ \hat{d}_n^{\mathrm{I}} = \dfrac{x_{4n+1} - x_{4n}}{\sqrt{2}} = \dfrac{x_n^{(1)} - x_n^{(0)}}{\sqrt{2}} & \text{(2.44c)} \end{cases}$$

and in the second:

$$\begin{cases} \hat{a}_n^{\mathrm{II}} = \dfrac{x_{4n} + x_{4n-1}}{\sqrt{2}} = \dfrac{x_n^{(0)} + x_{n-1}^{(3)}}{\sqrt{2}} & \text{(2.45a)} \\ \check{a}_n^{\mathrm{II}} = \dfrac{x_{4n-2} + x_{4n-3}}{\sqrt{2}} = \dfrac{x_{n-1}^{(2)} + x_{n-1}^{(1)}}{\sqrt{2}} & \text{(2.45b)} \\ \check{d}_n^{\mathrm{II}} = \dfrac{x_{4n-2} - x_{4n-3}}{\sqrt{2}} = \dfrac{x_{n-1}^{(2)} - x_{n-1}^{(1)}}{\sqrt{2}}. & \text{(2.45c)} \end{cases}$$

A schematic form of these equations is given in Fig. 2.6.

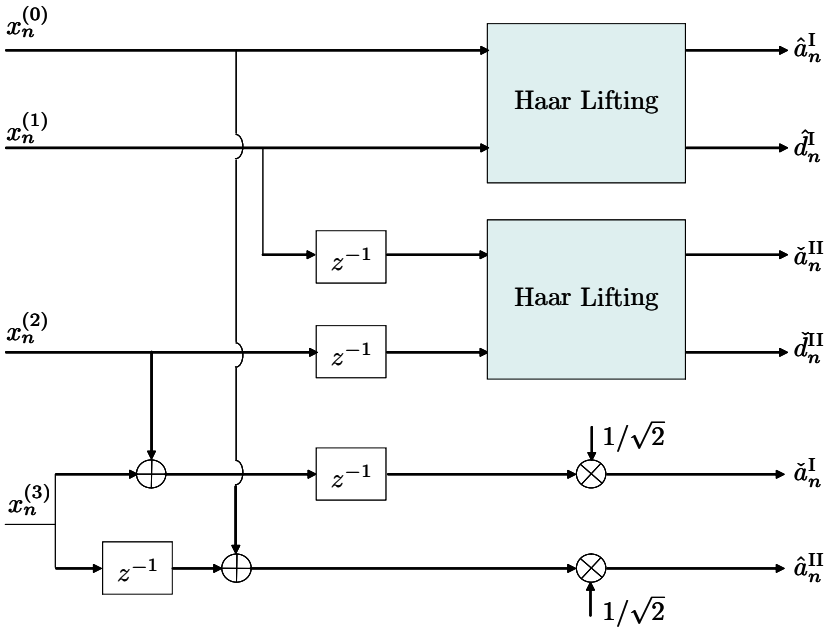


FIGURE 2.6: Redundant 4-band lifting scheme for the Haar filter bank encoder.

2.6 Encoder design for biorthogonal 5/3 filter banks

In the case of biorthogonal 5/3 filter banks, finding an equivalent 4-band lifting scheme as in the case of Haar filters is not possible, therefore we will only use the 2-band lifting scheme presented in the previous section, in order to find the six wavelet coefficient sets needed for the **MD3**-scheme at the coarsest level (see Figure 2.5).

If we replace the lifting parameters $p_1,\ u_1$ and γ in the analysis matrix $M(z)$ given in Section 2.4.3.2 with their values corresponding to the biorthogonal 5/3 filters we obtain the following transfer matrix:

$$\begin{bmatrix} 1.060 & 0.3535 & -0.1768\,z^{-1} - 0.1768 & 0.3535\,z^{-1} \\ -0.1768 - 0.1768\,z & 0.3535 & 1.060 & 0.3535 \\ 0.3535\,z & -0.1768 - 0.1768\,z & 0.3535 & 1.060 \\ 0.3535 & 1.060 & 0.3535 & -0.1768\,z^{-1} - 0.1768 \\ -0.3535 & 0.7070 & -0.3535 & 0.0 \\ 0.0 & -0.3535 & 0.7070 & -0.3535 \end{bmatrix}$$

that shall be further used in the implementation of the biorthogonal 5/3 scheme. The inversion of this matrix will give a decoding strategy for the central decoder whereas for the side decoders one should invert a sub-matrix given by the lines which form each side scheme (e.g. the matrix formed with the lines 1, 2 and 5 for the first side decoder).

2.7 Decoder design

Now that we have presented the actual transform stage serving in encoding the two descriptions, having taken the **MD3**-scheme as an example, let us focus on the corresponding decoding strategies. We are thus presenting in this section the reconstruction of each polyphase component of the input signal both for Haar and biorthogonal 5/3 filter banks.

As explained in the previous chapter, the next stage in the coding scheme is the operation of quantization. In our case we consider that uniform quantization is performed on the wavelet subbands, and we shall indicate this fact by the following notation $[c_n]$, where c stands for a generic wavelet coefficient subband.

In the generic case, our aim is to recover x_n, the input signal, from the subsampled wavelet coefficients corresponding to an MDC scheme at the central decoder, in which case the recovery error is the effect of quantization, and at the side decoders, where half of the subsampled wavelet subbands have been lost. Let us denote the recovered polyphase components of the signal by $\tilde{x}_n^{(i)}$.

In the following we shall obtain these polyphase components from the quantized subbands obtained from Haar and biorthogonal 5/3 filter banks. Then, the next sections will present an application of our MDC strategy to robust video coding and more coding/decoding details will be given for the central and side decoders, by incorporating motion into our schemes. Some structure improvements that lead to better reconstruction will also be presented for the case of Haar filter banks.

2.7.1 Haar decoders

Central decoder

The central decoder for Haar filter banks is obtained from the pseudo-inverse matrix computed in Section 2.4.3.1 in the case of the $MD3$-scheme. Thus, applying the inverse z-transform to Eq. (2.37) leads to the following reconstructed polyphase components of x_n:

$$\tilde{x}_n^{(0)} = \frac{5}{6}\left[\frac{1}{\sqrt{2}}(\hat{a}_n^{\mathrm{I}} - \hat{d}_n^{\mathrm{I}})\right] + \frac{1}{6}\left[\frac{1}{\sqrt{2}}(\check{a}_n^{\mathrm{II}} + \check{d}_n^{\mathrm{II}})\right] - \frac{\sqrt{2}}{6}(\check{a}_n^{\mathrm{I}} - \hat{a}_n^{\mathrm{II}})$$

$$\tilde{x}_n^{(1)} = \frac{1}{2}\left[\frac{1}{\sqrt{2}}(\hat{a}_n^{\mathrm{I}} + \hat{d}_n^{\mathrm{I}}) + \frac{1}{\sqrt{2}}(\check{a}_{n+1}^{\mathrm{II}} - \check{d}_{n+1}^{\mathrm{II}})\right]$$

$$\tilde{x}_n^{(2)} = \frac{1}{6}\left[\frac{1}{\sqrt{2}}(\hat{a}_{n+1}^{\mathrm{I}} - \hat{d}_{n+1}^{\mathrm{I}})\right] + \frac{5}{6}\left[\frac{1}{\sqrt{2}}(\check{a}_{n+1}^{\mathrm{II}} - \check{d}_{n+1}^{\mathrm{II}})\right] + \frac{\sqrt{2}}{6}(\check{a}_{n+1}^{\mathrm{I}} - \hat{a}_{n+1}^{\mathrm{II}})$$

$$\tilde{x}_n^{(3)} = -\frac{1}{2}\left[\frac{1}{\sqrt{2}}(\hat{a}_n^{\mathrm{I}} - \hat{d}_n^{\mathrm{I}}) - \frac{1}{\sqrt{2}}(\check{a}_{n+1}^{\mathrm{II}} - \check{d}_{n+1}^{\mathrm{II}})\right] + \frac{1}{\sqrt{2}}(\check{a}_{n+1}^{\mathrm{I}} + \hat{a}_{n+1}^{\mathrm{II}})$$

Note that most of these components are obtained by directly inverting the basic lifting schemes represented in Fig. 2.6 to which the contribution of the additional information given by the halves of approximation subbands is added. The immediate inversion of the basic lifting core is illustrated in Fig. 2.7.

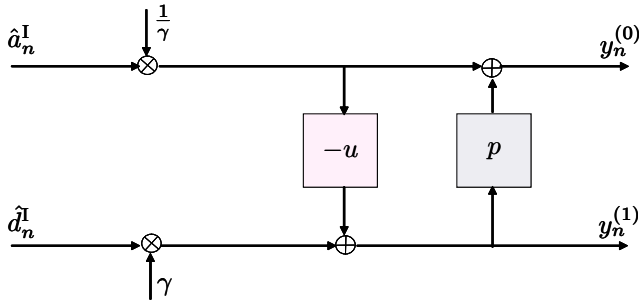


FIGURE 2.7: Inverse lifting scheme for the first description - gives the first two polyphase components of the input signal, reconstructed perfectly in the absence of quantization.

Side decoders

Concerning the side decoders, again from Fig. 2.6, we note that from each description we can partially recover the original sequence by immediate inversion of the scheme. For instance, if we only receive the first description, we can easily reconstruct the polyphase components $x_n^{(0)}$, $x_n^{(1)}$ from the first Haar lifting block. The last two polyphase components $x_n^{(2)}$ and $x_n^{(3)}$ are reconstructed by assuming that they are similar:

$$\tilde{x}_n^{(2)} = \tilde{x}_n^{(3)} = \frac{[\check{a}_{n+1}^{\mathrm{I}}]}{\sqrt{2}}.$$

Similarly, when receiving only the second description, we are able to directly reconstruct $x_n^{(1)}$, $x_n^{(2)}$ from the second Haar lifting block, while $x_n^{(0)}$ and $x_n^{(3)}$ are obtained by duplicating $\hat{a}_{n+1}^{\mathrm{II}}$:

$$\tilde{x}_{n+1}^{(0)} = \tilde{x}_n^{(3)} = \frac{[\hat{a}_{n+1}^{\mathrm{II}}]}{\sqrt{2}}.$$

In the next section we are considering the second example of filter bank formed with biorthogonal 5/3 filters and give the guidelines for the decoders.

2.7.2 Biorthogonal 5/3 decoders

Central decoder

The central decoder is obtained by using the pseudo-inverse computed in Section 2.4.3.2 in order to recover each of the four polyphase components of the input signal. At this point no motion estimation/compensation has been performed at the coarsest level. Therefore, the

recovered vector for the input signal is simply given by:

$$\begin{bmatrix} \tilde{x}_n^{(0)} \\ \tilde{x}_n^{(1)} \\ \tilde{x}_n^{(2)} \\ \tilde{x}_n^{(3)} \end{bmatrix} = \begin{bmatrix} 0.7567 & 0.02789 & -0.08500 & 0.1396 \\ -0.07465 & 0.05653 & 0.05633 & -0.2211 \\ -0.2034 & 0.2015 & 0.4580 & 0.1819 \\ 0.1081 & 0.003985 & -0.01214 & 0.01995 \\ 0.1396 & -0.08500 & 0.02789 & 0.7567 \\ -0.2211 & 0.05633 & 0.05653 & -0.07465 \\ 0.1819 & 0.4580 & 0.2015 & -0.2034 \\ 0.01995 & -0.01214 & 0.003985 & 0.1081 \\ -0.5499 & 0.4953 & -0.2199 & 0.03988 \\ 0.03732 & -0.02827 & -0.02816 & 0.1105 \\ -0.05405 & -0.001992 & 0.006071 & -0.009973 \\ 0.03988 & -0.2199 & 0.4953 & -0.5499 \\ -0.009973 & 0.006071 & -0.001992 & -0.05405 \\ 0.1105 & -0.02816 & -0.02827 & 0.03732 \end{bmatrix}^{\mathsf{T}} \begin{bmatrix} \hat{a}_n^{\mathrm{I}} \\ \hat{a}_{n+1}^{\mathrm{I}} \\ \check{a}_n^{\mathrm{I}} \\ \check{a}_{n-1}^{\mathrm{I}} \\ \hat{d}_n^{\mathrm{I}} \\ \hat{d}_{n-1}^{\mathrm{I}} \\ \hat{a}_n^{\mathrm{II}} \\ \hat{a}_{n+1}^{\mathrm{II}} \\ \check{a}_n^{\mathrm{II}} \\ \check{a}_{n+1}^{\mathrm{II}} \\ \check{a}_{n-1}^{\mathrm{II}} \\ \check{d}_n^{\mathrm{II}} \\ \check{d}_{n+1}^{\mathrm{II}} \\ \check{d}_{n-1}^{\mathrm{II}} \end{bmatrix}. \tag{2.46}$$

Side decoders

In the case of biorthogonal $5/3$ filter banks an equivalent four-band lifting implementation (as for Haar) is not possible. Moreover, using the pseudo-inverse for the reconstruction of the two side-decoders in the absence of motion estimation/compensation is not a feasible approach because of the delays in the synthesis filters. These delays would lead to an unrealistic reconstruction in which every current frame would be computed from frames in neighbouring GOPs, which are usually too far away in time to be exploitable in this case. In the case of Haar filter banks, the lifting structures forming the four-band MDC schemes allowed for the perfect reconstruction of two out of the three wavelet subbands in each description, thus leaving only one detail subband to interpolate.

Based on these considerations we have instead implemented a simple strategy consisting mainly in interpolating the missing frames from the immediately near ones, at the side decoders. We thus do not present the obtained inverse matrices for the side decoders, since this is irrelevant for the rest of the chapter.

This will be detailed in the next section, which considers an MDC application to video coding.

2.8 Application to robust video coding

In the theoretical part of this chapter we have mostly given considerations about the general MD encoding and decoding scheme for a 1D source signal. Let us now apply the described method to the robust coding of video sequences. If we consider a simple example ignoring the motion information usually inherent to video, the temporal signal in an animated sequence would look as in Figure 2.8. In other words, one temporal signal is given by the variations in the light intensity of each pixel s in a frame, throughout the whole sequence. Intuitively in this example the video is viewed simply as a sequence of still images without any link between them.

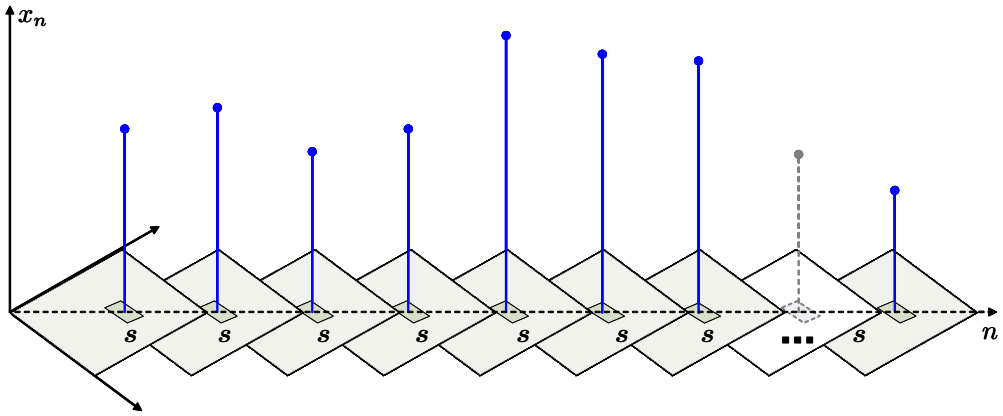


FIGURE 2.8: Example of temporal signal in a video viewed as a sequence of still images, and represented as a function of the frame number (the actual sampling period, $T_e[sec]$, of this signal is given by the frame rate, $F[fps]$ and the number of frames, N in the video sequence).

In practice however, a video is more than a chain of still images and important information related to motion needs to be incorporated. We are considering the so-called $t + 2D$ video codecs, such as MC-EZBC [HW00], [HW01], in which a Motion-Compensated Time Filtering (MCTF) transform is first applied in the temporal direction, this operation being followed by a spatial wavelet transform. The process is completed by an EZBC entropy coding stage.

The Motion Estimation/Compensation (ME/MC) operations performed on the video sequence prior to the temporal transform allow for building the temporal signals that we shall handle in the MDC context. More explicitly, each the motion compensated samples in each the video frame form temporal video signals, as illustrated in Figure 2.9.

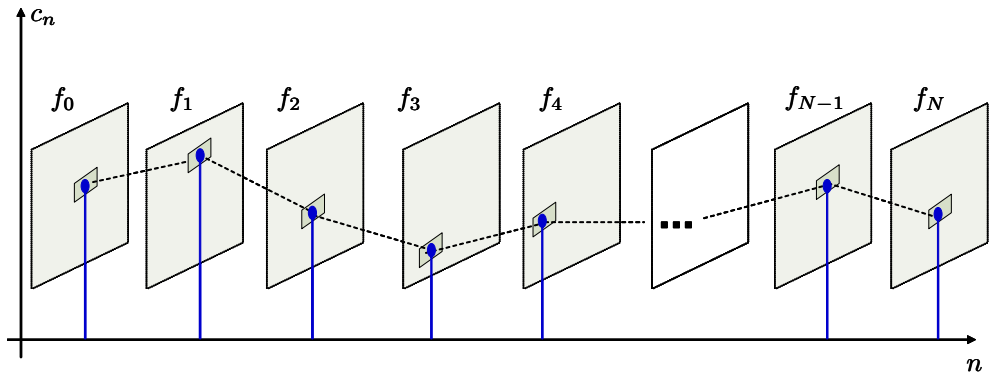


FIGURE 2.9: Temporal video signal after motion estimation. We have denoted the video frames by f_i, $i \in \{0, \ldots, N\}$.

In this application scenario the proposed wavelet frame decompositions have to be adapted to take into account the motion estimation and compensation between video frames, since these are an essential ingredient for the success of our temporal decompositions.

However, as shown in the case of critically-sampled 2-band and 3-band motion-compensated filter banks [PPB01], [TPP03], [PTPPH04], incorporating the ME/MC into the lifting scheme leads to non-linear spatio-temporal operators, and this needs to be taken into

consideration when designing our MD structures.

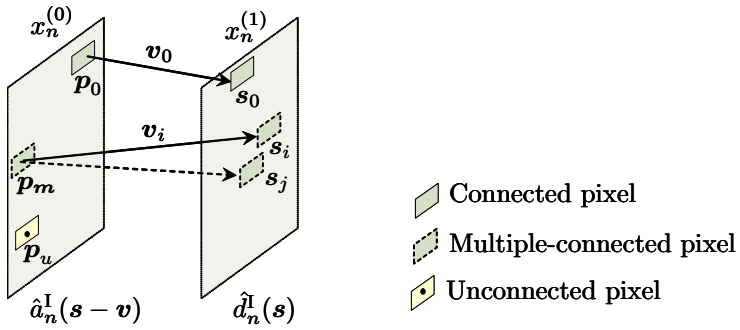


FIGURE 2.10: Motion compensated prediction in the Haar scheme.

2.8.1 Temporal video descriptions

Let us now explain the construction of two temporal video descriptions, using the Haar transform in the **MD3**-scheme. In this case, consider the motion-compensated prediction of a pixel $\boldsymbol{s}$ in the frame $x_n^{(1)}$ from the frame $x_n^{(0)}$ and denote by $\boldsymbol{v}$ the forward motion vector corresponding to $\boldsymbol{s}$ as illustrated in Figure 2.10. We also take into account in this figure the other two possible scenarios that might occur when estimating the motion across frames, which are the so-called multiple-connected pixels [CW99a], represented here as $\boldsymbol{p}_m$ and the unconnected pixels denoted by $\boldsymbol{p}_u$ in this example. In the first situation, several pixels $\boldsymbol{s}_i, i \in \{1\ldots,N\}$ in the current frame $x_n^{(1)}$ need to be predicted by a single pixel in the reference frame $x_n^{(0)}$ and we then have $\boldsymbol{s}_1 - \boldsymbol{v}_1 = \ldots = \boldsymbol{s}_i - \boldsymbol{v}_i = \ldots = \boldsymbol{s}_N - \boldsymbol{v}_N$, with $\boldsymbol{v}_i$ denoting the corresponding motion vectors. The motion vector is considered to be zero if a pixel is unconnected.

Writing now Eqs. (2.44a)-(2.44c) in a lifting form and incorporating the motion into the predict/update operators yields the following set of wavelet subbands at the coarsest resolution level (also corresponding to the first description in the **MD3**-scheme):

$$\begin{cases} \hat{d}_n^{\mathrm{I}}(\boldsymbol{s}) = \dfrac{x_n^{(1)}(\boldsymbol{s}) - x_n^{(0)}(\boldsymbol{s}-\boldsymbol{v})}{\sqrt{2}} \\ \hat{a}_n^{\mathrm{I}}(\boldsymbol{s}-\boldsymbol{v}) = \sqrt{2}x_n^{(0)}(\boldsymbol{s}-\boldsymbol{v}) + \hat{d}_n^{\mathrm{I}}(\boldsymbol{s}) \\ \check{a}_n^{\mathrm{I}}(\boldsymbol{s}) = \dfrac{x_{n-1}^{(2)}(\boldsymbol{s}) + x_{n-1}^{(3)}(\boldsymbol{s})}{\sqrt{2}}. \end{cases} \tag{2.47}$$

Note that the update step may involve all the details $\hat{d}_n^{\mathrm{I}}(\boldsymbol{s}_i),\ i \in \{1,\ldots,N\}$ while preserving the perfect reconstruction property and it has been shown [TPPvdS05] that the update step minimizing the reconstruction error is the one averaging all the detail contributions from the

connected pixels $\boldsymbol{s}_i$. With this remark, one can rewrite Eqs. (2.47) as follows:

$$\begin{cases} \hat{d}_n^{\mathrm{I}}(\boldsymbol{s}_i) = \dfrac{x_n^{(1)}(\boldsymbol{s}_i) - x_n^{(0)}(\boldsymbol{s}_i - \boldsymbol{v}_i)}{\sqrt{2}}, \quad i \in \{1, \ldots, N\} & \text{(2.48a)} \\ \hat{a}_n^{\mathrm{I}}(\boldsymbol{s}_i - \boldsymbol{v}_i) = \sqrt{2} x_n^{(0)}(\boldsymbol{s}_i - \boldsymbol{v}_i) + \dfrac{\sum_{\ell=1}^{N} \hat{d}_n^{\mathrm{I}}(\boldsymbol{s}_\ell)}{N} & \text{(2.48b)} \\ \check{a}_n^{\mathrm{I}}(\boldsymbol{s}) = \dfrac{x_{n-1}^{(2)}(\boldsymbol{s}) + x_{n-1}^{(3)}(\boldsymbol{s})}{\sqrt{2}} & \text{(2.48c)} \end{cases}$$

and with similar notations for multiple connections in the second description:

$$\begin{cases} \check{d}_n^{\mathrm{II}}(\boldsymbol{s}_i) = \dfrac{x_{n-1}^{(2)}(\boldsymbol{s}_i) - x_{n-1}^{(1)}(\boldsymbol{s}_i - \boldsymbol{v}_i)}{\sqrt{2}}, \quad i \in \{1, \ldots, M\} & \text{(2.49a)} \\ \check{a}_n^{\mathrm{II}}(\boldsymbol{s}_i - \boldsymbol{v}_i) = \sqrt{2} x_{n-1}^{(1)}(\boldsymbol{s}_i - \boldsymbol{v}_i) + \dfrac{\sum_{\ell=1}^{M} \check{d}_n^{\mathrm{II}}(\boldsymbol{s}_\ell)}{M} & \text{(2.49b)} \\ \hat{a}_n^{\mathrm{II}}(\boldsymbol{s}) = \dfrac{x_n^{(0)}(\boldsymbol{s}) + x_{n-1}^{(3)}(\boldsymbol{s})}{\sqrt{2}}. & \text{(2.49c)} \end{cases}$$

Since for video coding efficiency motion prediction is an important step, we propose an alternative scheme for building the two descriptions, in which we incorporate the motion estimation/compensation in the computation of the second approximation sequence ($\hat{a}_n^{\mathrm{I}}$, resp. $\check{a}_n^{\mathrm{II}}$).

This scheme is illustrated in Fig. 2.11. In this case an additional motion vector field needs

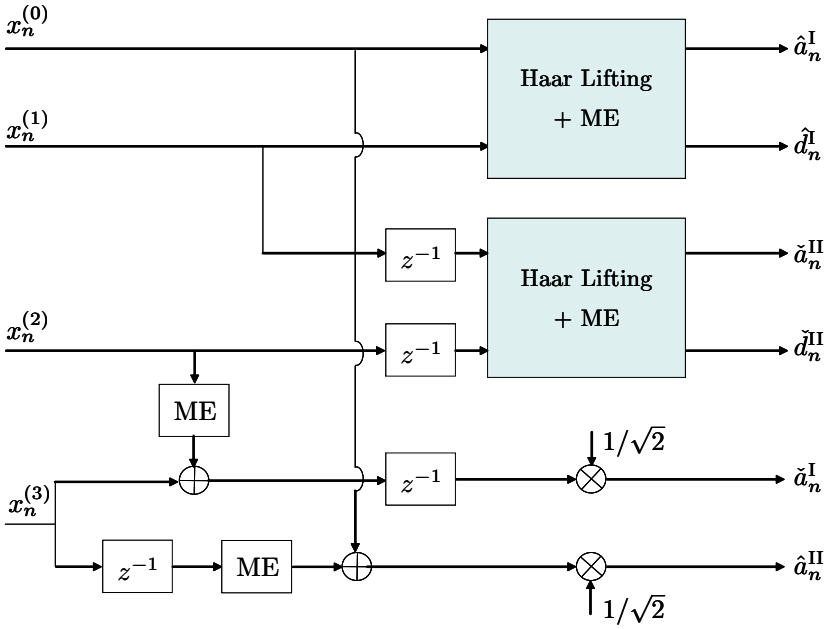


FIGURE 2.11: 4-band lifting scheme with motion estimation on the approximation subbands.

to be encoded per description. Therefore, in the following, this scheme will be referred to as 4B_1MV. Now, if we denote by $\boldsymbol{u}$ the motion vector predicting the pixel $\boldsymbol{s}$ in frame $x_{n-1}^{(3)}$ from $x_{n-1}^{(2)}$ and by $\boldsymbol{w}$ the motion vector predicting the pixel $\boldsymbol{s}$ in frame $x_n^{(0)}$ from $x_{n-1}^{(3)}$, the analysis

equations for $\hat{a}_n^{\mathrm{I}}$ and $\check{a}_n^{\mathrm{II}}$ can be rewritten as:

$$\check{a}_n^{\mathrm{I}}(\boldsymbol{s}-\boldsymbol{u}) = \frac{x_{n-1}^{(3)}(\boldsymbol{s}) + x_{n-1}^{(2)}(\boldsymbol{s}-\boldsymbol{u})}{\sqrt{2}}, \tag{2.50}$$

$$\hat{a}_n^{\mathrm{II}}(\boldsymbol{s}-\boldsymbol{w}) = \frac{x_{n-1}^{(3)}(\boldsymbol{s}-\boldsymbol{w}) + x_n^{(0)}(\boldsymbol{s})}{\sqrt{2}}, \tag{2.51}$$

for the connected pixels (here, only the first pixel in the scan order is considered in the computation), and

$$\check{a}_n^{\mathrm{I}}(\boldsymbol{s}) = \sqrt{2}x_{n-1}^{(2)}(\boldsymbol{s}),$$
$$\hat{a}_n^{\mathrm{II}}(\boldsymbol{s}) = \sqrt{2}x_{n-1}^{(3)}(\boldsymbol{s}),$$

for the non-connected pixels.

Furthermore, it can be noticed that the two polyphase components of the approximation signals that enter each description are temporally correlated. This suggested us to come up with a new coding scheme, illustrated in Fig. 2.12, where a motion-compensated temporal Haar transform is applied on $\hat{a}_n^{\mathrm{I}}$ and $\check{a}_n^{\mathrm{I}}$ (resp., on $\check{a}_n^{\mathrm{II}}$ and $\hat{a}_n^{\mathrm{II}}$). Compared to the original structure, two additional motion vector fields have to be transmitted. The scheme will thus be referred to as 4B_2MV.

We represent the temporal transforms involved in two decomposition levels of the 4B_2MV scheme in Figure 2.23. Note the temporal subsampling of the details on the first decomposition level and the redundancy introduced at the second decomposition level.

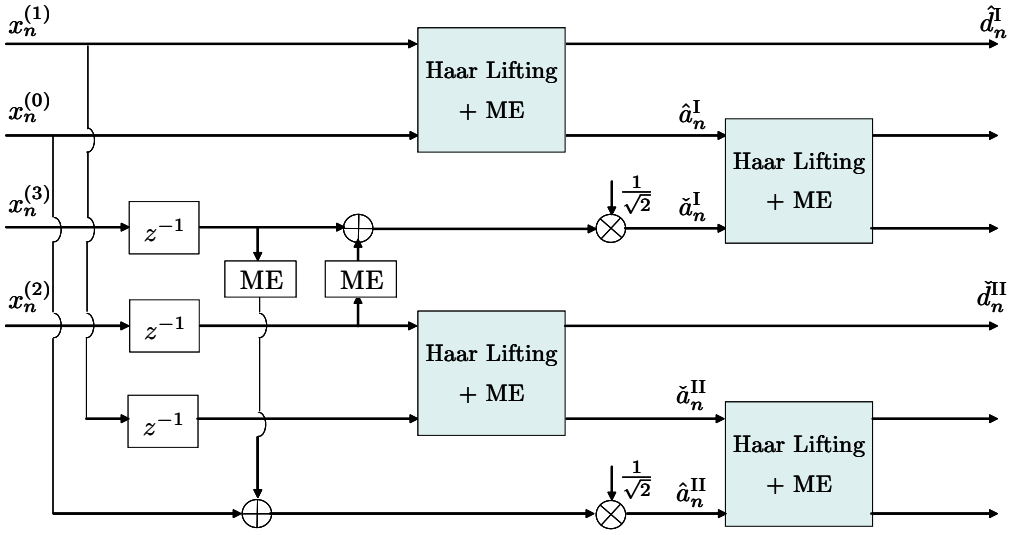


FIGURE 2.12: 4-band lifting scheme with motion estimation and Haar transform on the approximation subbands of each description.

Motion compensated prediction in the biorthogonal 5/3 filter bank case

In the biorthogonal 5/3 filter bank case an equivalent 4-band lifting structure can not be formulated. Moreover, including motion into the last stages of decomposition is a very tedious

task in the multiple description case. Therefore, for these filters we have considered that the coarsest level subbands are obtained as we presented in the theoretical part of this chapter, with no ME/MC. The higher resolution levels are obtained in the classical motion-compensated 5/3 temporal filtering scenario, in which the motion vectors fields are bi-directional. A schematic representation for the prediction stage is given in Figure 2.13.

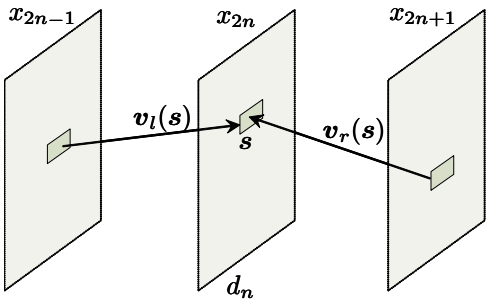


FIGURE 2.13: Motion compensated prediction in the biorthogonal 5/3 scheme. Motion vectors are denoted by $v(s)$ and their direction is given by the indices l or r, standing for "left" or "right", respectively. The detail subbands are synchronized with the even-indexed x_{2n} temporal frames.

At the decoder end we are using the following strategy. The central decoder uses the pseudo-inverse in Eq. (2.46) in order to recover the last decomposition level subbands and then the classical 2-band 5/3 lifting inversion scheme is used at higher levels. For the side decoders however, the reconstruction without motion could lead to much too poor results. We have therefore chosen to use the motion vectors computed at the encoding stage whenever it is possible, as it will be shown in the next section.

2.8.2 Central and side video decoders

Haar filter bank

The inversion of the Eqs. (2.48a) and (2.48b) is straightforward by the lifting scheme, allowing us to reconstruct the first two polyphase components. Using the same notations as in Section 2.4, the reconstructed polyphase components from the first description are as follows:

$$\begin{cases} \tilde{x}_n^{(0)}(\boldsymbol{s}_i - \boldsymbol{v}_i) = \dfrac{1}{\sqrt{2}} \left([\hat{a}_n^{\mathrm{I}}(\boldsymbol{s}_i - \boldsymbol{v}_i)] - \dfrac{1}{N} \displaystyle\sum_{\ell=1}^{N} [\hat{d}_n^{\mathrm{I}}(\boldsymbol{s}_\ell)] \right) \\ \tilde{x}_n^{(1)}(\boldsymbol{s}_i) = \dfrac{1}{\sqrt{2}} \left([\hat{a}_n^{\mathrm{I}}(\boldsymbol{s}_i - \boldsymbol{v}_i)] + 2[\hat{d}_n^{\mathrm{I}}(\boldsymbol{s}_i)] - \dfrac{1}{N} \displaystyle\sum_{\ell=1}^{N} [\hat{d}_n^{\mathrm{I}}(\boldsymbol{s}_\ell)] \right) \end{cases} \tag{2.52}$$

When analyzing the reconstruction of the connected pixels in the first two polyphase components, one can remark that it corresponds to the inverse lifting using the average update step.

A similar reasoning for the second description allows us to find the reconstruction of the sequence from the received frames $\check{a}_n^{\mathrm{II}}$, $\check{d}_n^{\mathrm{II}}$, and $\hat{a}_n^{\mathrm{II}}$. By inverting the Eqs. (2.49a) and (2.49b) we

obtain:

$$\begin{cases} \tilde{x}_n^{(1)}(\boldsymbol{s}_i - \boldsymbol{v}_i) = \dfrac{1}{\sqrt{2}} \left([\check{a}_{n+1}^{\mathrm{II}}(\boldsymbol{s}_i - \boldsymbol{v}_i)] - \dfrac{1}{M} \displaystyle\sum_{\ell=1}^{M} [\check{d}_{n+1}^{\mathrm{II}}(\boldsymbol{s}_\ell)] \right) \\ \tilde{x}_n^{(2)}(\boldsymbol{s}_i) = \dfrac{1}{\sqrt{2}} \left([\check{a}_{n+1}^{\mathrm{II}}(\boldsymbol{s}_i - \boldsymbol{v}_i)] + 2[\check{d}_{n+1}^{\mathrm{II}}(\boldsymbol{s}_i)] - \dfrac{1}{M} \displaystyle\sum_{\ell=1}^{M} [\check{d}_{n+1}^{\mathrm{II}}(\boldsymbol{s}_\ell)] \right) \end{cases} \tag{2.53}$$

For the non-connected pixels we have: $\tilde{x}_n^{(0)}(s_i) = \frac{1}{\sqrt{2}}[\hat{a}_n^{\mathrm{I}}(s_i)]$ and $\tilde{x}_n^{(1)}(s_i) = \frac{1}{\sqrt{2}}[\check{a}_{n+1}^{\mathrm{II}}(s_i)]$.

As it can be seen, $\tilde{x}_n^{(1)}$ can be recovered from both descriptions, and the final central reconstruction is obtained as the mean of these values. Also, one can remark that by knowing $x_{n-1}^{(2)}$ (resp. $x_n^{(0)}$) from the first (resp., second) description, it is possible to reconstruct $x_{n-1}^{(3)}$, by reverting Eqs. (2.50) and (2.51).

As for the side decoders of the initial scheme, the solution for the first description is given by Eqs. (2.52) and:

$$\tilde{x}_n^{(2)}(\boldsymbol{s}) = \tilde{x}_n^{(3)}(\boldsymbol{s}) = \frac{1}{\sqrt{2}} [\check{a}_{n+1}^{\mathrm{I}}(\boldsymbol{s})],$$

while for the second description it reads:

$$\tilde{x}_{n+1}^{(0)}(\boldsymbol{s}) = \tilde{x}_n^{(3)}(\boldsymbol{s}) = \frac{1}{\sqrt{2}} [\check{a}_{n+1}^{\mathrm{II}}(\boldsymbol{s})],$$

in addition to $\tilde{x}_n^{(1)}$ and $\tilde{x}_n^{(2)}$ obtained with Eqs. (2.53).

For the 4B_1MV scheme, the additional motion compensation involved in the computation of the approximation sequences requires reverting the motion vector field in one of the components. Thus, we have:

$$\tilde{x}_{n-1}^{(2)}(\boldsymbol{s}) = \frac{[\check{a}_n^{\mathrm{I}}(\boldsymbol{s})]}{\sqrt{2}},$$
$$\tilde{x}_{n-1}^{(3)}(\boldsymbol{s}) = \frac{[\check{a}_n^{\mathrm{I}}(\boldsymbol{s} - \boldsymbol{u})]}{\sqrt{2}},$$

for the first side decoder and

$$\tilde{x}_{n-1}^{(3)}(\boldsymbol{s}) = \frac{[\hat{a}_n^{\mathrm{II}}(\boldsymbol{s})]}{\sqrt{2}},$$
$$\tilde{x}_n^{(0)}(\boldsymbol{s}) = \frac{[\hat{a}_n^{\mathrm{II}}(\boldsymbol{s} - \boldsymbol{u})]}{\sqrt{2}},$$

for the second one.

For the scheme 4B_2MV, the temporal Haar transform being revertible, no additional difficulties appear for the central or side decoders.

Note that the reconstruction by one central and two side decoders corresponds to a specific application scenario, in which the user receives the two descriptions from two different locations (for example, two WiFi access points), but depending on its position, it can receive both or only one of the descriptions. In a more general scenario, the user may be in the reception zone of both access points, but packets may be lost from both descriptions (due to network congestion,

transmission quality etc). In this case, the central decoder will try to reconstruct the sequence by exploiting the information in all the received packets. It is therefore clear that an important issue for the reconstruction quality will be the packetization strategy. Even though the complete description of the different situations which can appear in the decoding (depending on the type of the lost packets) cannot be done here, it is worth noting that in a number of cases an efficient usage of the received information can be employed: for instance, even if we do not receive the spatio-temporal subbands of one of the descriptions, but only a packet containing its motion vectors, these vectors can be exploited in conjunction with the other description for improving the fluidity of the reconstructed video. We also take advantage of the redundancy existing at the last level to choose, for the frames which can be decoded from both descriptions, the version which has the best quality, and thus to limit the degradations appearing in one of the descriptions.

Biorthogonal 5/3 filter bank

Let us consider a two level scheme based on the biorthogonal 5/3 filter bank, as in Figure 2.14.

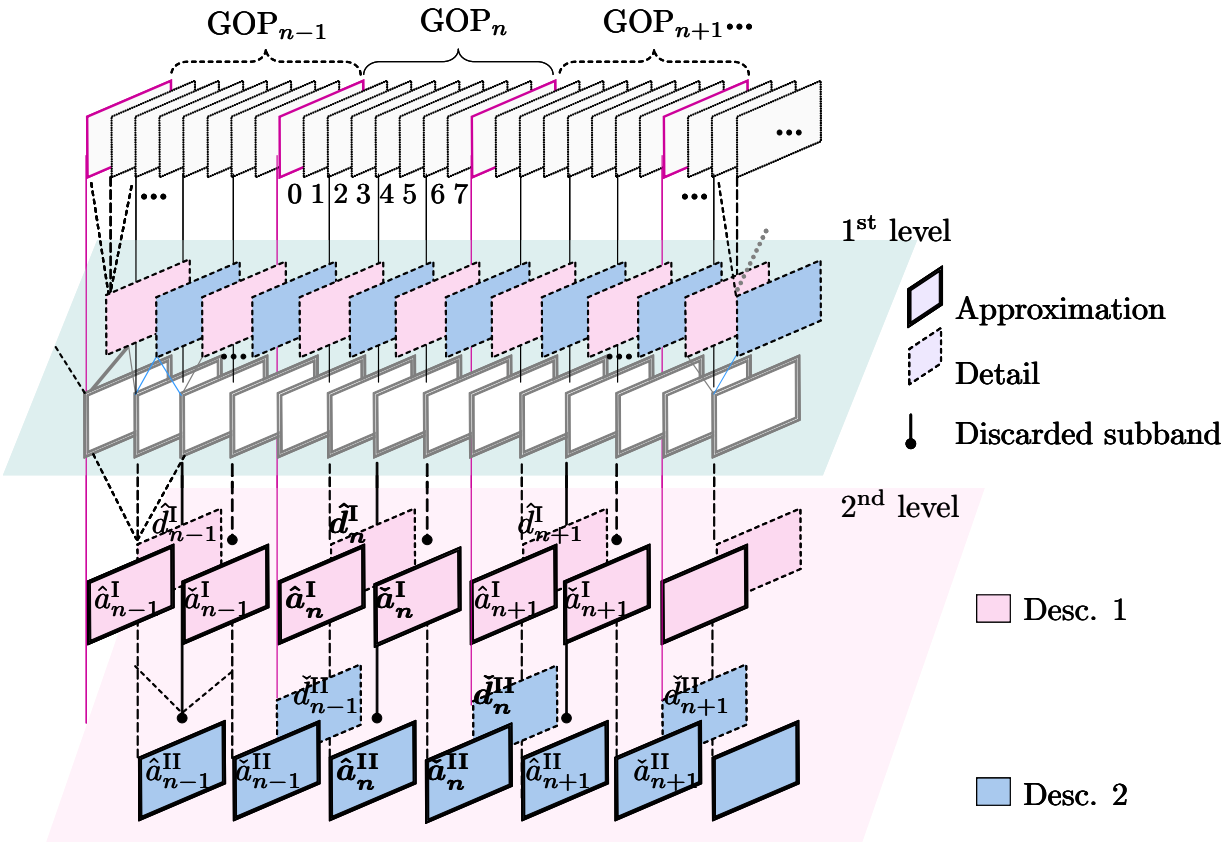


FIGURE 2.14: Two levels of biorthogonal 5/3 temporal decomposition of a video sequence.

Here, three successive GOPs of a video sequence are first temporally decomposed using the 5/3 filter banks. In this figure we highlight the multiple description decomposition at each level. Thus at the first level of decomposition we split the details into the two descriptions according to their index parity. The even indexed details go to the first description and the odd ones go to the second one. The approximation subbands are further decomposed to the next

level, at which we obtain the two descriptions as presented in the beginning of this chapter. We also indicate the positions of the discarded coefficients at this (coarsest) level. The vertical and oblique lines between decomposition levels show the video frames involved in the computation of one wavelet subband (the oblique lines also point out the vector fields calculated at the encoder).

As we said before, no motion estimation/compensation is used at the last level for the central reconstruction. Thus, at the central decoder we simply recover the input frames by using the Eq. (2.46).

The side reconstruction based on this example implies the reconstruction of the eight frames that form the video GOP, n, denoted here by $\tilde{x}_n^{\#i}(\boldsymbol{s})$, $i \in \{0, \ldots 7\}$, only from the received subset of wavelet coefficients (those forming the Description 1 for the Side A decoder, for instance). Since some of the detail subbands are missing, we can not make much use of the bi-directional motion vectors, thus we are only going to use one of the two motion vectors, whenever possible.

We first recover the frames which are synchronous with the approximation subbands in the first description, by renormalizing them twice (since the decomposition is done on two levels). We reuse the notation $[.]$ to indicate that the subbands are quantized. Thus we obtain:

$$\tilde{x}_n^{\#0}(\boldsymbol{s}) = \frac{1}{2}[\hat{a}_n^{\mathrm{I}}(\boldsymbol{s})],$$
$$\tilde{x}_n^{\#4}(\boldsymbol{s}) = \frac{1}{2}[\check{a}_n^{\mathrm{I}}(\boldsymbol{s})].$$

Then we can use these two frames to recover $\tilde{x}_n^{\#2}$ by averaging $\tilde{x}_n^{\#0}$ and $\tilde{x}_n^{\#4}$:

$$\tilde{x}_n^{\#2}(\boldsymbol{s}) = \frac{1}{2}(\tilde{x}_n^{\#0}(\boldsymbol{s}) + \tilde{x}_n^{\#4}(\boldsymbol{s})). \tag{2.54}$$

Similarly, we recover $\tilde{x}_n^{\#3}$ as an average of $\tilde{x}_n^{\#2}$ and $\tilde{x}_n^{\#4}$:

$$\tilde{x}_n^{\#3}(\boldsymbol{s}) = \frac{1}{2}(\tilde{x}_n^{\#2}(\boldsymbol{s}) + \tilde{x}_n^{\#4}(\boldsymbol{s})). \tag{2.55}$$

For the reconstruction of $\tilde{x}_n^{\#1}(\boldsymbol{s})$ we can take advantage of the motion vector $\boldsymbol{v}_l$ (as depicted in Figure 2.13). On the other hand, the right motion field $\boldsymbol{v}_r$ which would compensate $\tilde{x}_n^{\#1}(\boldsymbol{s})$ from $\tilde{x}_n^{\#2}(\boldsymbol{s})$ is unusable, since $\tilde{x}_n^{\#2}$ has been already quite damaged by the averaging operation. Therefore, $\tilde{x}_n^{\#1}$ will be given by:

$$\tilde{x}_n^{\#1}(\boldsymbol{s}) = \tilde{x}_n^{\#0}(\boldsymbol{s} + \boldsymbol{v}_{l_0}),$$

where we have denoted by $\boldsymbol{v}_{l_0}$ the motion vector between the frame number 0 and the frame number 1. The same problem in using the vector field computed at the first decomposition level arises for the reconstruction of the $\tilde{x}_n^{\#5}$ video frame. We proceed similarly to $\tilde{x}_n^{\#1}$, and thus obtain:

$$\tilde{x}_n^{\#5}(\boldsymbol{s}) = \tilde{x}_n^{\#4}(\boldsymbol{s} + \boldsymbol{v}_{l_4}),$$

with a similar notation for the motion vector between the frames number 4 and 5. The only frames that need yet to be reconstructed in this GOP are $\tilde{x}_n^{\#6}$ and $\tilde{x}_n^{\#7}$. These are recovered in

the same way as those in Eqs. (2.54) and (2.55), but using the the corresponding frames in the next GOP. More precisely, we have:

$$\tilde{x}_n^{\#6}(\boldsymbol{s}) = \frac{1}{2}(\tilde{x}_n^{\#4}(\boldsymbol{s}) + \frac{1}{2}[\hat{a}_{n+1}^{\mathrm{I}}(\boldsymbol{s})]),$$
$$\tilde{x}_n^{\#7}(\boldsymbol{s}) = \frac{1}{2}(\tilde{x}_n^{\#6}(\boldsymbol{s}) + \frac{1}{2}[\hat{a}_{n+1}^{\mathrm{I}}(\boldsymbol{s})]).$$

The border GOPs are handled in the usual manner, by making mirror extension of the GOP (in fact, equivalently modifying the filtering coefficients).

2.8.3 Simulation experiments and results

Let us begin the presentation of our experimental framework by recalling that when comparing the four MDC schemes based on wavelet-frames, in terms of global noise power of the reconstructed sequences (in Section 2.4.4), the **R** and **MD2** schemes appeared to provide lower performances than the **MD1** and **MD3** ones. Therefore, in the following we are only presenting results concerning the latter two schemes.

This section is organized as follows. In a first part we are presenting comparisons between the **MD1** and **MD3** schemes for the Haar transform. This will give a choice criterion based on their side performances, since the two schemes exhibit similar central performances as it will be shown shortly. We will find that the scheme **MD3** is the best compromise in terms of central/side reconstruction. Therefore, the rest of the section concentrates entirely on this scheme.

Then we have considered two application scenarios. In the first one, an entire description might be lost during transmission and this corresponds to the "on-off channels" scenario. We have compared the reconstruction results on several video sequences for Haar and 5/3 filter banks in a reduced motion estimation/compensation scenario, namely by excluding this operation from the coarsest resolution level. This simplification is needed since a motion compensated lifting scheme has not been found for the 5/3 filters in the multiple description context. This comparison gives a choice criterion between these two filter banks. We have concluded that even though the central performances are better for the 5/3 filter banks the side ones are much too poor for the introduced complexity. Therefore in a second application scenario we only consider the Haar filter bank.

This new application scenario is called "Packet-losses" transmission, and it involves forming packets inside each of our two temporal descriptions and authorizing the loss of a subset. This scenario corresponds to the video transmission over Ethernet networks, for instance, without QoS.

Choice between schemes

We have implemented $J = 3$ levels of motion-compensated temporal lifting Haar decomposition [PPB01], the last level consisting of one of the two analyzed schemes: **MD1** or **MD3**. Recall that the detail frames obtained at resolution levels $j < J$ have been alternately distributed between the two descriptions in an identical manner for the two schemes. They only differ at the coarsest resolution level and the overall redundancy of the structure has the size of an approximation subband.

The proposed schemes have been tested on several CIF sequences at 30fps. See for instance Figure 2.15 for an example of four frames extracted from the test video sequence "Foreman" and Figure 2.16 for a similar excerpt of the test video sequence "Mobile".

(a) Frame 1 (b) Frame 2

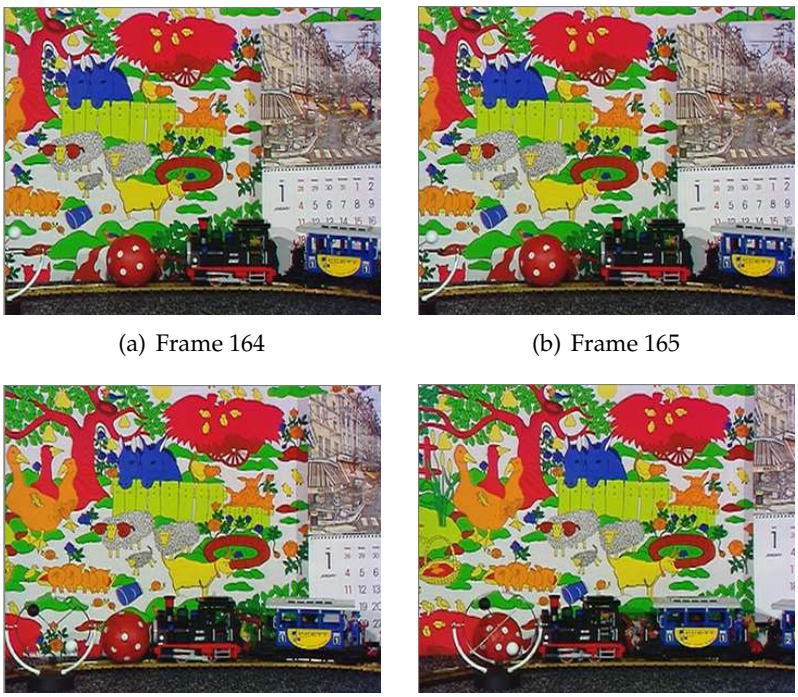
(c) Frame 3 (d) Frame 4

FIGURE 2.15: The first four frames extracted from the test video sequence "Foreman".

On the first two temporal decomposition levels a full-pixel motion compensation is involved in the lifting transform, while at the last level no motion estimation is performed. The temporal subband frames have been decomposed with the 9/7 biorthogonal wavelets. The spatio-temporal wavelet coefficients and motion vectors have been coded as for the non redundant codec, by using the MC-EZBC algorithm [ezb].

In Figures 2.17 and 2.18 we compare the PSNR-rate performances of the central and side decoders for the **MD1** and **MD3** schemes, on the Y component of the signal. The exact figures are given in Table 2.5. Note that the central decoder of the **MD1** scheme slightly outperforms the central decoder of **MD3**, as predicted by the theoretical framework. One of the side decoders (denoted by "A" in Figure 2.17) is identical for the two schemes. However, due to an asymmetrical construction of the two descriptions in the **MD1** scheme, one of its side decoders (denoted by "B" in Figure 2.17) exhibits a poorer performance.

The motion estimation is performed using Hierarchical Variable Size Block Matching (HVBSM) algorithm with block sizes ranging from 64×64 to 4×4. An integer pel accuracy is used for motion compensation. Spatio-temporal coefficients and motion vectors (MV) are encoded within the MC-EZBC framework [CW99a, ezb], where MV fields are first represented as quad-tree maps and MV values are encoded with a 0-order arithmetic coder, in raster-scan order.



(a) Frame 164 (b) Frame 165

(c) Frame 228 (d) Frame 275

FIGURE 2.16: Four frames extracted from the test video sequence "Mobile".

"FOREMAN" **MD1** scheme						
bitrate	250	500	750	1000	1500	3000
central	29.48	32.19	33.85	34.98	36.85	40.53
side A	26.05	27.20	27.78	28.13	28.66	29.51
side B	24.32	24.84	25.06	25.16	25.29	25.43
"FOREMAN" **MD3** scheme						
bitrate	250	500	750	1000	1500	3000
central	29.27	32.01	33.68	34.79	36.68	40.39
side A	26.05	27.20	27.78	28.13	28.66	29.51
side B	25.26	26.16	26.62	26.88	27.28	27.96
"MOBILE" **MD1** scheme						
bitrate	250	500	750	1000	1500	3000
central	19.89	22.18	23.54	24.88	26.55	30.61
side A	18.96	20.15	20.78	21.24	21.91	23.07
side B	18.36	19.24	19.70	19.93	20.37	20.90
"MOBILE" **MD3** scheme						
bitrate	250	500	750	1000	1500	3000
central	19.70	21.95	23.33	24.61	26.33	30.43
side A	18.96	20.15	20.78	21.24	21.91	23.07
side B	18.81	19.83	20.35	20.72	21.28	22.23

TABLE 2.5: Rate-distortion comparison: YSNR (dB) at different bitrates (Kbs), for "FOREMAN" and "MOBILE" sequences (CIF at 30fps) on three levels of wavelet decomposition.

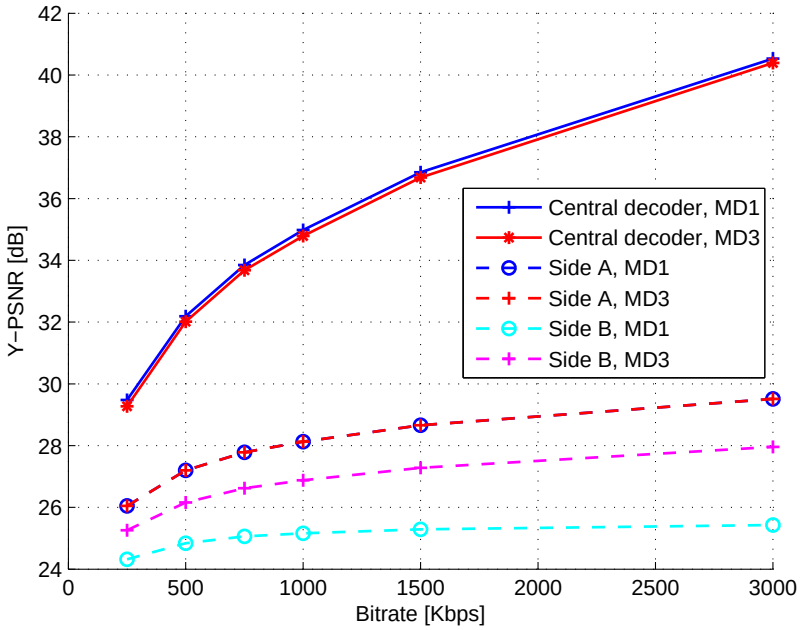


FIGURE 2.17: Central and side Y-PSNR vs. rate curves for the schemes **MD1** and **MD3** ("Foreman" CIF sequence, 30fps).

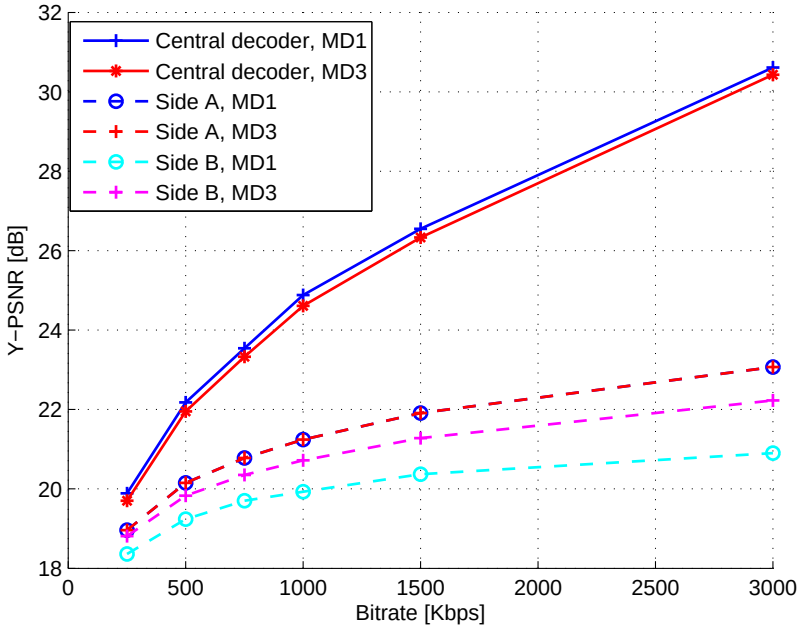


FIGURE 2.18: Central and side Y-PSNR vs. rate curves for the schemes **MD1** and **MD3** ("Mobile" CIF sequence, 30fps).

2.8.3.1 On-off channels scenario

As we previously said, in this scenario we have considered the loss of an entire description and we have tested the **MD3** scheme on several video sequences.

Choice of the filter bank

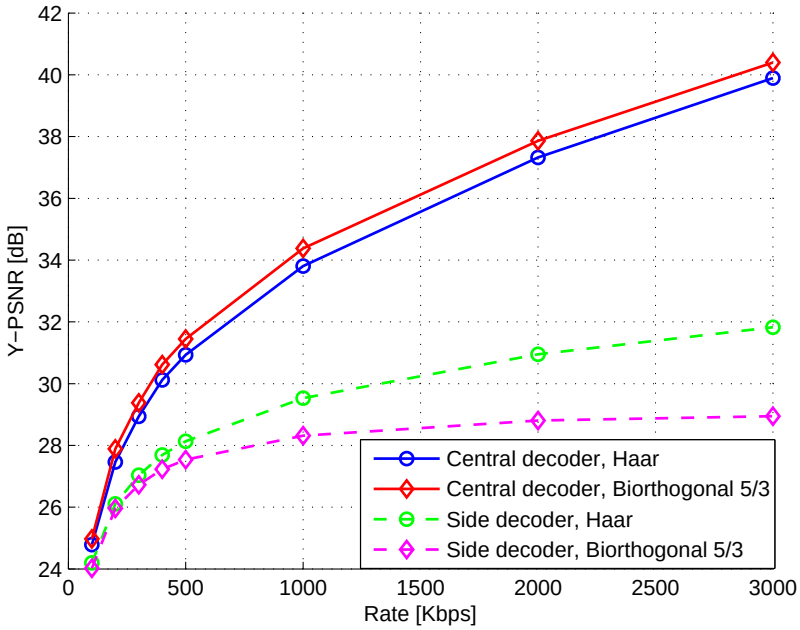


FIGURE 2.19: Y-PSNR vs. rate curves("Foreman", CIF at 30 fps). Comparison between the Haar and the 5/3 biorthogonal FB scheme, on two levels of decomposition.

In Figures 2.19 and 2.20 we present the PSNR-rate performance comparison for Haar and biorthogonal 5/3 filter banks on two decomposition levels, for the test sequences "Foreman" and "Mobile" in CIF format at 30fps. As expected from Tables 2.3 and 2.4, the central decoder performs better in the 5/3 case.

Robust Haar MD coder

In Fig. 2.21 we compare the rate-distortion performance of the non-robust Haar scheme with that of the MDC central decoder on the "Foreman" video test sequence in QCIF format at 30fps. The bitrate corresponds to the global rate for the robust codec (both descriptions). Three temporal decomposition levels have been used in this experiment ($J = 3$). We can observe that even the loss of one description still allows for acceptable quality reconstruction especially at low bitrates and also that the global redundancy does not exceed 30% of the bitrate.

Fig. 2.22 illustrates the central Y-PSNR vs. rate curves for different levels of redundancy and, together with Fig. 2.21 shows the narrowing of the gap with respect to the non redundant version when the number of decomposition levels increases.

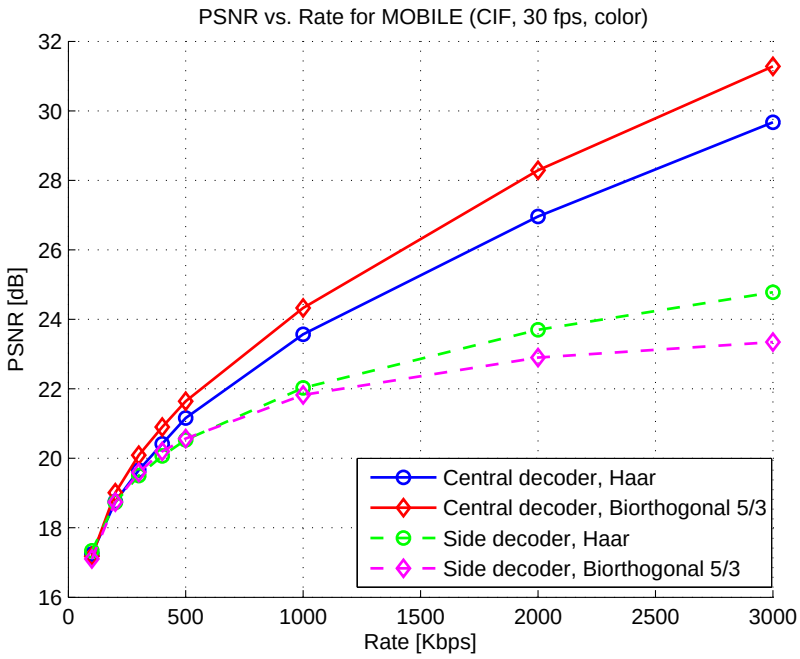


FIGURE 2.20: Y-PSNR vs. rate curves("Mobile", CIF at 30 fps). Comparison between the Haar and the 5/3 biorthogonal FB scheme, on two levels of decomposition.

The difference in performance between the two descriptions is a phenomenon appearing only if the scheme involves three or more decomposition levels, since it is related to an asymmetry in the GOF structure of the two descriptions when performing the decimation. Indeed, as illustrated in Fig. 2.23, when the first description is lost, some of the motion information in the second description cannot be used to improve the reconstruction, while this does not happen when loosing the second description.

In Figs. 2.24-2.25, we present the Y-PSNR vs. rate curves for the central and side decoders, in the absence of packet losses. The performances of the scheme without ME/MC in the computation of the approximation sequences $\check{a}_n^{\mathrm{I}}$ and $\hat{a}_n^{\mathrm{II}}$ are compared with the 4B_1MV and 4B_2MV schemes. One can remark that the addition of the ME/MC step in the computation of $\check{a}_n^{\mathrm{I}}$ and $\hat{a}_n^{\mathrm{II}}$ (4B_1MV) does not lead to an increase in the coding performance of the central decoder, since the expected gain is balanced by the need to encode an additional MV field. On the other hand, the final MC-Haar transform (4B_2MV) leads to much better results, since instead of two correlated approximation sequences we now only have transformed subbands. For the side decoders however, the introduction of the motion-compensated average in the computation of $\check{a}_n^{\mathrm{I}}$ and $\hat{a}_n^{\mathrm{II}}$ leads to a significant improvement in coding performances (increasing with the bitrate from 1 to 2.5 dB), and the MC-Haar transform adds another 0.3 dB of improvement.

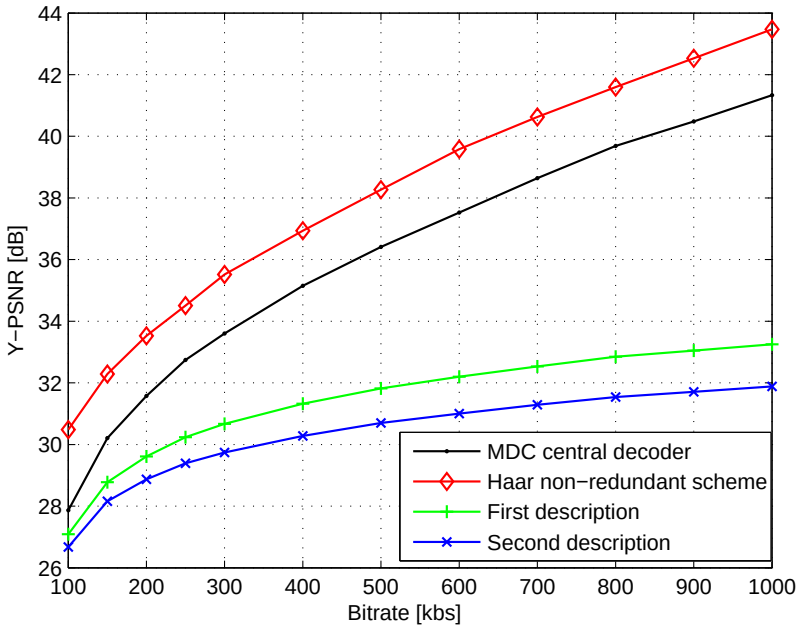


FIGURE 2.21: Central and side Y-PSNR vs. rate curves of the MDC scheme compared with the non-robust Haar codec ("Foreman" QCIF sequence, 30fps).

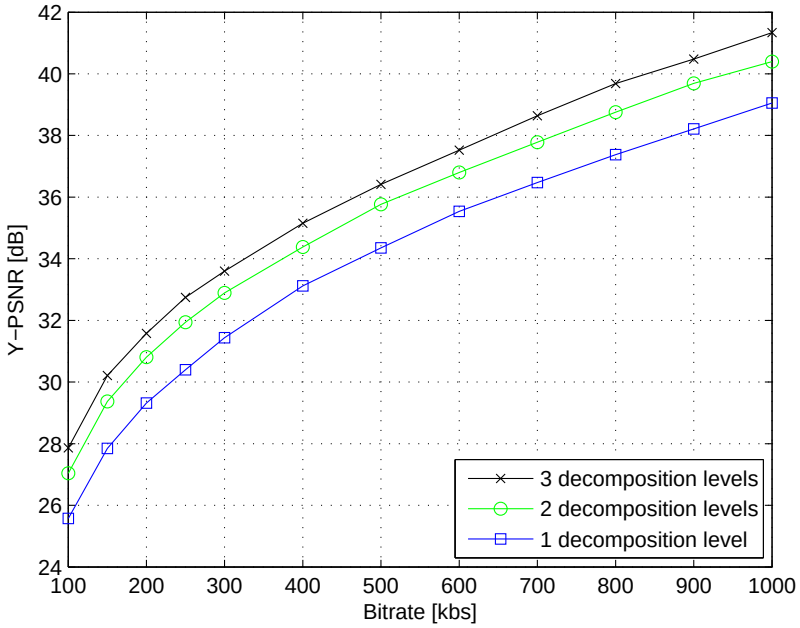


FIGURE 2.22: Y-PSNR vs. rate curves at the central decoder for several levels of decomposition (redundancy).

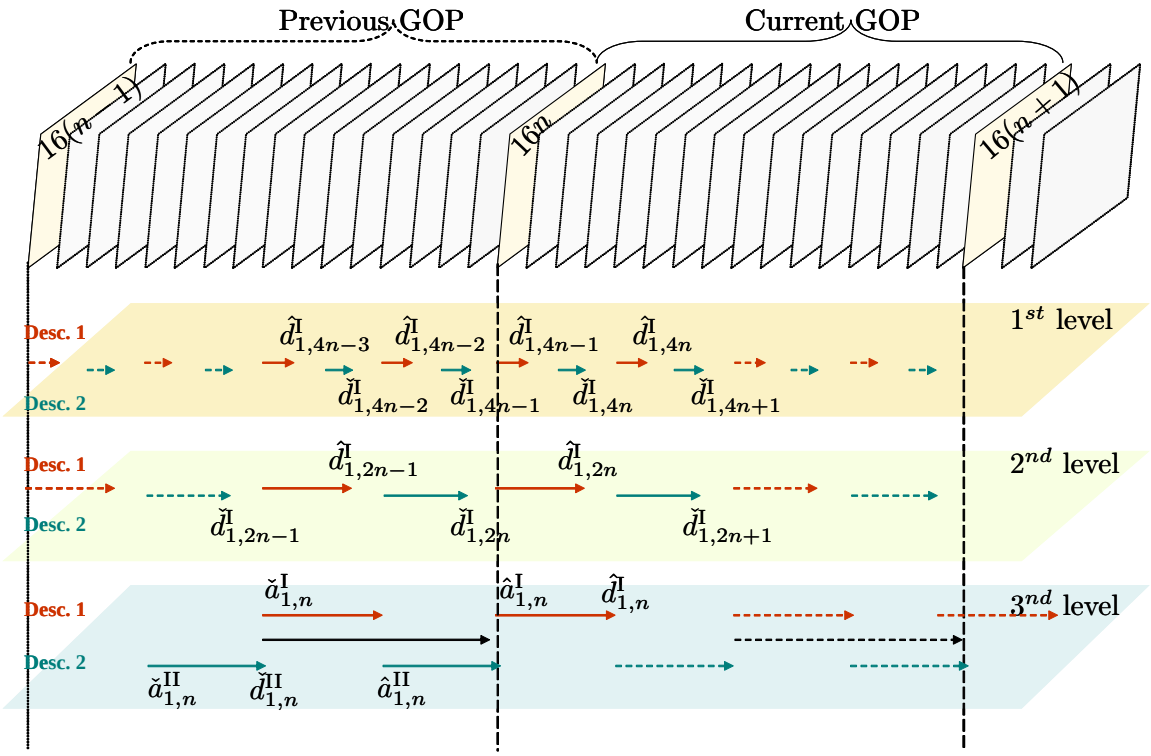


FIGURE 2.23: 4B_2MV scheme over 3 levels (GOP size = 16). Motion-compensated temporal operations are represented by arrows (solid lines for the current GOP, dashed lines for the adjacent GOPs).

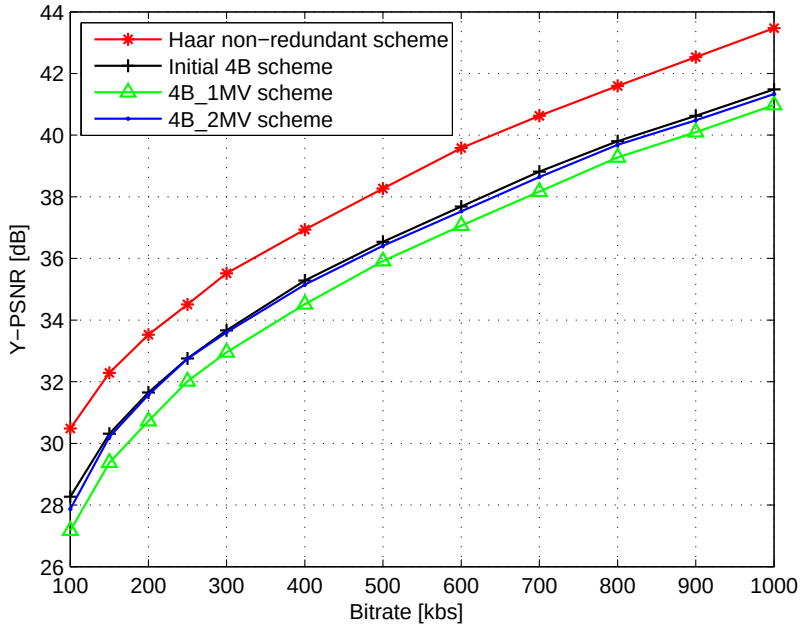


FIGURE 2.24: Y-PSNR vs. rate curves for different reconstruction strategies, central decoder ("Foreman" QCIF sequence, 30 fps).

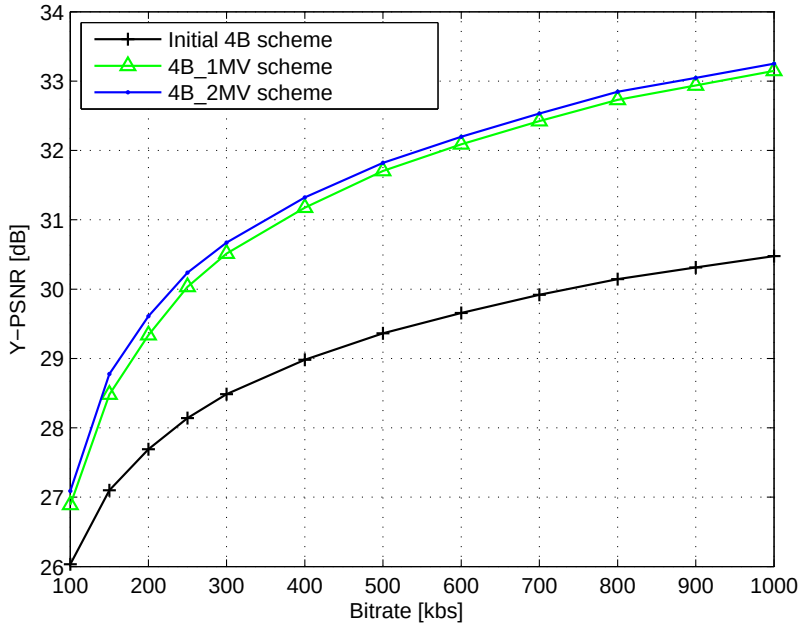


FIGURE 2.25: Y-PSNR vs. rate curves for different reconstruction strategies, first side decoder ("Foreman" QCIF sequence, 30 fps).

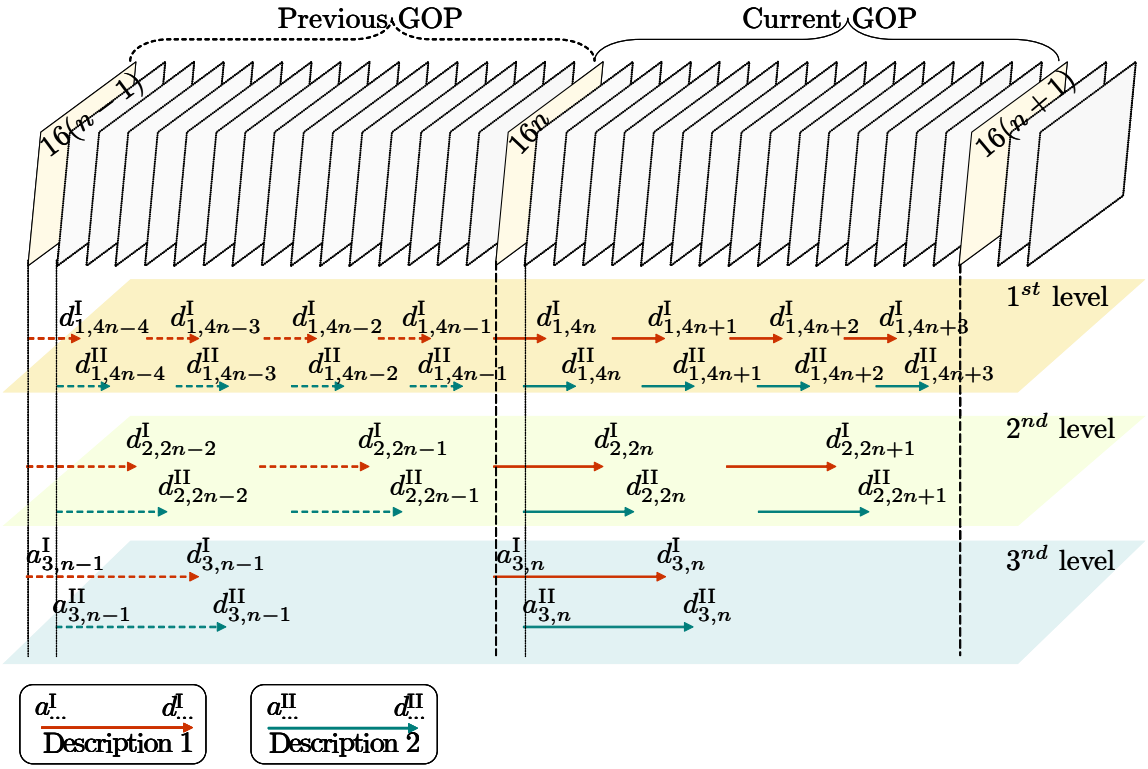


FIGURE 2.26: Three levels of decomposition in the temporal splitting scheme.

2.8.3.2 Packet-losses scenario

In this second scenario, we have tested our **MD3** scheme for transmission over a packet-loss network, like Ethernet. In this case, the bitstreams of the two descriptions are separated in packets of maximal size of 1500 bytes. For each GOP, separate packets are created for the motion vectors and for each spatio-temporal subband. If the packet with motion vectors is lost, or if the packet with the spatial approximation subband of the temporal approximation subband is lost, then we consider that the entire GOP is lost (it cannot be reconstructed).

We compare our scheme with a non redundant MCTF one and also with another very well-known temporal MDC scheme, consisting in a temporal splitting of the initial video sequence. Odd and even frames are separated into two descriptions which are encoded with a Haar MCTF coder (Fig. 2.26 illustrates the motion vectors and temporal transforms for this structure).

The coding performance as a function of the packet loss rate is illustrated in Figs. 2.27 and

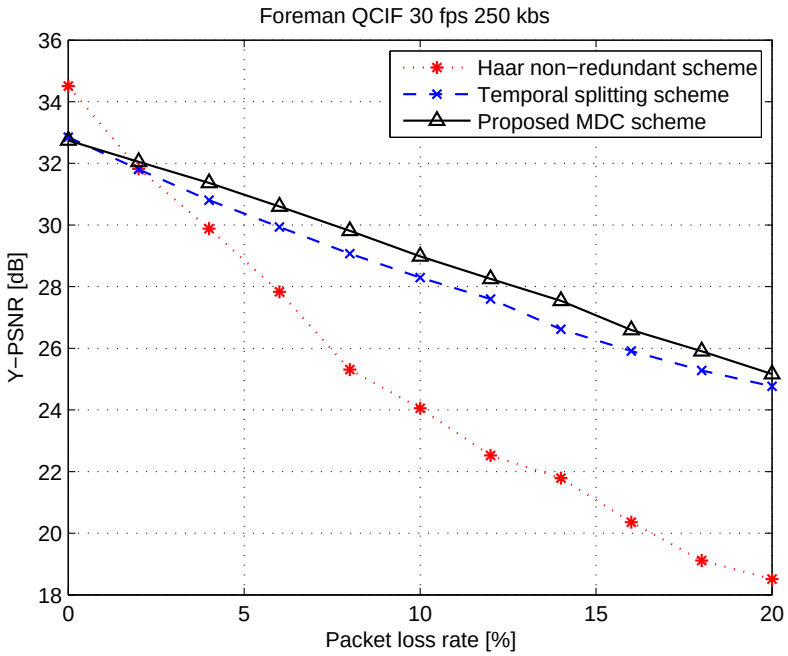


FIGURE 2.27: Distortion vs. packet loss rate ("Foreman" QCIF sequence, 30 fps at 250 Kbs).

2.28 for the "Foreman" and "Mobile" video test sequences at 250 Kbs. As expected, when there is no loss, the non redundant coding is better than both MDC schemes (which have comparable performances). However, as soon as the packet loss rate gets higher than 2%, our scheme overpasses by 0.5-1dB the temporal splitting and the non robust coding by up to 4dB.

Moreover, we have noticed that the MDC splitting scheme exhibits a flickering effect, due to the fact that a lost packet will degrade the quality of one over two frames. In our scheme this effect is not present, since the errors in one description have limited influence thanks to the existing redundancies, and also to a different propagation during the reconstruction process.

Fig. 2.29 presents the influence of the average update operator, with gains of about 0.2 dB over the entire range of packet loss rates. Finally, we have compared in Fig. 2.30 the Y-PSNR vs.

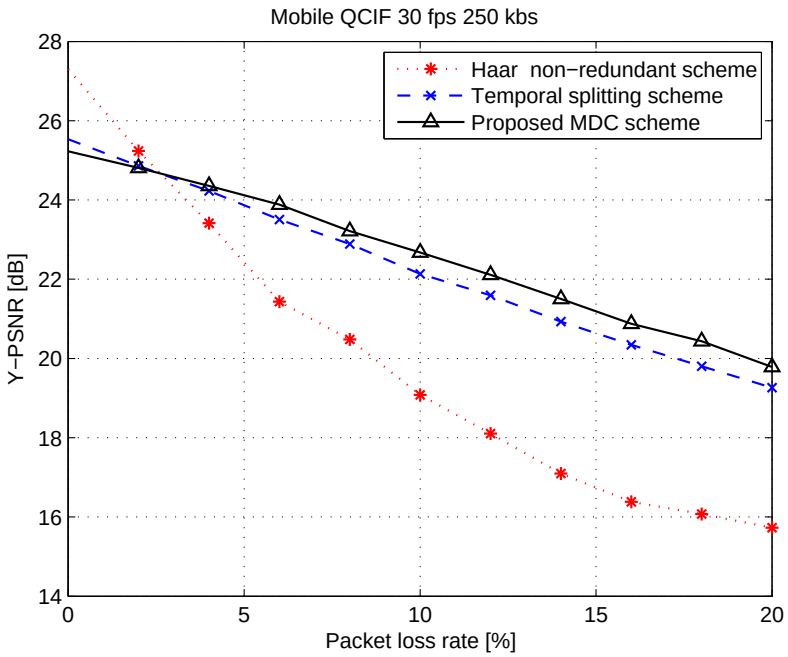


FIGURE 2.28: Distortion vs. packet loss rate ("Mobile" QCIF sequence, 30 fps).

rate curves of the temporal splitting and the proposed MDC schemes for a fixed packet loss rate (10%). One can remark a difference of 0.5-1.3 dB at medium and high bitrates (150-1000 Kbs) and slightly smaller at low bitrates (100 Kbs). It is noticeable that the PSNR of the reconstructed sequence is not monotonically increasing with the bitrate: a stiff increase in PSNR until 250 Kbs is followed by a "plateau" effect which appears at higher bitrates. This is due to the loss of the information in the spatial approximation of the temporal approximation subband. Indeed, for low bitrates, this spatio-temporal subband can be encoded into a single packet, so for uniform error distribution, the rate-distortion curve increases monotonically. At a given threshold (here, it happens at about 250 Kbs for packets of 1500 bytes), the approximation subband has to be coded into two packets. Moreover, we considered that if any of these two packets is lost, the GOF cannot be reconstructed. Therefore, we see a drop in performance. From this point, with the increasing bitrate, the performance improves till a new threshold where the subband needs to be encoded into three packets and so on. A better concealment scheme in the spatial domain, allowing to exploit even a partial information from this subband, would lead to a monotonic increase in performance.

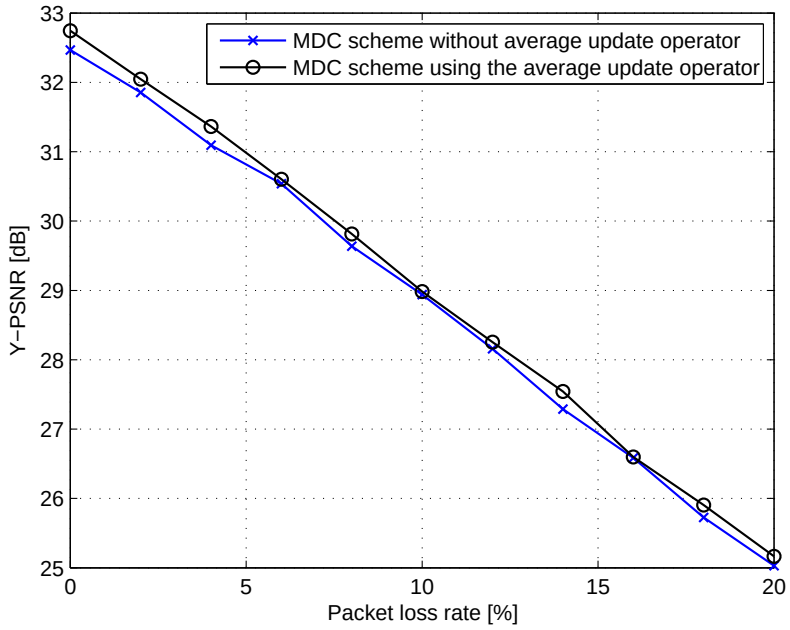


FIGURE 2.29: Influence of average update operator on the performance ("Foreman" QCIF sequence, 30 fps).

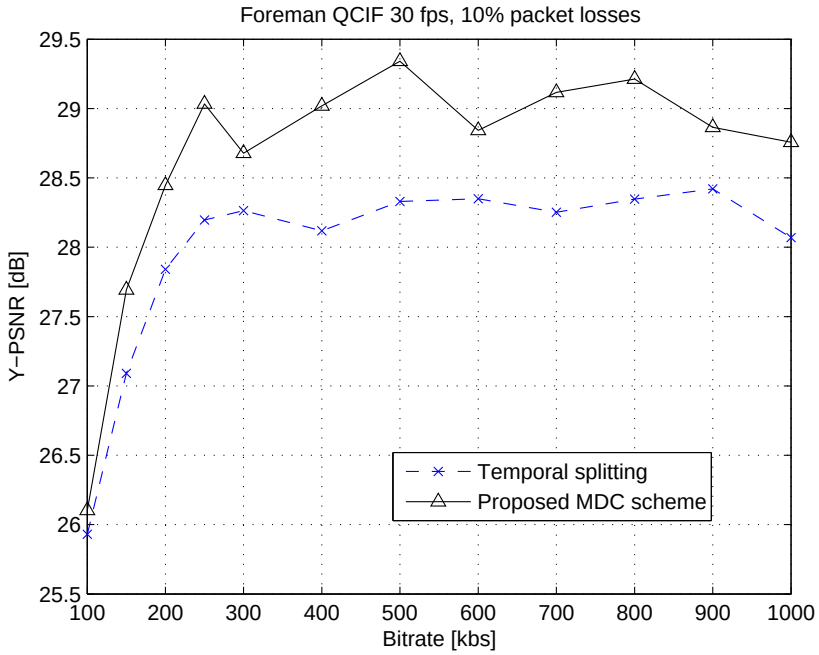


FIGURE 2.30: Y-PSNR vs. Rate curves at 10% packet loss rate ("Foreman" QCIF sequence, 30 fps).

2.9 Conclusions

In this chapter we have presented a study of temporal multiple description coding schemes based on redundant wavelets forming a frame. Using a union of classical orthonormal or biorthogonal wavelet bases at the coarsest resolution of the one-dimensional signal, we first discussed all the possible two-description schemes in this framework and checked their invertibility.

Considering the filter bank equivalence with these wavelet transforms, we have given practical implementation structures for two categories of filter banks, Haar and biorthogonal 5/3, which are compliant to current video coding scenarios involving wavelets. The proposed structures are based on oversampled filter banks and in the Haar case we have equally proven the existence of a 4-band lifting structure enabling the construction of a motion-compensated two-description wavelet scheme. Such an equivalent scheme was not possible to exhibit for the biorthogonal 5/3 filters.

Our study continued with decoding solutions for the schemes that have proven to be potentially interesting in terms of MD encoding. We have thus discussed solutions for the inversion of the oversampled system and we have given optimality criteria for these inverses with respect to the quantization noise.

In the last part of this chapter, we have presented a new multiple description scalable video coding scheme based on a motion-compensated redundant temporal analysis related to Haar and biorthogonal 5/3 wavelets.

The redundancy of the scheme can be reduced by increasing the number of temporal decomposition levels. Reversely, it can be increased either by reducing the number of temporal decomposition levels, or by using non-decimated versions of some of the detail coefficients.

The performances of the proposed MDC schemes have been mainly tested in two scenarios: on-off channels and packet losses. We have compared the performances of the two filter banks in the on-off channels scenario and have chosen the Haar filter bank for the second one based on the compromise it offers in terms of performances versus computational complexity. In the packet-losses scenario we compared our approach to an existing temporal splitting solution.

Note that the presented schemes build the descriptions in the temporal domain of the video, but they can be combined with structures introducing the redundancy in the spatial domain, for which many more solutions have been proposed in the literature. The increased flexibility thus achieved may be exploited to better adapt the packetization to different situations of network losses and also to improve the reconstruction at different levels.

Chapter 3

Spatial MDC schemes

In this chapter we present the extension of our frame-based multiple description schemes to two-dimensional signals, such as still images. If one uses a separable analysis for the classical frame decomposition, this extension is simple. However, as in the temporal case, we propose an additional subsampling of some of the wavelet subbands in order to obtain reduced redundancy multiple description schemes. We saw that the post-transform subsampling operation raises the problem of perfect reconstruction and proven that in the one-dimensional case the subsampled schemes still satisfy the frame property. In the two-dimensional case the structures obtained after the additional subsampling require more attention, first of all because several subsampling strategies (by a factor of two in each spatial direction) are possible, and secondly because not all of the resulting subsets of coefficients form an invertible structure.

In the following, we elaborate a new encoder design in view of obtaining two spatial descriptions having reduced redundancy. We discuss criteria to choose among the possible MDC schemes and evaluate their performance in different transmission scenarios.

Another contribution that we present in this chapter is a post-processing stage at the decoder side based on iterative projections. The quantization information enclosed in the bitstream can be used in order to enhance the reconstruction of the transmitted image in the event of losses - whole description loss or random losses. The technique has some similarities with the one proposed by Chou in [CMW99], which is based on Projections Onto Convex Sets [Roc70], but it considers a more evolved iterative scheme as well as biorthogonal wavelet frames. The proposed method is derived from the general algorithm allowing to minimize a quadratic convex function under convex constraints, which was developed in [Com03].

The work elaborated in this chapter was partially presented in [PPPP05b], [PPPP05a].

3.1 Multiple spatial representations in the wavelet domain

In the following, as in the previous chapter, we start by presenting the wavelet frame decomposition, which uses in this case separable pairs of two-dimensional dyadic filters.

We adapt the notations introduced in Chapter 2 to rejoin those from the classical wavelet decomposition of images, thus denoting the approximation and the three detail subband coefficients at resolution level $j \in \{1, \ldots, J\}$ by a_j, dh_j, dv_j and dd_j, respectively. The second letter in the detail subbands corresponds to their spatial orientation: horizontal, vertical or diagonal.

The letter J is used as before to indicate the resolution of the last level of wavelet decomposition (coarsest resolution).

For the filter bank approach we consider, as before, $(h[n])_{n\in\mathbb{Z}}$ and $(g[n])_{n\in\mathbb{Z}}$ to be the impulse responses of the analysis low-pass and high-pass filters on one of the dimensions. Under the hypothesis of separability, the wavelet subbands in the two-dimensional case are computed by cascading convolutions followed by decimations by a factor of 2:

$$\begin{aligned} a_j[n,m] &= \sum_{k,l} a_{j-1}[k,l]h[2n-k]h[2m-l] \\ dh_j[n,m] &= \sum_{k,l} a_{j-1}[k,l]h[2n-k]g[2m-l] \\ dv_j[n,m] &= \sum_{k,l} a_{j-1}[k,l]g[2n-k]h[2m-l] \\ dd_j[n,m] &= \sum_{k,l} a_{j-1}[k,l]g[2n-k]g[2m-l], \end{aligned} \tag{3.1}$$

at each resolution level, $j \in \{1, \dots J\}$, where a_{j-1} stands for the approximation subband at the upper (finer) resolution.

Our two-description schemes will be based on the classical wavelet decomposition given in Eq. (3.1) up to the next-to-last level of decomposition, $J-1$, and on a un-decimated decomposition at the coarsest resolution level (as in the temporal MDC case). This is obtained by considering a shift in the impulses responses of the filters in one spatial direction at a time or in both.

Note: As in the temporal MDC case, the decomposition onto the shifted wavelet basis can occur at any level of resolution, but, since we want to perform a post-transform subsampling on some of the subbands, then the simplest case in terms of number of possible subband combinations in the two descriptions is obtained with the shifting occurring at the coarsest resolution level, J. In this manner, it is sufficient to consider the invertibility of our redundant schemes at the last level of decomposition in order to asses the global ability of perfect reconstruction of our structures.

Based on these considerations, let us give the equations leading to the wavelet subbands at the coarsest resolution, in an oversampled context.

$$\begin{aligned} a_{J_{(s,s')}}[n,m] &= \sum_{k,l} a_{J-1}[k,l]h[2n+s-k]h[2m+s'-l] \\ dh_{J_{(s,s')}}[n,m] &= \sum_{k,l} a_{J-1}[k,l]h[2n+s-k]g[2m+s'-l] \\ dv_{J_{(s,s')}}[n,m] &= \sum_{k,l} a_{J-1}[k,l]g[2n+s-k]h[2m+s'-l] \\ dd_{J_{(s,s')}}[n,m] &= \sum_{k,l} a_{J-1}[k,l]g[2n+s-k]g[2m+s'-l], \end{aligned} \tag{3.2}$$

where we have pointed out the shift parameters $(s,s') \in \{0,1\}$ introduced in the impulse responses of the analysis filters in each spatial direction.

If we want to link these new notations with those introduced in Chapter 2, then the pair $(s,s') = (0,0)$ will correspond to coefficients obtained from $\mathcal{B}^{\mathrm{I}}$ whereas a pair $(s,s') \neq (0,0)$ indicates that $\mathcal{B}^{\mathrm{II}}$ has been used in at least one spatial direction.

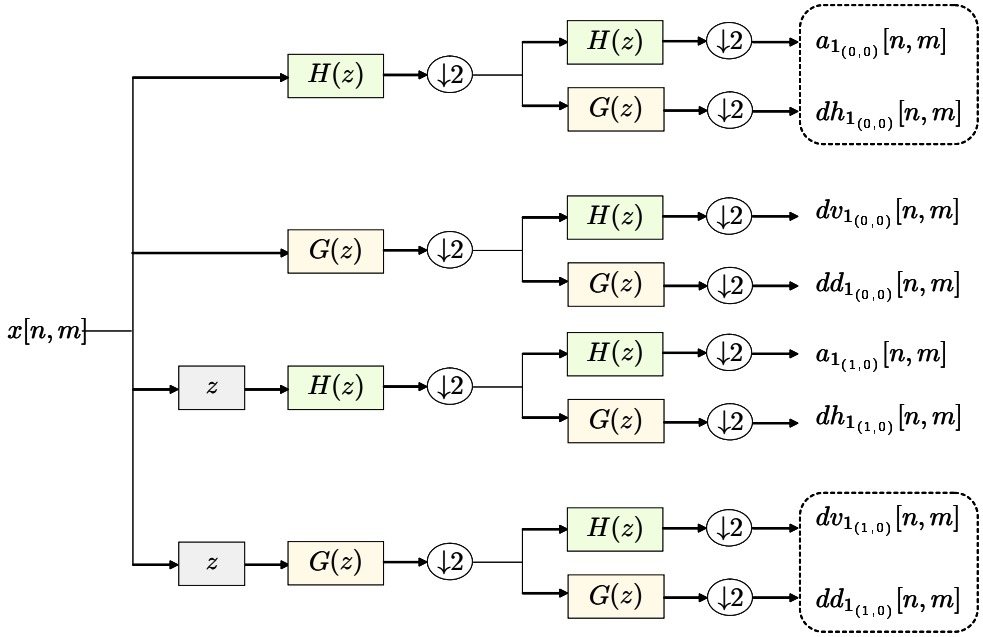


FIGURE 3.1: An example of oversampled filter bank for a separable 1-level wavelet decomposition of the image x (obtained with a shift by one in the horizontal direction of impulse responses of the analysis filters).

In Figure 3.1 we depict an example of two-dimensional oversampled wavelet decomposition on one level and for the following choice of shifting parameters in the impulse responses of the analysis filters: $(s, s') = (0, 0)$ for the first two branches and $(s, s') = (1, 0)$ for the next two. This means that a second wavelet decomposition is obtained from shifted by one impulse responses of the analysis filters in the horizontal direction. In the vertical one, the usual decomposition filters are used.

Two remarks

A first obvious remark is that, as in the 1D-case presented in Chapter 2, the shifted representation is identical to the first one except for the coarsest resolution where we are introducing some information diversity by considering the complementary positions for the decimation operation. By keeping these two representations we have an oversampled scheme that forms a wavelet frame and thus a shift-invariant structure, [PKC96]. Intuitively, such a decomposition amounts to by-passing the decimation step in the transforms at the last level of decomposition, thus leading to four times as many coefficients at the coarsest resolution. Other methods considering frame-based multiple descriptions are giving algorithms for efficient inversion of the system, [GKV99], [BR05]. In [BR05] a restoration stage of the missing coefficients is used before the synthesis oversampled filter bank. Instead of calculating a global pseudo-inverse for the system, the problem is split into blocks and block inverses are computed such that lost samples can be recovered before the actual synthesis stage.

A second remark concerns the chosen decimation positions in the previous approximation subband. Going back to Figure 3.1, we note that any 4-branch subset of the oversampled decomposition, containing one approximation and three detail subbands forms a complete

wavelet basis (in other words any coefficient group given by cascading: $H(z)$ with $H(z)$, $H(z)$ with $G(z)$, $G(z)$ with $H(z)$ and $G(z)$ with $G(z)$, respectively), thus a directly invertible system.

This is highlighted in Figure 3.1 by the dashed rectangles, for a non-"standard" combination of wavelet subbands. We shall retrieve these considerations a little further when we shall present the exhaustive list of interesting invertible combinations in each of the three shifting situations in the impulse responses of the analysis filters. Also, we have found out that some of these non-classical combinations (in terms of the diversity in the inherent subsampling operation associated with the wavelet transform) yield a smaller reconstruction error at the same quantization step, when compared to the homogeneous subsampling, classically proposed.

Toward forming the descriptions

As announced at the beginning of the chapter, we eliminate the overlap between the two representations by down-sampling all the detail coefficients from Eqs. (3.1) and (3.2) on a quincunx grid. Thus, unlike other existing schemes [MG04b], it is worth noticing that with the proposed scheme the redundancy in terms of coefficient number will be limited to the size of an approximation subband at the coarsest resolution.

If we consider only the first two branches in Figure 3.1, an illustration of the quincunx subsampling in the final subbands is shown in Figure 3.2. A detailed explanation of this strategy follows below.

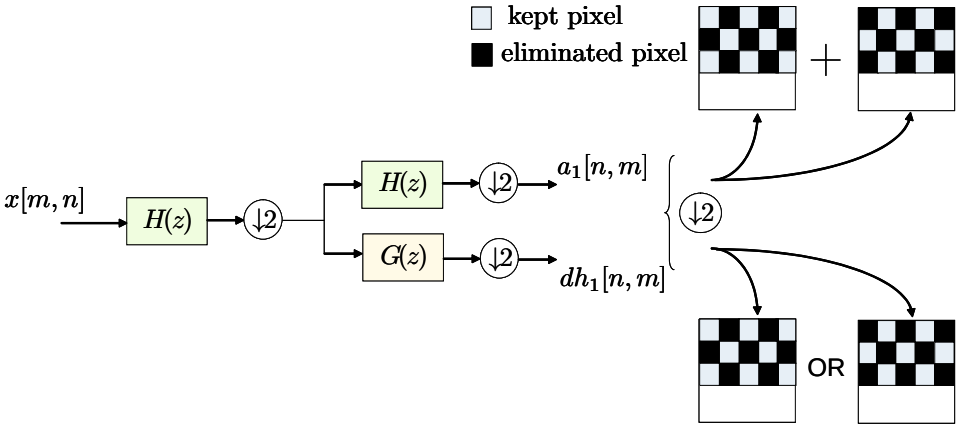


FIGURE 3.2: Illustration of the proposed additional subsampling of the wavelet subbands (note that the resulting approximation subbands are entirely kept in the multiple description scheme, while the detail subbands are quincunx subsampled and only one of the components of this subsampling is kept in the final representation).

In the two-dimensional case also, by discarding some wavelet coefficients we endanger the frame property and thus the perfect reconstruction of our schemes. This issue is addressed a little further and its resolution leads in the same time to a criterion for subband choice in each description.

But for the moment let us explain more thoroughly the forming of the two descriptions in the case of still images applications.

First of all, it is again convenient to use the polyphase representation in order to express the transfer matrix of our multiple description system and study its invertibility.

Recall also that by the quincunx sampling of a 2-D field $(x[n,m])_{n,m}$ one gets two quincunx polyphase components, that can be written as:

$$x^{(q)}[n,m] = x[n+m+q, n-m] \tag{3.3}$$

where $q \in \{0,1\}$.

The polyphase transfer matrix associated to the redundant system based on the impulse responses used in Eqs. (3.1) and in Eq. (3.2), can be written in the frequency domain as a tensor product of the polyphase matrices corresponding to the filter bank operating along one of the dimensions.

Let

$$\boldsymbol{M}_0(\omega) \quad = \begin{bmatrix} H_0(\omega) & H_1(\omega) \\ G_0(\omega) & G_1(\omega) \end{bmatrix}$$

be the polyphase matrix corresponding to the filter bank operating along one of the dimensions. H_0 and H_1 are the two polyphase components of H as follows:

$$H_0(\omega) = \sum_n h_{2n} e^{-\imath n\omega} = \frac{1}{2}[H(\frac{\omega}{2}) + H(\frac{\omega}{2}+\pi)]$$
$$H_1(\omega) = \sum_n h_{2n+1} e^{-\imath n\omega} = \frac{e^{\frac{\imath\omega}{2}}}{2}[H(\frac{\omega}{2}) - H(\frac{\omega}{2}+\pi)]$$

and similar notations are used for G.

For the shifted filters (s or s' equal to 1) we obtain:

$$\boldsymbol{M}_1(\omega) \quad = \begin{bmatrix} H_1(\omega) & e^{\imath\omega} H_0(\omega) \\ G_1(\omega) & e^{\imath\omega} G_0(\omega) \end{bmatrix}.$$

The polyphase matrix for the 2D separable representation in Eq. (3.1) at the coarsest level or in Eq. (3.2) is thus given by the Kronecker tensor product: $\mathcal{M}_{(r,r')}(\omega_x,\omega_y) = \boldsymbol{M}_r(\omega_x) \otimes \boldsymbol{M}_{r'}(\omega_y)$, where $(r,r') \in \{(0,0),(s,s')\}$.

With these considerations, the convolutions followed by decimations from Eqs. (3.1) and (3.2) can be put into the following matrix form:

$$\begin{bmatrix} \mathcal{C}_{(0,0)}(\omega_x,\omega_y) \\ \mathcal{C}_{(s,s')}(\omega_x,\omega_y) \end{bmatrix} = \begin{bmatrix} \mathcal{M}_{(0,0)}(\omega_x,\omega_y) \\ \mathcal{M}_{(s,s')}(\omega_x,\omega_y) \end{bmatrix} \mathcal{X}(\omega_x,\omega_y) \ , \tag{3.4}$$

where we have denoted by $\mathcal{X}(\omega_x,\omega_y)$, the vector of the Fourier transforms of the 4 polyphase components of the approximation coefficients obtained at the next to last level, temporally denoted by $(x[2n-k,2m-l])_{n,m}$ with $(k,l) \in \{0,1\}^2$, for easier reading. The left-hand side term in Eq. (3.4) is the vector of the Fourier transforms of each coefficient subband, $c_{(s,s')}$, at the last level of decomposition in the two representations.

By highlighting the quincunx polyphase components of the coefficients with the quincunx notation from Eq. (3.3), Eq. (3.4) can be rewritten under the form:

$$\begin{bmatrix} \mathcal{C}^{(0)}_{(0,0)}(\omega_x,\omega_y) \\ \mathcal{C}^{(1)}_{(0,0)}(\omega_x,\omega_y) \\ \mathcal{C}^{(0)}_{(s,s')}(\omega_x,\omega_y) \\ \mathcal{C}^{(1)}_{(s,s')}(\omega_x,\omega_y) \end{bmatrix} = \begin{bmatrix} \widetilde{\mathcal{M}}_{(0,0)}(\omega_x,\omega_y) \\ \widetilde{\mathcal{M}}_{(s,s')}(\omega_x,\omega_y) \end{bmatrix} \begin{bmatrix} \mathcal{X}^{(0)}(\omega_x,\omega_y) \\ \mathcal{X}^{(1)}(\omega_x,\omega_y) \end{bmatrix}, \tag{3.5}$$

where, $\mathcal{X}^{(q)}$, $q \in \{0,1\}$, denote the two polyphase components of the Fourier transform of the input signal, $x^{(q)}$. Also, for $(r,r') \in \{(0,0),(s,s')\}$, we have used the notation:

$$\widetilde{\mathcal{M}}_{(r,r')}(\omega_x,\omega_y) = \begin{bmatrix} \widetilde{\mathcal{M}}^{(0)}_{(r,r')}(\omega_x,\omega_y) & \widetilde{\mathcal{M}}^{(1)}_{(r,r')}(\omega_x,\omega_y) \\ \widetilde{\mathcal{M}}^{(1)}_{(r,r')}(\omega_x,\omega_y) & e^{\imath(\omega_x+\omega_y)}\widetilde{\mathcal{M}}^{(0)}_{(r,r')}(\omega_x,\omega_y) \end{bmatrix},$$

with

$$\widetilde{\mathcal{M}}^{(0)}_{(r,r')}(\omega_x,\omega_y) = \frac{1}{2}(\mathcal{M}_{(r,r')}(\nu_x,\nu_y) + \mathcal{M}_{(r,r')}(\nu_x+\pi,\nu_y+\pi))$$
$$\widetilde{\mathcal{M}}^{(1)}_{(r,r')}(\omega_x,\omega_y) = \frac{e^{\imath\nu_x}}{2}(\mathcal{M}_{(r,r')}(\nu_x,\nu_y) - \mathcal{M}_{(r,r')}(\nu_x+\pi,\nu_y+\pi)),$$

where $\nu_x = (\omega_x+\omega_y)/2$ and $\nu_y = (\omega_x-\omega_y)/2$.

In the left-hand side of Eq. (3.5) we end up with a subband coefficient vector having 16 components while the vectors $\mathcal{X}^{(q)}(\omega_x,\omega_y)$, $q \in \{0,1\}$, have 4 components each.

3.1.1 Forming low redundancy descriptions

Let us now describe several multiple description schemes each based on a specific subsampling in Eqs. (3.1) and (3.2), as mentioned before.

We are focusing on equally important descriptions, therefore we shall favor the combinations that contain in each description coefficients from both wavelet representations. According to Figure 3.2, and in order to avoid obtaining too damaged side reconstructions, we preserve one entire approximation subband in each description and distribute the subsampled detail subbands among descriptions, such that the global redundancy of the MDC scheme equals the size of one coarsest-level subband in the wavelet decomposition tree.

In the former section we have shown that the two representations only differ at the coarsest resolution. Each of the two proposed descriptions contains one of the polyphase components of the quincunx sampling of the detail subbands from finer resolutions. The coarsest level in each description will be detailed further.

Considering Eq. (3.3), we build two descriptions as follows:

1. **Description I** is formed by a set of coefficients C_J^{I} defined at resolution level J and the detail subbands $\{dh_j^{(0)},\ dv_j^{(0)},\ dd_j^{(0)}\}$ defined at resolution levels $j \in \{1,\ldots,J-1\}$;
2. **Description II** contains a set of coefficients C_J^{II} defined at resolution level J as well the other quincunx polyphase components of each detail subband: $\{dh_j^{(1)},\ dv_j^{(1)},\ dd_j^{(1)}\}$, for $j \in \{1,\ldots,J-1\}$.

Here, we have denoted by C_J^{I} (resp. C_J^{II}) the set of all subband coefficients at the coarsest resolution in the first (resp. second) description. These sets will be of the form:

$$C_J^{\mathrm{I}} = \{a_{J,(0,0)},\ dh^{(p_1)}_{J,(r_1,r_1')},\ dv^{(p_2)}_{J,(r_2,r_2')},\ dd^{(p_3)}_{J,(r_3,r_3')}\}$$
$$C_J^{\mathrm{II}} = \{a_{J,(s,s')},\ dh^{(p_4)}_{J,(r_4,r_4')},\ dv^{(p_5)}_{J,(r_5,r_5')},\ dd^{(p_6)}_{J,(r_6,r_6')}\}$$

where $p_i \in \{0,1\}$, $i \in \{1,\ldots,6\}$, denotes the selected quincunx polyphase component for the i-th detail coefficient sequence at resolution level J. Also, for all $i \in \{1,\ldots,6\}$, we have either $(r_i, r_i') = (0,0)$ or $(r_i, r_i') = (s, s')$.

Scheme redundancy: Note that if we would keep all the coefficients in the subbands obtained at the coarsest resolution we would introduce a redundancy factor of 2. But, by transmitting only 6 of the 12 polyphase components of the high-pass subbands, we are not introducing any redundancy in the detail coefficients. In this way, the redundancy factor at the coarsest resolution level is limited to $10/8 = 1.25$, or in other words, the overall redundancy of the MDC scheme is of the size of an approximation subband.

We will consider all the possible overcomplete expansions, based on the proposed translated impulse responses of the filters and discuss their perfect reconstruction ability. In the next section we present the framework that led to the following conclusions concerning these possible overcomplete expansions:

- The first one is given by $s = s' = 1$ and will be denoted later on by the index $(1,1)$. In this case only two possible sets of wavelet subbands from the two representations provide perfect reconstruction. These sets are given by the critically sampled decomposition from Eq. (3.1) to which we added the approximation subband from Eq. (3.2) or by the similar structure considering all of the second basis coefficients and the approximation from the first basis as redundancy. In this situation perfect reconstruction is not an issue since both schemes include the critically sampled decomposition. In addition, under i.i.d. hypotheses on the quantization noise, we have determined numerically that these schemes lead to smaller reconstruction errors than the critically sampled decomposition at the same quantization step, as it will be shown later on. In this situation the central decoder does indeed exploit the introduced redundancy in order to increase the quality of the reconstruction.
- The other two possible shifts are a more interesting case. They are given by $s = 1 - s'$. For each of these combinations we obtain at least 12 schemes that can be perfectly recoverable. These 12 schemes also yield a smaller reconstruction error as compared with the critically sampled scheme.

These considerations are detailed further.

3.1.2 Perfect reconstruction issues

By discarding some of the detail coefficients, the global system no longer has a frame structure for all combinations of polyphase components in the detail subbands. It is therefore important to identify the combinations which ensure the perfect reconstruction. To this end we study the invertibility of the polyphase transfer matrix of our system.

Following the strategy discussed in Section 3.1.1, we have to keep only 10 of the resulting 16 quincunx subsampled wavelet subbands: all 4 approximation components and 6 detail ones. Once this choice has been made, let us denote by $\bar{\mathcal{M}}(\omega_x, \omega_y)$ the submatrix of size 10×8 formed by the corresponding selected lines of the polyphase transfer matrix in Eq. (3.5). The perfect reconstruction of the proposed scheme is guaranteed if and only if $\bar{\mathcal{M}}(\omega_x, \omega_y)$ is left-invertible for all $(\omega_x, \omega_y) \in [0, 2\pi)^2$. We designate this matrix as the quincunx polyphase transfer matrix in the following.

3.1.2.1 Invertibility of the polyphase matrix

The left invertibility of the polyphase transfer matrix can be studied by considering its singular value decomposition. A necessary and sufficient condition for perfect reconstruction is that none of the eight so-obtained singular values vanishes on the unit bi-circle. For the three possible combinations of s and s' (different from $(0,0)$) we have studied the evolution on the unit bi-circle of the minimum singular values of each quincunx polyphase transfer matrix corresponding to one of the considered low-redundancy schemes. Considering the shift of the

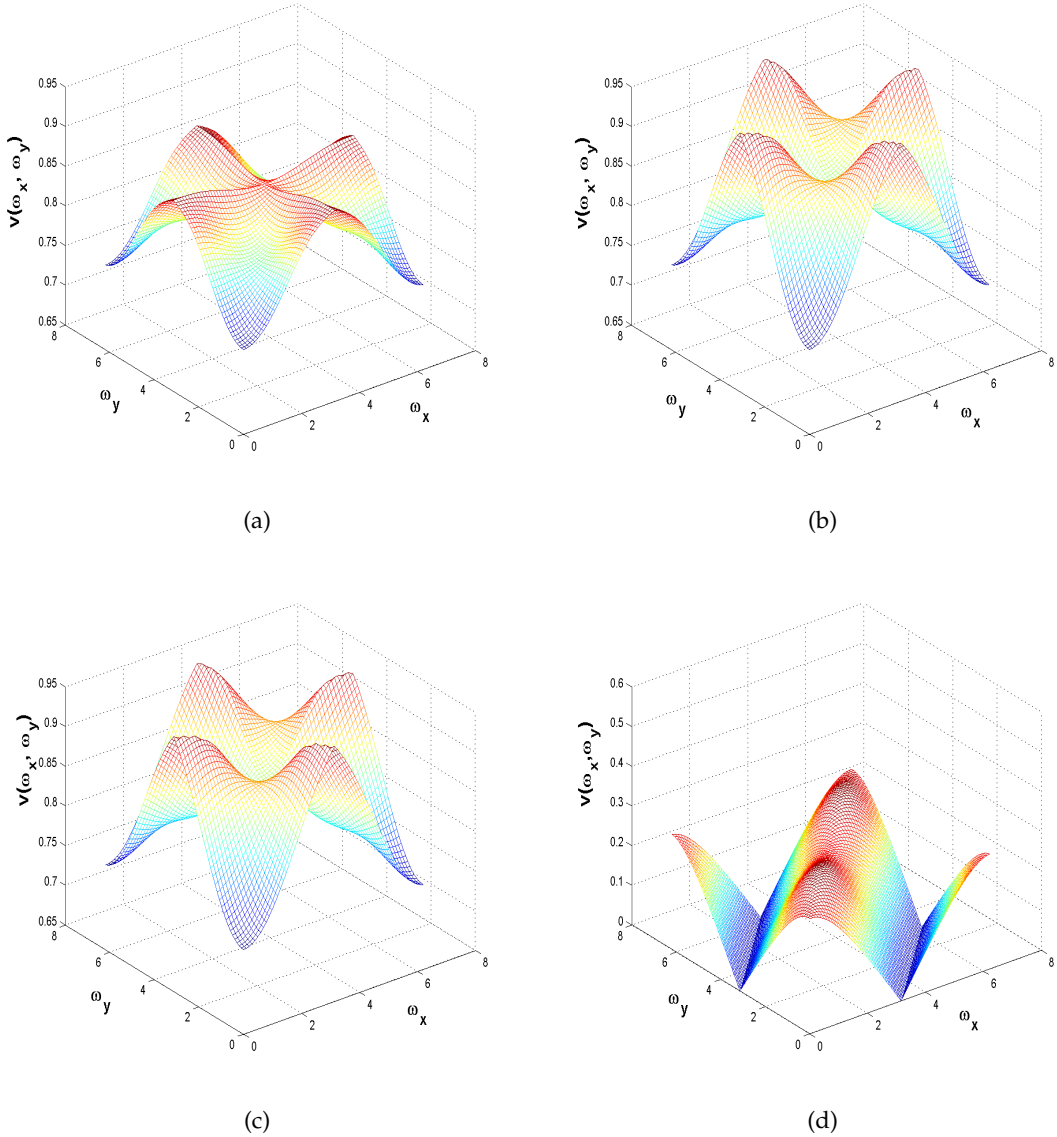


FIGURE 3.3: Minimum singular value, v, of the quincunx polyphase transfer matrix as a function of frequency for schemes: (a) $D^{\mathrm{I}} \cup D^{\mathrm{II}} = \{a_{(0,0)}, dh_{(0,0)}, dv_{(0,0)}, dd_{(0,0)}, a_{(1,1)}\}$, (b) the critically sampled decomposition $\{a_{(0,0)}, dh_{(0,0)}, dv_{(0,0)}, dd_{(0,0)}\}$, (c) $D^{\mathrm{I}} \cup D^{\mathrm{II}} = \{a_{(0,0)}, dh_{(0,1)}, dv_{(0,0)}, dd_{(0,0)}, a_{(0,1)}\}$, (d) one of the combinations that do not yield perfect reconstruction.

filters impulse responses by $(1,1)$, we show in Fig. 3.3 the variation w.r.t. the frequency of the minimum singular value of one of the possible choices for the matrix $\bar{\mathcal{M}}(\omega_x, \omega_y)$. The invertibility of the system is ensured, since for all $(\omega_x, \omega_y) \in [0, 2\pi)^2$ the minimum singular value is nonzero. In Fig. 3.3 we also show a less obvious combination of polyphase quincunx detail subbands, that yields perfect reconstruction, as well as a combination that does not.

The two descriptions forming our scheme will be denoted by D^{I} and D^{II} in Fig. 3.3. Let us now give two examples of MD schemes. A first scheme corresponds to $(s, s') = (1,1)$ and it has the following distribution of the coefficients between the two descriptions at the last level: $D^{\mathrm{I}}_{(1,1)} = \{a_{(0,0)}, dh^{(0)}_{(0,0)}, dv^{(0)}_{(0,0)}, dd^{(0)}_{(0,0)}\}$ and $D^{\mathrm{II}}_{(1,1)} = \{a_{(1,1)}, dh^{(1)}_{(0,0)}, dv^{(1)}_{(0,0)}, dd^{(1)}_{(0,0)}\}$. In this case, the perfect reconstruction that is reflected by Fig. 3.3 can be deduced more directly by observing that the $a_{(1,1)}$ approximation sequence comes in addition to the decomposition onto a basis and thus the overall decomposition is clearly invertible.

A second perfect reconstruction scheme, obtained with $(s, s') = (0,1)$, is formed by the following descriptions: $D^{\mathrm{I}}_{(0,1)} = \{a_{(0,0)}, dh^{(0)}_{(0,1)}, dv^{(0)}_{(0,0)}, dd^{(0)}_{(0,0)}\}$ and $D^{\mathrm{II}}_{(0,1)} = \{a_{(0,1)}, dh^{(1)}_{(0,1)}, dv^{(1)}_{(0,0)}, dd^{(1)}_{(0,0)}\}$. This combination leads to a smaller reconstruction error than in the critically sampled case at the same quantization step, as will be illustrated in the next section.

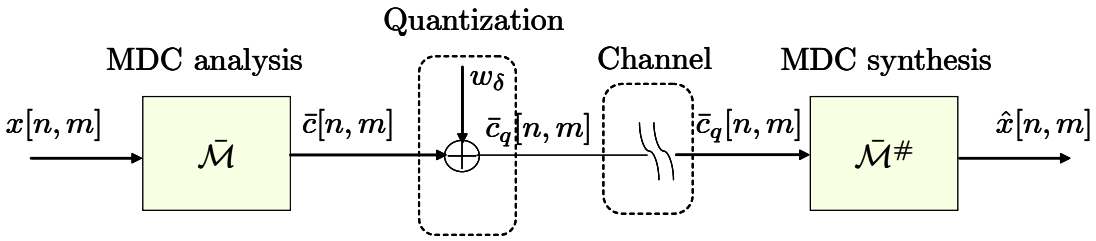


FIGURE 3.4: Example of the transmission chain for one of the MDC schemes given by the transfer matrix $\bar{\mathcal{M}}$, in which we highlight the hypothesis that the quantization can be viewed as an additive noise on the MDC coefficients $\bar{c}[n,m]$.

The study of the singular values of the matrix $\bar{\mathcal{M}}(\omega_x, \omega_y)$ in each situation also provides a means to evaluate the mean squared reconstruction error (MSE), that will be denoted here by $e_{\bar{\mathcal{M}}^{\#}}$. Indeed, if we consider the pseudoinverse of the quincunx transfer matrix, denoted by $\bar{\mathcal{M}}^{\#}$, such that $\bar{\mathcal{M}}^{\#}\bar{\mathcal{M}} = \mathbf{I}_8$, then the reconstruction error that we will be showing to depend on the singular values of $\bar{\mathcal{M}}^{\#}$ can be cast on the singular values of $\bar{\mathcal{M}}$.

Let us also denote the Fourier transform of the coefficient vector representing the chosen combination of subbands that form the two descriptions by $\bar{C}(\omega_x, \omega_y)$. Then, using one of the possible quincunx transfer matrices, $\bar{\mathcal{M}}$, this coefficient vector is obtained as follows:

$$\bar{C}(\omega_x, \omega_y) = \bar{\mathcal{M}}(\omega_x, \omega_y)\mathcal{X}(\omega_x, \omega_y)$$

to which corresponds the spatial domain coefficient vector $\bar{c}[n,m]$.

The so-obtained wavelet coefficients are further uniformly quantized with a step of δ and let us consider that the quantization operation introduces the additive i.i.d. noise w_δ as in Figure 3.4. In other words the quantized coefficients, denoted by $\bar{c}_q[n,m]$ in the spatial domain, are given by:

$$\bar{c}_q[n,m] = \bar{c}[n,m] + w_\delta[n,m]$$

Let us now express the mean squared reconstruction error for the signal x as being $e_x = E\|x - \hat{x}\|^2$, and by considering the frame operator F associated to the polyphase quincunx transfer matrix $\bar{\mathcal{M}}$, and its dual operator denoted by $\tilde{F}$, we get:

$$e_x = E\{\|\tilde{F}\bar{c} - \tilde{F}\bar{c}_q\|^2\} = E\{\|\tilde{F}w_\delta\|^2\} \leq E\{\|\tilde{F}\|^2\|w_\delta\|^2\}$$

If the quantization step, δ, equals one[1], and by expressing w_δ as $w_\delta = \bar{c}_q - \bar{c}$, then $E\{\|w_\delta\|^2\} = \sigma^2_{w_\delta}$ and the reconstruction error will be mainly bounded by the norm of the dual operator $\tilde{F}$ for an i.i.d quantization noise of zero mean and standard deviation of 1. Moreover, we can relate to the norm of the associated inverse quincunx transfer matrix $\bar{\mathcal{M}}^\#$ and thus to the norm of the analysis quincunx transfer matrix $\bar{\mathcal{M}}$. And by using the singular values decomposition of $\bar{\mathcal{M}}$ we can write that:

$$\|\bar{\mathcal{M}}\|^2 = trace(\bar{\mathcal{M}}^t\bar{\mathcal{M}}) = \sum_{i=1}^{8} {v_i}^2,$$

where we have denoted by v_i, $i \in \{1, \ldots, 8\}$ the singular values of $\bar{\mathcal{M}}$.

If we consider N_f points in each spatial frequency direction and compute the singular values of $\bar{\mathcal{M}}$ in all of these points, then it immediately follows that the mean squared error (MSE) generated by the reconstructed signal $\hat{x}$ for a specific MDC scheme given by this transfer matrix can be expressed as:

$$e_{\bar{\mathcal{M}}^\#} = \frac{1}{N_f^2} \sum_{\omega_{x_k}=1}^{N_f} \sum_{\omega_{y_l}=1}^{Nf} \sum_{i \in \{1, \ldots 8\}} \frac{1}{v_i(\omega_{x_k}, \omega_{y_l})^2}, \tag{3.6}$$

where $v_i(\omega_x, \omega_y)$ stands for the i^{th} singular value of $\bar{\mathcal{M}}(\omega_x, \omega_y)$ at each spatial frequency.

The next section presents all the combinations of wavelet subbands, into a two description scheme, that lead to interesting invertible structures from the MSE point of view. By a slight abuse of notations we are designating the MSE computed with Eq. (3.6) by $e_{\bar{\mathcal{M}}^\#}$ for all the MDC schemes, even though each one is given by a different quincunx transfer matrix $\bar{\mathcal{M}}$. We are equally computing in the same way the MSE for a critically sampled decomposition and we are designating it by $e_{\mathcal{M}^\#_{(0,0)}}$.

3.1.2.2 The invertible combinations of subbands

We now verify that a smaller mean square reconstruction error can be obtained with the proposed redundant schemes in the absence of network losses and for the same quantization step at the central decoder.

We have performed this study for three possible frame constructions based on the shifting parameters $(s, s') \in \{(0,0), (1,0), (0,1), (1,1)\}$ in the analysis filters (each one including the pair $(s, s') = (0,0)$) that we have discussed in this chapter.

From a practical point of view, recall that all the possible two-description schemes formed on this wavelet frame are given by the 10×8 minors of the 16×8 global transfer matrix of the system, corresponding to an overcomplete decomposition (or in other words to an undecimated wavelet decomposition). Recall also that the factor 16 is given by the two polyphases of the proposed quincunx subsampling of the wavelet subbands.

[1] We are choosing this stepsize in order to satisfy a high rate approximation for which the quantization noise can be considered as an uniform i.i.d random variable.

However, as we said before, not all of the 10×8 minors of the global polyphase matrix are invertible and, among those that are, not all lead to a smaller reconstruction error than the critically sampled wavelet scheme. Therefore, we identify and present here only those combinations that satisfy both properties.

Let us recall the global wavelet coefficient vector corresponding to the fully overcomplete representation for the scheme given by the shifting parameters $(s, s') \in \{(0,0),(1,0)\}$, in which we have highlighted the quincunx components as before (see Eq. (3.5) in Fourier domain), denoted here by:

$$C = [\underbrace{\mathbf{C}^{(0)}_{(0,0)}}_{1\times4} \quad \underbrace{\mathbf{C}^{(1)}_{(0,0)}}_{5\times8} \quad \underbrace{\mathbf{C}^{(0)}_{(1,0)}}_{9\times12} \quad \underbrace{\mathbf{C}^{(1)}_{(1,0)}}_{13\times16}]^t$$

with

$$\mathbf{C}^{(i)}_{(s,s')} = [A^{(i)}_{(s,s')} \quad Dh^{(i)}_{(s,s')} \quad Dv^{(i)}_{(s,s')} \quad Dd^{(i)}_{(s,s')}]$$

The positions (indexes) of each quincunx polyphase component in the global coefficient vector C of the overcomplete representation are highlighted with the braces under each coefficient block $C^{(i)}_{(s,s')}$.

We are going to present the most interesting combinations of subbands by referring to them through these indexes. These are the combinations that lead to a smaller reconstruction error than a complete wavelet basis (critically sampled and encoded at the same quantization step) and they are given in Table 3.1 for this scheme. Moreover, we want to highlight the indexes corresponding to the kept polyphase quincunx components of the detail subbands from the two representations, therefore we are regrouping in the first four "Subband Index" columns the ones that correspond to the two approximation subbands[2], thus leaving the last six "Subband Index" columns for the indexes of the detail subsets of coefficients - that change with the combination number.

In order to have a clearer idea of the actual subsets of wavelet subbands that correspond to the indices given in Table 3.1, we are equally giving the Table 3.2, in which we have, in addition, highlighted the most efficient detail combinations in terms of reconstruction error.

Note that the smallest reconstruction error in Table 3.2 is obtained for the central decoder combinations which contain whole detail subbands given by a certain analysis filter instead of only halves of them, or in other words only one of the two quincunx polyphase components from each basis (such as the combination $\#2 = \{1,\ 5,\ 9,\ 13,\ 2,\ 3,\ 4,\ 6,\ 8,\ 11\}$, in Table 3.1, corresponding to the coefficients: $\{a_{(0,0)}, dh_{(0,0)}, dd_{(0,0)}, a_{(1,0)}, dv^{(0)}_{(0,0)}, dv^{(0)}_{(1,0)}\}$).

For the critically sampled wavelet basis, which would correspond for instance to the set of indices $\{1,\ 5,\ 2,\ 3,\ 4,\ 6,\ 7,\ 8\}$ in Table 3.1, we obtain $e_{\mathcal{M}^{\#}_{(0,0)}} = 8.1882$ at a quantization step of 1, according to Eq (3.6).

In this manner we retrieve the result announced before, in which non-homogeneous subsampling (with respect to the chosen positions) in the detail subbands as compared to the classical critically sampled decomposition can form complete bases and, if taken with an additional approximation subband, can lead to better reconstruction performances.

[2]Again, the quincunx components of the approximation subbands are explicitly given for the homogeneity of the representation.

No.	Subband Index										$e_{\bar{\mathcal{M}}\#}$
1	1	5	9	13	2	3	4	6	7	8	6.9801
2	1	5	9	13	2	3	4	6	8	11	7.8747
3	1	5	9	13	2	3	4	6	8	15	7.8747
4	1	5	9	13	2	4	6	7	8	11	7.8747
5	1	5	9	13	2	4	6	7	8	15	7.8747
6	1	5	9	13	2	4	6	8	11	15	6.9801
7	1	5	9	13	3	7	10	12	14	16	6.9801
8	1	5	9	13	3	10	11	12	14	16	7.8747
9	1	5	9	13	3	10	12	14	15	16	7.8747
10	1	5	9	13	7	10	11	12	14	16	7.8747
11	1	5	9	13	7	10	12	14	15	16	7.8747
12	1	5	9	13	10	11	12	14	15	16	6.9801

TABLE 3.1: Subband choices (i.e. lines forming the $\bar{\mathcal{M}}_{(s,s')}$ minor in the global polyphase transfer matrix from Eq. (3.5)) for invertible two-description schemes based on the shift parameters $(s, s') \in \{(0,0), (1,0)\}$.

No.	Approximation Subbands		Detail Subbands				$e_{\bar{\mathcal{M}}\#}$
1	$a_{(0,0)}$	$a_{(1,0)}$	$\mathbf{dh}_{(0,0)}$	$\mathbf{dv}_{(0,0)}$	$\mathbf{dd}_{(0,0)}$	-	**6.9801**
2	$a_{(0,0)}$	$a_{(1,0)}$	$dh_{(0,0)}$	$dd_{(0,0)}$	$dv^{(0)}_{(0,0)}$	$dv^{(0)}_{(1,0)}$	7.8747
3	$a_{(0,0)}$	$a_{(1,0)}$	$dh_{(0,0)}$	$dd_{(0,0)}$	$dv^{(0)}_{(0,0)}$	$dv^{(1)}_{(1,0)}$	7.8747
4	$a_{(0,0)}$	$a_{(1,0)}$	$dh_{(0,0)}$	$dd_{(0,0)}$	$dv^{(1)}_{(0,0)}$	$dv^{(0)}_{(1,0)}$	7.8747
5	$a_{(0,0)}$	$a_{(1,0)}$	$dh_{(0,0)}$	$dd_{(0,0)}$	$dv^{(1)}_{(0,0)}$	$dv^{(1)}_{(1,0)}$	7.8747
6	$a_{(0,0)}$	$a_{(1,0)}$	$\mathbf{dh}_{(0,0)}$	$\mathbf{dd}_{(0,0)}$	$\mathbf{dv}_{(1,0)}$	-	**6.9801**
7	$a_{(0,0)}$	$a_{(1,0)}$	$\mathbf{dv}_{(0,0)}$	$\mathbf{dh}_{(1,0)}$	$\mathbf{dd}_{(1,0)}$	-	**6.9801**
8	$a_{(0,0)}$	$a_{(1,0)}$	$dv^{(0)}_{(0,0)}$	$dh_{(1,0)}$	$dd_{(1,0)}$	$dv^{(0)}_{(1,0)}$	7.8747
9	$a_{(0,0)}$	$a_{(1,0)}$	$dv^{(0)}_{(0,0)}$	$dh_{(1,0)}$	$dd_{(1,0)}$	$dv^{(1)}_{(1,0)}$	7.8747
10	$a_{(0,0)}$	$a_{(1,0)}$	$dv^{(1)}_{(0,0)}$	$dh_{(1,0)}$	$dd_{(1,0)}$	$dv^{(0)}_{(1,0)}$	7.8747
11	$a_{(0,0)}$	$a_{(1,0)}$	$dv^{(1)}_{(0,0)}$	$dh_{(1,0)}$	$dd_{(1,0)}$	$dv^{(1)}_{(1,0)}$	7.8747
12	$a_{(0,0)}$	$a_{(1,0)}$	$\mathbf{dh}_{(1,0)}$	$\mathbf{dv}_{(1,0)}$	$\mathbf{dd}_{(1,0)}$	-	**6.9801**

TABLE 3.2: Combinations of subbands at the coarsest resolution, which lead to invertible two-description schemes for the shifting parameters $(s, s') \in \{(0,0), (1,0)\}$ in the analysis filters.

The same tables are presented for the frame given by the analysis filters $h_{(s,s')}$ and $g_{(s,s')}$, with $(s, s') \in \{(0,0), (0,1)\}$ this time. By exploiting the symmetry between the two schemes ($(s, s') = (1,0)$ or $(s, s') = (0,1)$) we can immediately give the following results[3].

The 10×8 minors of the global polyphase transfer matrix that lead to good invertible combinations in terms of MSE at the same quantization step are given by the line combinations presented in Table 3.3. The equivalent combinations for the actually considered wavelet subbands are presented in Table 3.4.

[3]See also Figure 3.1.

No.	Subband index										$e_{\bar{\mathcal{M}}\#}$
1	1	5	9	13	2	3	4	6	7	8	6.9801
2	1	5	9	13	2	3	4	7	8	10	7.8747
3	1	5	9	13	2	3	4	7	8	14	7.8747
4	1	5	9	13	2	6	11	12	15	16	6.9801
5	1	5	9	13	2	10	11	12	15	16	7.8747
6	1	5	9	13	2	11	12	14	15	16	7.8747
7	1	5	9	13	3	4	6	7	8	10	7.8747
8	1	5	9	13	3	4	6	7	8	14	7.8747
9	1	5	9	13	3	4	7	8	10	14	6.9801
10	1	5	9	13	6	10	11	12	15	16	7.8747
11	1	5	9	13	6	11	12	14	15	16	7.8747
12	1	5	9	13	10	11	12	14	15	16	6.9801

TABLE 3.3: Subband choices for invertible two-description schemes based on the shift parameters $(s, s') \in \{(0,0),(0,1)\}$

No.	Approximation Subbands		Detail Subbands				$e_{\bar{\mathcal{M}}\#}$
1	$a_{(0,0)}$	$a_{(0,1)}$	$\mathbf{dh}_{(\mathbf{0},\mathbf{0})}$	$\mathbf{dv}_{(\mathbf{0},\mathbf{0})}$	$\mathbf{dd}_{(\mathbf{0},\mathbf{0})}$	-	**6.9801**
2	$a_{(0,0)}$	$a_{(0,1)}$	$dh^{(0)}_{(0,0)}$	$dv_{(0,0)}$	$dd_{(0,0)}$	$dh^{(0)}_{(0,1)}$	7.8747
3	$a_{(0,0)}$	$a_{(0,1)}$	$dh^{(0)}_{(0,0)}$	$dv_{(0,0)}$	$dd_{(0,0)}$	$dh^{(1)}_{(0,1)}$	7.8747
4	$a_{(0,0)}$	$a_{(0,1)}$	$\mathbf{dh}_{(\mathbf{0},\mathbf{0})}$	$\mathbf{dv}_{(\mathbf{0},\mathbf{1})}$	$\mathbf{dd}_{(\mathbf{0},\mathbf{1})}$	-	**6.9801**
5	$a_{(0,0)}$	$a_{(0,1)}$	$dh^{(0)}_{(0,0)}$	$dv_{(0,1)}$	$dd_{(0,1)}$	$dh^{(0)}_{(0,1)}$	7.8747
6	$a_{(0,0)}$	$a_{(0,1)}$	$dh^{(0)}_{(0,0)}$	$dv_{(0,1)}$	$dd_{(0,1)}$	$dh^{(1)}_{(0,1)}$	7.8747
7	$a_{(0,0)}$	$a_{(0,1)}$	$dv_{(0,0)}$	$dd_{(0,0)}$	$dh^{(1)}_{(0,0)}$	$dh^{(0)}_{(0,1)}$	7.8747
8	$a_{(0,0)}$	$a_{(0,1)}$	$dv_{(0,0)}$	$dd_{(0,0)}$	$dh^{(1)}_{(0,0)}$	$dh^{(1)}_{(0,1)}$	7.8747
9	$a_{(0,0)}$	$a_{(0,1)}$	$\mathbf{dv}_{(\mathbf{0},\mathbf{0})}$	$\mathbf{dd}_{(\mathbf{0},\mathbf{0})}$	$\mathbf{dh}_{(\mathbf{0},\mathbf{1})}$	-	**6.9801**
10	$a_{(0,0)}$	$a_{(0,1)}$	$dh^{(1)}_{(0,0)}$	$dh^{(0)}_{(0,1)}$	$dv_{(0,1)}$	$dd_{(0,1)}$	7.8747
11	$a_{(0,0)}$	$a_{(0,1)}$	$dh^{(1)}_{(0,0)}$	$dv_{(0,1)}$	$dd_{(0,1)}$	$dh^{(1)}_{(0,1)}$	7.8747
12	$a_{(0,0)}$	$a_{(0,1)}$	$\mathbf{dh}_{(\mathbf{0},\mathbf{1})}$	$\mathbf{dv}_{(\mathbf{0},\mathbf{1})}$	$\mathbf{dd}_{(\mathbf{0},\mathbf{1})}$	-	**6.9801**

TABLE 3.4: Combinations of subbands at the coarsest resolution, which lead to invertible two-description schemes for the shifting parameters $(s, s') \in \{(0,0),(0,1)\}$ in the analysis filters.

The only other possible two-descriptions scheme given by the proposed frame decomposition is the one using filters with shifting parameters $(s, s') \in \{(0,0),(1,1)\}$. For this scheme we have obtained only two combinations that lead to a smaller MSE than the critically sampled decomposition. These include the critically sampled decomposition for the corresponding (s, s') analysis filters combination to which the approximation subband for the other basis is added. The MSE obtained in this case is of 6.8901 as compared to 8.1882 for the critically sampled basis at the same quantization step.

In Table 3.5 we give the corresponding transfer matrix minors. As before, this is equivalent to the subband choices given in Table 3.6.

From this study we can conclude that the proposed MDC schemes offer two reconstruction

No.	Subband Index										$e_{\bar{\mathcal{M}}\#}$
1	1	5	9	13	2	3	4	6	7	8	6.9801
2	1	5	9	13	10	11	12	14	15	16	6.9801

TABLE 3.5: Subband choices for invertible two-description schemes based on the shift parameters $(s, s') \in \{(0,0),(1,1)\}$

No.	Approximation Subbands		Detail Subbands			$e_{\bar{\mathcal{M}}\#}$
1	$a_{(0,0)}$	$a_{(1,1)}$	$dh_{(0,0)}$	$dv_{(0,0)}$	$dd_{(0,0)}$	6.9801
2	$a_{(0,0)}$	$a_{(1,1)}$	$dh_{(1,1)}$	$dv_{(1,1)}$	$dd_{(1,1)}$	6.9801

TABLE 3.6: Combinations of subbands at the coarsest resolution, which lead to invertible two-description schemes for the shifting parameters $(s, s') \in \{(0,0),(1,1)\}$ in the analysis filters.

quality levels which can be expected better than the critically sampled decomposition (for the same encoding conditions, under a high rate hypothesis and in the absence of losses). These values of $e_{\bar{\mathcal{M}}\#}$ confirm numerically the inherent symmetries that exist amongst schemes considering a frame representation based on unions of shifted wavelet basis.

For all these schemes that represent good two-description choices we have studied reconstruction strategies based on convex optimization techniques.

3.1.3 Transform implementation considerations

As in the previous chapter, the separable wavelet decomposition can be efficiently implemented via the popular lifting scheme. In the following we give a quick overview of this aspect, highlighting the particular implementation for biorthogonal 9/7 filter banks that we are considering for the application proposed in this chapter.

Since we are in a separable case, we are going to give the announced results only on one of the image dimensions and we are recalling some of the notations that we have used in the temporal MDC schemes[4].

The basic one stage transform of a one-dimensional signal split into odd and even samples can be represented as in Figure 3.5.

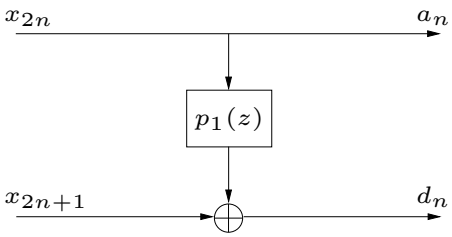


FIGURE 3.5: One stage transform of the input signal.

[4]e.g. the spatial index in either polyphase component of the signal x_{2n}, x_{2n+1} or in the wavelet subbands is becoming a subscript index again.

Then from Figure 3.5 and with the previous notations we have:

$$\begin{aligned} a_n &= x_{2n} \\ d_n &= x_{2n+1} + \sum_k p_1[k] x_{2n-2k} \end{aligned}$$

for one set of approximation and detail coefficients respectively.

For the considered 9/7 biorthogonal filter pairs used in the two representations we choose a lifting factorization [DS98], that can be implemented as in Figure 3.6, where $p_k(z), k \in \{1, \ldots, 4\}$ are Laurent polynomials of the form:

$$\begin{aligned} p_1(z) &= p_1(1+z), \\ p_2(z) &= p_2(1+z^{-1}), \\ p_3(z) &= p_3(1+z), \\ p_4(z) &= p_4(1+z^{-1}), \end{aligned}$$

and the factors $p_i, i \in \{1, \ldots, 4\}$ are computed according to [DS98]:

$$\begin{aligned} p_1 &\approx -1.586134342, \\ p_2 &\approx -0.05298011854, \\ p_3 &\approx 0.8829110762, \\ p_4 &\approx 0.4435068522, \\ p_5 &\approx 1.149604398. \end{aligned}$$

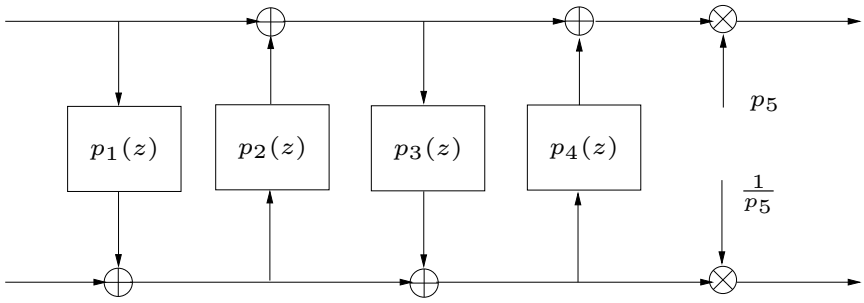


FIGURE 3.6: Lifting formulation of the 9/7 biorthogonal filter pair on one of the dimensions.

In order to compute the adjoint operator, we express the scalar product in which we can

identify the recovered input samples as follows:

$$\begin{aligned}
&\sum_n a_n\tilde{a}_n + \sum_n d_n\tilde{d}_n = \\
&\sum_n \tilde{a}_n x_{2n} + \sum_n \tilde{d}_n(x_{2n+1} + \sum_k p_1(k)x_{2n-2k}) = \\
&\sum_n \tilde{a}_n x_{2n} + \sum_n \tilde{d}_n x_{2n+1} + \sum_n \tilde{d}_n \sum_k p_1(k)x_{2(n-k)} = \\
&\sum_n \tilde{a}_n x_{2n} + \sum_n \tilde{d}_n x_{2n+1} + \sum_n \tilde{d}_n \sum_l p_1(n-l)x_{2l} = \\
&\sum_n \tilde{a}_n x_{2n} + \sum_n \tilde{d}_n x_{2n+1} + \sum_l \sum_n p_1(n-l)\tilde{d}_n x_{2l} = \\
&\sum_n \tilde{a}_n x_{2n} + \sum_n \tilde{d}_n x_{2n+1} + \sum_n \sum_l p_1(l-n)\tilde{d}_l x_{2n} = \\
&\sum_n [\tilde{a}_n + \sum_l p_1(l-n)\tilde{d}_l]x_{2n} + \sum_n \tilde{d}_n x_{2n+1}
\end{aligned}$$

and where we have denoted the dual coefficients by $\tilde{\cdot}$.

The outputs of the adjoint operator are thus given by:

$$\begin{aligned}
\tilde{x}_{2n+1} &= \tilde{d}_n \\
\tilde{x}_{2n} &= \tilde{a}_n + \sum_l p_1(l-n)\tilde{d}_l = \\
&\tilde{a}_n + \sum_l \tilde{p}_1(l-n)\tilde{d}_l
\end{aligned}$$

which can be schematically represented as in Figure 3.7.

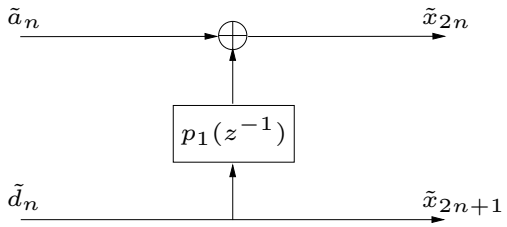


FIGURE 3.7: Adjoint one stage lifting transform of the input signal.

By extension, the dual lifting of the structure in Figure 3.6 is shown in Figure 3.8.

3.1.3.1 Boundary effects in the lifting scheme

Finding the adjoint of the lifting operator as in section 3.1.3 requires some caution on the image borders. If its computation is done in the same manner all over the image, one is confronted with serious boundary effects. To overcome this drawback we apply a special treatment on the borders which is given by:

$$\begin{aligned}
a_n &= x_{2n} \\
d_n &= x_{2n+1} + \sum_k p_1(k,n)x_{2n-2k}
\end{aligned}$$

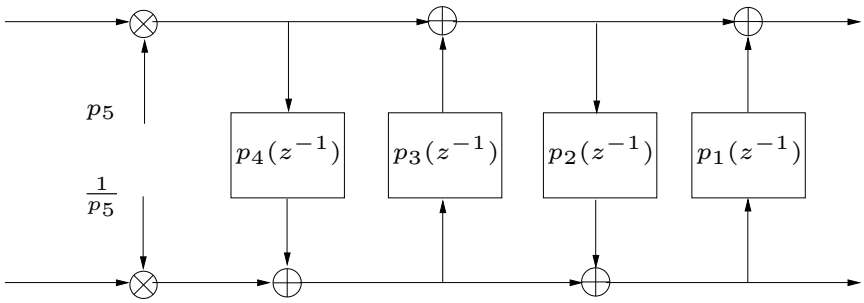


FIGURE 3.8: The dual lifting scheme.

where the corresponding operator also depends on the borders.

Thus the scalar product will be:

$$\sum_n a_n \tilde{a}_n + \sum_n d_n \tilde{d}_n =$$
$$\sum_n \tilde{a}_n x_{2n} + \sum_n \tilde{d}_n (x_{2n+1} + \sum_k p_1(k,n) x_{2n-2k}) =$$
$$\sum_n \tilde{a}_n x_{2n} + \sum_n \tilde{d}_n x_{2n+1} + \sum_n \tilde{d}_n \sum_k p_1(k,n) x_{2n-2k} =$$
$$\sum_n \tilde{a}_n x_{2n} + \sum_n \tilde{d}_n x_{2n+1} + \sum_n \tilde{d}_n \sum_l p_1(n-l,n) x_{2l} =$$
$$\sum_n \tilde{a}_n x_{2n} + \sum_n \tilde{d}_n x_{2n+1} + \sum_l \sum_n \tilde{d}_l p_1(l-n,l) x_{2n} =$$
$$\sum_n x_{2n} [\tilde{a}_n + \sum_l \tilde{d}_l p_1(l-n,l)] + \sum_n \tilde{d}_n x_{2n+1}$$

The outputs of the adjoint operator are therefore:

$$\tilde{x}_{2n} = \tilde{a}_n + \sum_l \tilde{d}_l p_1(l-n,l)$$
$$\tilde{x}_{2n+1} = \tilde{d}_n$$

3.2 Optimized MD reconstruction

At the decoder stage we can enhance the quality of the reconstruction by solving an optimization problem under convex constraints. These constraints are provided by the quantization operation, which can be incorporated as prior knowledge on our system at the decoder. Other constraints such as the light intensity of each pixel can equally be considered.

In the following we are, at first, briefly reviewing some notions of convex optimization mainly through a set theoretic approach and, subsequently, we are presenting our proposed optimization algorithm.

For more in depth information about convex analysis and the related techniques one can refer to [Roc70], [BV04], for instance.

3.2.1 A brief overview on convex set theoretic estimation

We are presenting here the guidelines for a set theoretic approach to convex optimization. This is useful in order to grasp the terminology that we are employing in the next section, as well as the motivations behind the chosen algorithm. We are mainly relying on the theory and results presented in [Com93] and [Com97].

When a signal recovery problem needs to be solved one can resort to convex analysis techniques for instance, provided that a convex formulation can be given.

Consider a signal x in a real Hilbert space, $\mathcal{H}$, endowed with a norm $\|.\|$, that is degraded by some operation (either physical noise or some analysis transform such as Fourier or wavelet followed by some compression operation, for example). In this situation, the optimization problem to solve consists in finding the best possible estimation for x and, most of the time, it is a constrained one.

Choosing an objective function for the optimization problem is not always an easy task. This is due to the fact that, even for simple practical goals, it is sometimes hard to formalize in a strict, mathematical way the optimization criterion. Moreover, the interest in reaching optimality of the solution might sometimes disagree with the subjective perception that governs the expected result. It is known, for instance, that the human eye is not an optimal least-squares detector which makes the choice of an objective function in the sense of squared estimation error a poor candidate even if it solves efficiently the optimization problem. Besides all that, one should make use of the available a priori information about the process. The most reasonable approach to obtain acceptable solutions is to incorporate all available information into the problem formulation. This could lead to an unsolvable optimization problem though, since the constraint information may be very diverse.

In order to determine the acceptability of a solution, the information about the problem should be appropriately classified. Combettes proposes in [Com93] the following classification:

- information about the solution itself (meaning physical constraints that can be imposed on the solution);
- information about the system (properties of the physical system that generated the data) - generally this is incorporated in the problem formulation;
- information about the external factors (model uncertainty, noise etc.).

Suppose that a solution for the optimization problem exists and that it lies in a space $\Xi \subset \mathcal{H}$. The solution space should be chosen as the one that contains the objects directly described by most of the available information. Suppose also that we can define families $(\psi_i)_{i\in I}$ of convex constraints for the objective function, where I is a given finite or countably infinite index set. These constraints define the following closed and convex property sets:

$$(\forall i \in I) \quad S_i = \{a \in \Xi \mid a \text{ satisfies } \psi_i\}.$$

Note that if such a formulation is not possible one can replace the property sets by their convex hull.

The set theoretic estimation approach is given by the pair $(\Xi, (S_i)_{i\in I})$ and is based on finding a point in the so-called "feasibility" set, given by:

$$S = \bigcap_{i\in I} S_i,$$

instead of finding the strict optimum of the problem [Com93]. The set theoretic formulation is said to be *consistent* if and only if the intersection of these property sets is non-void: $S \neq \emptyset$, *fair* if the original signal x lies in S and *ideal* if $S = \{x\}$. Inconsistency arises for instance if mutually exclusive constraints are included. In general, consistency is difficult to check analytically and is often revealed by the convergence behavior of the solution algorithm. Note also that averaging feasible solutions yields a feasible solution in the convex case and this might be useful in order to speed up the convergence of the optimization algorithm.

Feasibility algorithms

Finding a^* in the feasibility set S, has mainly been tackled by the so-called "Projections Onto Convex Sets" (POCS) method, [Roc70], which consists in sequentially projecting an initial estimate of x onto the individual convex property sets, following a periodic schedule, assuming that the index set is finite: $I = \{1, \ldots, M\}$. The point a^* will be in this situation the weak limit of the iterative sequence $(a_n)_{n\geq 0}$ given by:

$$\forall n \in \mathbb{N}, \quad a_{n+1} = P_{(n(mod\ M)+1)_{S_i}}(a_n)$$

with $P_{n_{S_i}}$ designating the projection operator of a_n onto S_i at iteration n.

Recall that, given a non-empty set $A \subset \mathcal{H}$ and a point $z \in \mathcal{H}$, the distance between z and A is given by:

$$d_A(z) = \inf\{\|z - y\| \mid y \in A\}.$$

Moreover, the projection of z onto A, defines the unique point $P_A(z) \in A$ satisfying:

$$\forall y \in A, \quad \langle y - P_A(z) | z - P_A(z) \rangle \leq 0. \tag{3.7}$$

Obviously,

$$\|z - P_A(z)\| = d_A(z).$$

POCS methods encounter some drawbacks however (as explained in [Com97]):

- the convergence of the algorithm is usually quite slow;
- only one set is processed by iteration, therefore parallelizing the algorithm is not possible;
- an exact projection is computed at each iteration, and this might be numerically chaleng-ing;
- the method is limited to problems having a finite number of constraints.

More efficient algorithms, [Com97], can be proposed under some relaxations of the problem formulation and some additional assumptions, as we will see shortly. The main idea in this new optimization strategy can be summarized as follows. In order to find $a^* \in S$ as before, a sequence of signals $(a_n)_{n\in\mathbb{N}}$ is built according to the following steps:

1. At each iteration a series of approximate projections $(P_{i,n}(a_n))_{i\in I_n}$ of (a_n) onto a subfamily of property sets $(S_i)_{i\in I_n}$ with $I_n \subset I$ is computed;

2. The obtained projections are averaged:

$$d_n = \sum_{i \in I_n} w_{i,n} P_{i,n}(a_n).$$

3. The new estimation is built as follows:

$$a_{n+1} = a_n + \lambda_n (d_n - a_n),$$

where λ_n is a relaxation parameter $\lambda_n \in (0, 2L_n)$ and $L_n \geq 1$. The usual values in convex optimization for L_n are lying in $(0, 2)$. With this new relaxation parameter the increased speed for the convergence of the algorithm is ensured.

Other common algorithms encountered in the literature for convex optimization are Kaczmarz's and Cimmino's algorithms, the relaxation method of Agmon-Motzkin-Schoenberg, relaxed POCS, Dijkstra's algorithm and anchor point methods, (see [Gas66], [MS54], [Agm54], [BD86] and other references within [Com93]).

Quadratic signal recovery

We are now going to present some considerations that are closer to the methodology employed in solving our specific reconstruction problem. These paragraphs are based on recent advances in convex programming given in [Com03].

Suppose as before that we have to find an estimation for a signal x in a real Hilbert space, $\mathcal{H}$, endowed with the norm $\|.\|$ and associated scalar product $\langle \cdot, \cdot \rangle$, such that this estimation satisfies all the M convex constraints that can be a priori defined for the system, and suppose that these constraints are quadratic. If we consider the linear operators L_j, and the reference signals r_j, then these constraints can be put in the form:

$$\|L_j x - r_j\|^2 \leq \xi_j, \tag{3.8}$$

where ξ_j are some given bounds. Normally j would take values in $\{1, \ldots, M\}$, and Eq. (3.8) would define the convex property sets, S_j, as before, which in their turn would define the feasibility set $S_M = \bigcap_{j \in \{1,\ldots,M\}} S_j$ for the searched estimation of x.

However, in reality, some of these constraints may not have known bounds. Assume that p of the M constraints have unknown bounds, then the so-called "hard" feasibility set can be defined as the intersection of only the convex sets that are precisely defined:

$$S = \bigcap_{k \in \{p+1,\ldots,M\}} S_k.$$

The convex optimization problem that can be solved in this situation is the following.

Problem 1.
Find $\hat{x} \in S$ such that $J(\hat{x}) = \inf_{x \in S} J(x)$, where the objective function J is defined as a weighted average over the p un-precise quadratic constraints:

$$J(x) = \sum_{j=1}^{p} \alpha_j \|L_j x - r_j\|^2,$$

with $(\alpha_j)_{j\in\{1,\dots,p\}}$ *in* $(0,+\infty)$. *The optimization must be done under the constraints:*

$$\|L_k x - r_k\|^2 \leq \xi_k, \quad k \in \{p+1,\dots,M\}.$$

This problem is thus tantamount to solving a constrained quadratic minimization, which is usually solved iteratively since a closed form solution for it is not guaranteed. From a practical point of view, finding the feasible signal $\hat{x}$ in Problem 1 translates to finding the signal whose linear transformations $L_j x$ are closest, in a least squares sense, to the reference signals r_j, with $j \in \{1,\dots,p\}$.

The methods discussed in [Com03], and on which our proposed algorithm relies, are particularly suitable for large scale applications of signal recovery under convex constraints. First of all, a block iterative parallel structure is defined, which helps to speed up the computation by taking advantage of parallel processing units. In addition, by using local linearizations, the exact enforcement of the constraints is relaxed to an approximate one, thus contributing again to the convergence speed.

Considering again Eq. (3.7) that defines the projection of x onto a closed convex set $A \subset \mathcal{H}$, this is usually achieved through an iterative procedure. In the particular case in which A is in fact a half-space of $\mathcal{H}$ denoted by: $A_h = \{y \in \mathcal{H} \parallel \langle y,u\rangle \leq \eta\}$, with $u \in \mathcal{H}\ \{0\}$ and $\eta \in \mathbb{R}$, the projection $P_{A_h}(x)$ is given explicitly by:

$$P_{A_h}(x) = \begin{cases} x + \frac{\eta - \langle y,u\rangle}{\|u\|^2} u, & \text{if } \langle y,u\rangle > \eta \\ x, & \text{if } \langle y,u\rangle \leq \eta. \end{cases}$$

It is thus useful to be able to express the problem with this approximation.

A geometrical representation of the block-iterative algorithm, that we are reproducing from [Com03], is given in Figure 3.9. The algorithm is initialized with the point x_0 and at each iteration i an outer approximation A_i is built at the intersection of two closed half-spaces that contain the feasibility set S. The ellipses centered in x_0 represent the level curves for the criterion J.

3.2.2 Optimized decoding problem formulation

Let us now discuss the specific application of the above theory to the quality enhancement of the received description(s) at the decoder.

We adopt more concise notations for simplicity:

- $\alpha_i = (\alpha_i[j])_{1\leq j\leq N} = F_i x$, $i \in \{1,2\}$ is the vector of coefficients resulting from the decomposition of an image x onto one of two biorthogonal wavelet bases; the image x is viewed as a column vector in $\mathbb{R}^N$ where N is the number of pixels of x.
- $\hat{\alpha}_i = Q_\delta(\alpha_i)$ denotes the vector of quantized coefficients using a uniform quantizer Q_δ with quantization step $\delta \in \mathbb{R}_+^*$ (the extension to non-uniform/distinct quantizers is immediate). From the received description(s), a subset $(\hat{\alpha}_i[j])_{j\in J_i}$ of quantized coefficients is known, where $J_i \subset \{1,\dots,N\}$.

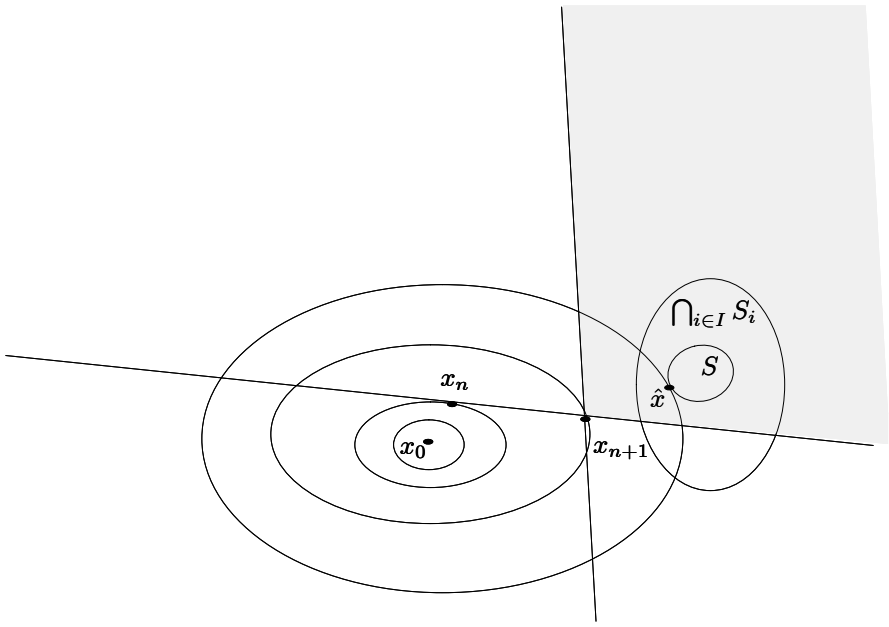


FIGURE 3.9: An illustration for the iterative projections algorithm which minimizes the objective function J over the shaded area by decomposing the global minimization problem into a sequence of simpler minimizations (over outer approximations to the feasibility set S).

Note that, when $J_1 = \emptyset$ or $J_2 = \emptyset$, the reconstruction of x is achieved by directly inverting F_2 or F_1. This situation however never arises for the central decoder and it also happens only in specific cases for the side decoders. Subsequently, we address the general case when both J_1 and J_2 are non empty.

In our case the *a priori* information about our problem is modeled by the quantization constraints:

$$\forall i \in \{1,2\}, \forall j \in J_i, \qquad |\alpha_i[j] - \hat{\alpha}_i[j]| \leq \frac{\delta}{2}. \tag{3.9}$$

The constraints (3.9) define the closed hyperslabs[5]

$$S_{i,j} = \{x \mid \ -\frac{\delta}{2} \leq f_i[j]^\mathsf{T} x - \hat{\alpha}_i[j] \leq \frac{\delta}{2}\}$$

where $f_i[j]$ is the j-th basis function of the i-th representation (the j-th column of the matrix F_i). The decoded image should therefore belong to the closed convex set $S = \bigcap_{i\in\{1,2\}} \bigcap_{j\in J_i} S_{i,j}$.

Let x_0 be a reference image we expect the decoded image to be close to. Such a reference image may correspond to an initial estimate of the original image.

The decoding problem can be cast as:

Problem 2.
Find $\hat{x} = \arg\min_{x\in S} \|x - x_0\|$, *where* $\|.\|$ *denotes the euclidean norm of* $\mathbb{R}^N$.

This means that $\hat{x}$ is the projection of x_0 onto S. As we have mentioned in the previous section, usually, no closed form expression of $\hat{x}$ exists and an iterative optimization algorithm needs to be used to compute it.

[5] A hyperslab is defined as the set $\{a \in \Xi \,|\, \alpha_i \leq f_i(a) \leq \beta_i\}$ for a non-zero continuous real functional f_i defined over a vector space Ξ and $\alpha_i, \beta_i \in \mathbb{R}$

Before presenting such an algorithm, we note that the projection onto each set $S_{i,j}$ is easily expressed as

$$P_{S_{i,j}}(x_0) = x_0 - \gamma_i[j]\frac{f_i[j]}{\|f_i[j]\|}$$

where $\gamma_i = T_i(\alpha_{i,0})$ with $\alpha_{i,0} = F_i x_0$. Each operator T_i is such that, for all $j \notin J_i$, $\gamma_i[j] = 0$ and, for all $j \in J_i$,

$$\gamma_i[j] = \begin{cases} \dfrac{\alpha_{i,0}[j] - \frac{\delta}{2} - \hat{\alpha}_i[j]}{\|f_i[j]\|}, & \text{if } \alpha_{i,0}[j] > \frac{\delta}{2} + \hat{\alpha}_i[j] \\ 0, & \text{if } |\alpha_{i,0}[j] - \hat{\alpha}_i[j]| \leq \frac{\delta}{2} \\ \dfrac{\alpha_{i,0}[j] + \frac{\delta}{2} - \hat{\alpha}_i[j]}{\|f_i[j]\|}, & \text{if } \alpha_{i,0}[j] < -\frac{\delta}{2} + \hat{\alpha}_i[j]. \end{cases}$$

A similar multiple description scheme for images has been developed in [CMW99] and it uses overcomplete expansions obtained with frame decompositions obtained by concatenation of two invertible decompositions as in our case. This method also is inspired on the works in [GKAV98] and [GKV99]. In this MDC approach however, each of the coefficients resulted from the frame decomposition is viewed as an individual description, contrary to our strategy in which we form descriptions of mixed coefficients. Another difference with our scheme pertains to the amount of introduced redundancy. In [CMW99] the redundancy is of a factor of 2 (since two critically sampled decompositions are involved in the overcomplete expansion), whereas in our case the redundancy is tunable with the number of decomposition levels, being for instance of 1.0156 for a 3-level wavelet decomposition in each description (or of 25% more coefficients at the coarsest resolution level as we have seen in Section 3.1).

At the reconstruction a POCS algorithm is proposed in [CMW99]. This algorithm, however, uses the global projections onto S_i and not onto $S_{i,j}$ and this is only possible for orthogonal decompositions. In our case a POCS implementation is not possible since the number of convex sets involved is too high, and this is the motivation behind using the proposed algorithm that we will describe in the sequel, after giving the general iterative algorithm that we have adapted to our given framework. Moreover, as we have said before, the main shortcomings of a POCS method are the fact that it does not converge to the best approximation of x_0 in S and also its convergence is slow [Com97].

3.2.3 Iterative projections algorithm

The proposed method is derived from the general algorithm allowing to minimize a quadratic convex function under convex constraints, which was developed in [Com03]:

Algorithm 1 (Combettes, 2003)**.**

① *Fix* $\epsilon \in (0, 1)$. *Set* $x_0 = r$ *and* $n = 0$.

② *Take a nonempty finite index set* $I_n \subset I$.

③ *Set* $z_n = x_n + \lambda_n R^{-1}(\sum_{i \in I_n} w_{i,n} p_{i,n} - x_n)$, *where:*

a) R is defined as: $R = \sum_{j \in \{1,\ldots,p\}} \alpha_j L_j^* L_j$ *and* L_j^* *stands for the adjoint operator of* L_j *(with* L_j *and the index set* $\{1, \ldots, p\}$ *as in the previous section);*

b) For every $i \in I_n$, $p_{i,n}$ is a subgradient[6] *projection given by:*

$$p_{i,n} = \begin{cases} x_n - \frac{f_i(x_n)t_{i,n}}{\|t_{i,n}\|^2}, & \text{if } f_i(x_n) > 0 \\ x_n, & \text{if } f_i(x_n) \leq 0 \end{cases}$$

with $f_i : \mathcal{H} \to \mathbb{R}$ a continuous and convex function and $t_{i,n} \in \partial f_i(x_n)$[7] *a subgradient;*

c) The weights $(w_{i,n})_{i\in I_n}$ belong to $[0, 1]$ and satisfy $\sum_{i\in I_n} w_{i,n} = 1$;

d) The relaxation parameter λ_n belongs to $[\epsilon L_n, L_n]$, with L_n given by:

$$L_n = \begin{cases} \frac{\sum_{i\in I_n} w_{i,n}p_{i,n}\|p_{i,n}-x_n\|^2}{\langle R^{-1}u_n \mid u_n\rangle}, & \text{if } x_n \notin \bigcap_{i\in I_n} S_i \\ \frac{1}{\|R^{-1}\|}, & \text{otherwise.} \end{cases}$$

④ *Set: $\pi_n = \langle x_0 - x_n \mid x_n - z_n\rangle$, $\mu_n = \|x_0 - x_n\|^2$, $\nu_n = \|x_n - z_n\|^2$ (all w.r.t R) and $\rho_n = \mu_n\nu_n - \pi_n^2$. Compute:*

$$x_{n+1} = \begin{cases} z_n, & \text{if } \rho_n = 0 \text{ and } \pi_n > 0 \\ x_0 + (1 + \frac{\pi_n}{\nu_n})(z_n - x_n), & \text{if } \rho_n > 0 \text{ and } \pi_n\nu_n \geq \rho_n \\ x_n + \frac{\nu_n}{\rho_n}(\pi_n(x_0 - x_n) + \mu_n(z_n - x_n)), & \text{if } \rho_n > 0 \text{ and } \pi_n\nu_n < \rho_n. \end{cases}$$

⑤ *Set $n \leftarrow n + 1$ and go to ②.*

Since in our case we are using only differentiable operators the subgradient computations amount to gradients. Our proposed algorithm derived from the Algorithm 1 is given in the sequel. Since we are mainly handling 8-bit images we will introduce an additional constraint to keep the computed samples in the range 0 to 255. This will be done by simple projection.

Algorithm 2.

① *Set $k = 0$ and choose an initial estimate x_0*[8].

② *Calculate $\alpha_{i,k} = F_i x_k$, $i \in \{1, 2\}$.*

③ *Set $\gamma_{i,k} = T_i(\alpha_{i,k})$ and $\lambda_{i,k} = \|\gamma_{i,k}\|^2$, for $i \in \{1, 2\}$. Calculate $\tilde{\gamma}_{i,k} = (\gamma_{i,k}[j]/\|f_i[j]\|)_j$.*

④ *Set $a_{i,k} = -F_i^{\mathsf{T}}\tilde{\gamma}_{i,k}$, $i \in \{1, 2\}$*

⑤ *Set $a_{pm} = x_k$. Find outlier values (below 0 and above 255) in a_{pm} and put them in masks m_p and m_m. Set* $a_{pm} = \begin{cases} -x_k & \forall m_p > 0, \\ 255 - x_k & \forall m_m > 0, \\ 0 & \text{otherwise.} \end{cases}$

⑥ *Set $\omega_{1,k} = \frac{2\lambda_{1,k}}{3(\lambda_{1,k}+\lambda_{2,k})}$, $\omega_{2,k} = \frac{2\lambda_{2,k}}{3(\lambda_{1,k}+\lambda_{2,k})}$, and $\omega_{pm} = \frac{1}{3}$. Compute $L_k = \omega_{1,k}\lambda_{1,k} + \omega_{2,k}\lambda_{2,k} + \omega_{pm}\|a_{pm}\|^2$*

[6] A vector t is called a subgradient projection of a function g in a vector space Ξ at a point a if the continuous affine functional $f_{a,t}$ which is given by $b \mapsto \langle b - a\ t\rangle + g(a)$, having the "slope" t takes the same value as g in a and minorizes g on Ξ

[7] $(\partial f_i(x))_{i\in I}$ is called the subdifferential of f_i in x

[8] possible choices of x_0 will be presented bellow.

⑦ *If* $L_k = 0$*, exit iteration.* $v_k = \omega_{1,k}a_{1,k} + \omega_{2,k}a_{2,k} + \omega_{pm}a_{pm}$ *and* $d_k = \frac{L_k}{\|v_k\|^2}v_k$.

⑧ *Set* $b_k = x_0 - x_k$, $\pi_k = -b_k^{\mathsf{T}}d_k$, $\mu_k = \|b_k\|^2$, $\nu_k = \|d_k\|^2$, *and* $\rho_k = \mu_k\nu_k - \pi_k^2$.

⑨ *Set* $x_{k+1} = \begin{cases} x_k + d_k & \text{if } \rho_k = 0 \text{ and } \pi_k \geq 0, \\ x_0 + (1 + \pi_k/\nu_k)d_k & \text{if } \rho_k > 0 \text{ and } \pi_k\nu_k \geq \rho_k, \\ x_k + \dfrac{\nu_k}{\rho_k}\left(\pi_k b_k + \mu_k d_k\right) & \text{if } \rho_k > 0 \text{ and } \pi_k\nu_k < \rho_k. \end{cases}$

⑩ *Set* $k \leftarrow k + 1$ *and go to* ②.

It can be noted that the adjoint operators F_i^{T} involved in step ④ can be implemented by using filter bank structures and the developments on the lifting scheme in Section 3.1.3. Another interesting characteristic of this algorithm is that the computations in ②, ③ and ④ can be parallelized on a bi-processing unit.

3.2.4 Reference image in the objective function

Before discussing the performances of the proposed iterative algorithm we need to make some remarks on the reference image, x_0, used in the first run of step ②. This corresponds to a first estimate of the original image obtained from the decoding of the two descriptions. There are however many ways to combine this recovered information and we are dealing with this subject in this section. This discussion is necessary because the proposed iterative algorithm is sensitive to the choice of the reference image used in the objective function. It turns out that the best performances are achieved by choosing x_0 as the global minimizer for the unconstrained optimization Problem 2.

In the following we are presenting several strategies that lead to the initialization image for the Algorithm 2, from the simplest to the most complex.

We start with a simple weighted mean reconstruction from the two descriptions to which we are referring as WMR in the sequel. This is indeed the simplest and quickest initialization that one can imagine but it leads to poorer performances than those theoretically expected for the optimization method and it has motivated our research for other reference images.

3.2.4.1 WMR from two descriptions

In this first, very simple strategy, the two received descriptions are individually recovered and the obtained images are averaged in order to give an initialization point for the iterative algorithm. However, one must take into account the fact that in some of the invertible cases the distribution of subbands resulting from one or the other wavelet basis is not balanced. Therefore a weighted average would be a better solution for combining the two reconstructions. We shall refer to this technique as WMR1 from now on. Let the two reconstructed images from each description be denoted by $\widehat{x}_i$, with $i \in \{1, 2\}$:

$$\widehat{x}_i = F_i^{-1}\widehat{\alpha}_i$$

where F_i^{-1} is the synthesis operator for each of the two representations and $\hat{\alpha}_i$ stands for the quantized coefficients in each description, as before (Section 3.2.2).

Thus, if we denote by $\widehat{x}_m$ the average reconstruction from $\widehat{x}_1$ and $\widehat{x}_2$ we get:

$$\widehat{x}_m = p_1\widehat{x}_1 + p_2\widehat{x}_2$$

where the weights are given by the energy of the reconstruction:

$$p1 = \frac{||\widehat{x}_1||^2}{||\widehat{x}_1||^2 + ||\widehat{x}_2||^2}$$

and $p_2 = 1 - p_1$.

The starting point x_0 in the presented algorithm is the so-computed $\widehat{x}_m$. As we shall see after having presented the other strategies, this one has very poor performances as compared to a critically sampled decomposition and it systematically contradicts the results presented in Section 3.1.2.2 for the expected mean square error at the same quantization step as the non-redundant decomposition.

We have thus extended this averaging method by considering a least squares algorithm giving the optimal weights for each one of the two reconstructions from the descriptions.

3.2.4.2 WMR with least squares (LS) weights

In this new method we thus compute the weights for each coefficient set with a least squares algorithm.

Let us denote the initialization image in this case by $\widehat{x}_0$.

We have to minimize:

$$\mathrm{E}\{|x(\mathbf{n}) - \widehat{x}_0(\mathbf{n})|^2\}$$

for any $\mathbf{n} \in \mathbb{Z}^2$, by setting $\widehat{x}_0(\mathbf{n}) = \boldsymbol{p}_1\widehat{x}_1(\mathbf{n}) + \boldsymbol{p}_2\widehat{x}_2(\mathbf{n})$.

This amounts to minimizing:

$$\mathrm{E}\{|x(\mathbf{n}) - p_1\widehat{x}_1(\mathbf{n}) - p_2\widehat{x}_2(\mathbf{n})|^2\}$$

We thus have to solve the following system of equations:

$$\begin{bmatrix} \mathrm{E}\{\widehat{x}_1(\mathbf{n})\}^2 & E\{\widehat{x}_1(\mathbf{n})\widehat{x}_2(\mathbf{n})\} \\ \mathrm{E}\{\widehat{x}_2(\mathbf{n})\widehat{x}_1(\mathbf{n})\} & E\{\widehat{x}_2(\mathbf{n})\}^2 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} = \begin{bmatrix} \mathrm{E}\{x(\mathbf{n})\widehat{x}_1(\mathbf{n})\} \\ \mathrm{E}\{x(\mathbf{n})\widehat{x}_2(\mathbf{n})\} \end{bmatrix}$$

By choosing the reconstruction weights in this manner we observe experimentally that for some of the schemes the weights computed in the last section were the optimal ones in a least squares sense, therefore a real improvement cannot be expected for all the possible invertible combinations of subbands.

Therefore a further enhancement of the initialization has been proposed and it is given in the following. This is also motivated by the fact that the global multiple description structure introduces a lot of zero coefficients when each of the descriptions is taken individually, thus biasing the computation of the global reconstruction weights for the initialization.

3.2.4.3 Selective WMR - LS weights

This initialization can be further extended by computing least squares weights for each of the received wavelet subbands in the descriptions. This is referred to as selective WMR in this section in order to highlight the fact that the weights depend only on the subbands which are non-zero in the global MD representation. Note that according to the chosen scheme a subband can either contain all of the samples or only half of them, when only one of the two possible quincunx components where included in the MDC scheme. The number of weights shall thus vary according to these two situations. More precisely, regarding the coarsest decomposition level in which subbands may no longer be the outcome of a single basis decomposition, we have a total of 10 halves of subbands which, in the central reconstruction, are either grouped into 5 whole subbands (two approximation subbands and 3 detail subbands) or into 4 whole subbands (two approximations and two details) and 2 half-subbands (if the remaining quincunx subsampled details were not the outcome of a single basis decomposition).

In this situation the initialization of the iterative algorithm can be written as:

$$\widehat{x}_0 = \sum_{k=1}^{K} w_k \widehat{r}_k$$

with K being the maximum number of kept subbands in the two descriptions at the last level of decomposition, and $\widehat{r}$ the reconstruction from a single subband only, when all the other coefficients are set to zero.

Note that, based on the considerations given above (as well as in section 3.1.2.2) the number of weights to be computed is not higher than 6 in any of the favorable invertible schemes.

Transmission considerations

One major drawback of this method is that, in order to compute the least squares optimal weights, the original image must be known. Therefore, this computation must take place at the encoder side, since this knowledge is no longer available at the decoder. By doing that, additional information must be included in the bitstream containing these weights. Even though this does not encumber the transmission, the option of transmitting side information might not be available for all codecs by default. This operation obviously reduces the flexibility of our scheme which is something we would like to avoid. A solution that overcomes this difficulty is to set these K weights at fixed values (for example, estimated off-line on a large database) and use these values at the decoder for any image. This leads us to a new strategy, detailed in the following.

3.2.4.4 Selective WMR from each of the received subbands - fixed weights

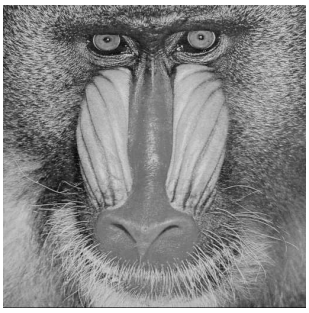
In order to chose fixed values for the subband reconstruction weights we need to verify that the obtained values do not vary a lot from one image to another, at the same quantization step. Therefore we have computed those weights for several images of different types: natural, synthetic, satellite, biomedical, astronomical. Some of the most common test images in this database[9] are given in Fig 3.10.

[9] The size of the considered data base is of 15 images.

(a) Lena

(b) Barbara

(c) Pentagon



(d) Mandrill

(e) Boat

(f) Moon

(g) Goldhill

(h) Aerial

(i) Montage

(j) Paris Spot 5 (512×512 fragment)

FIGURE 3.10: Ten of the test images used to compute the LS weights for the initialization based on individual subband reconstruction (512×512 pixels).

In Table 3.7 we can see the evolution of these values for ten of the database images at a quantization step of 1, for the MDC combination $(s, s') \in \{(0,0),(0,1)\}$ on three levels of decomposition.

Image	$w_1 a_{(0,0)}$	$w_2(dv_{(0,0)})$	$w_3(dd_{(0,0)})$	$w_4(a_{(0,1)})$	$w_5(dh_{(0,1)})$
Lena	0.0000	0.9998	1.0000	1.0000	1.0000
Barbara	0.0001	0.9999	0.9996	0.9998	1.0000
Pentagon	0.0002	1.0000	1.0000	0.9998	1.0000
Mandrill	-0.0001	1.0001	1.0000	1.0001	1.0000
Boat	-0.0000	1.0001	0.9999	1.0000	1.0000
Moon	0.0001	0.9991	1.0001	0.9999	0.9999
Goldhill	0.0001	1.0000	0.9996	0.9999	0.9999
Aerial	0.0001	1.0000	1.0001	0.9999	0.9999
Montage	0.0000	1.0001	0.9995	1.0000	1.0000
Paris Spot 5	0.0000	1.0000	1.0000	1.0000	1.0000

TABLE 3.7: Least squares weights for several test images uniformly quantized with the step of 1

Quantization step	$w_1(a_{(0,0)})$	$w_2(dv_{(0,0)})$	$w_3(dd_{(0,0)})$	$w_4(a_{(0,1)})$	$w_5(dh_{(0,1)})$
1	0.0001	1.0000	0.9996	0.9999	0.9999
4	0.0019	0.9996	0.9981	0.9981	0.9982
16	0.0198	0.9903	0.9174	0.9802	0.9816
32	0.0625	0.9683	0.8934	0.9372	0.9424
64	0.1851	0.8948	0.7849	0.8153	0.8402
128	0.3487	0.8047	0.6925	0.6525	0.6898

TABLE 3.8: The evolution of the computed least squares weights on each subband for the image "Goldhill"

As it can be seen, choosing the 5-tuple $(0, 1, 1, 1, 1)$ for the combination of subbands $\{a_{(0,0)}, dv_{(0,0)}, dd_{(0,0)}, a_{(0,1)}, dh_{(0,1)}\}$ as weights corresponding to each subband is a reasonable approximation. Moreover the variation of weights across the different quantization steps is sufficiently small as it can be seen in Table 3.8 for the image "Goldhill", for instance.

Note that these weights confirm the remark we have made in the beginning of this chapter, concerning the fact that the subsampling positions in the critically sampled transform do not alter the basis property and the perfect reconstruction in the absence of quantization. In the above situation our computation has confirmed the fact that the set of wavelet subbands $a_{(0,1)}, dv_{(0,0)}, dd_{(0,0)}, dh_{(0,1)}$ suffices for optimal reconstruction in the absence of quantization.

3.3 Case study

In order to draw some conclusions about the proposed decoding enhancements let us look at a case study for the test image "Mandrill" (512×512 pixels). We are considering two scenarios: the first one represents the ideal MDC functioning with on/off channels and the second one considers the case of random (or deterministic) losses inside both descriptions. In this case study we do not concentrate on the encoder used in order to determine the bitrate points in the PSNR vs. Rate curves. The encoding algorithm that we are using will be discussed in the next section.

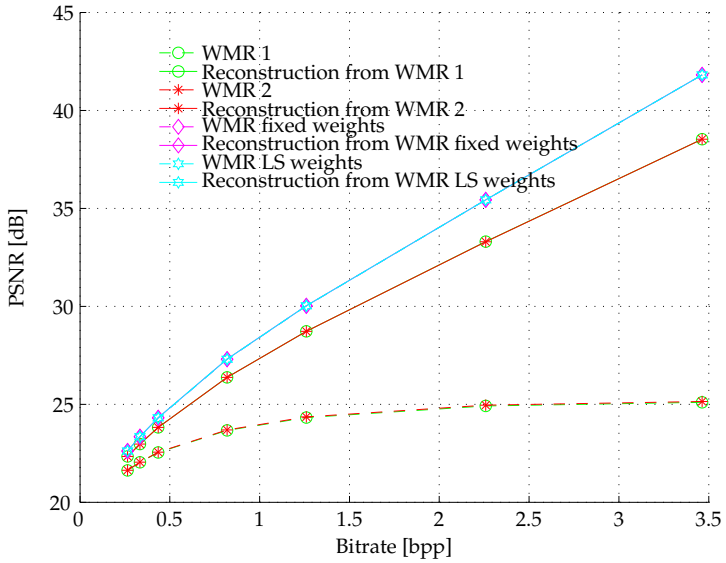


FIGURE 3.11: Central decoder comparison between the different iterative reconstructions for the image "Mandrill".

3.3.1 On/off channels - loss of an entire description

In this scenario we are aiming to offer choice criteria for the optimized decoding.

Let us consider the MD scheme given by $D^{\mathrm{I}} \cup D^{\mathrm{II}} = \{a_{(0,0)}, dv_{(0,0)}, a_{(1,0)}, dh_{(1,0)}, dd_{(1,0)}\}$. The central decoder in this situation is presented in Figure 3.11 for the different methods of initialization presented in Section 3.2.4. Here we have denoted by $WMR1$ and $WMR2$ the weighted mean reconstructions (initialization points) for which there are only two weights to be computed, one for each description. These correspond to the first two sections in Section 3.2.4. In this situation we confirm what we have announced before, that is the energy-dependent weights are the very close to the optimal values in a least squares sense. Therefore, no visible difference can be observed in Figure 3.11 between these two methods either before or after the convex optimization.

On the other hand, the dependency of the quadratic algorithm to the initialization is obvious in this scheme. Indeed, four of the eight curves represented in Figure 3.11, which correspond to the points for selective-weights for each of the non-zero subbands in the central reconstruction, weights that are chosen to be either optimal in a least squares (LS) sense or in a fixed manner (using a high-rate approximation of the LS weights). Both the initial values for each of these two curves and the iterated values coincide.

This is important because it validates the selective WMR with fixed weights technique we have discussed above. Thus, there is no need for the additional costs incurred by transmitting a vector of reconstruction weights to the decoder along with the two descriptions. A fixed set of weights offers an acceptable sub-optimal solution.

Similar results are obtained for the other schemes that contain only full subbands at the last level of decomposition instead of only one of the quincunx polyphase components coming each from a different basis for the third detail subband[10].

[10]Recall that the only good invertible combinations of detail subbands at the coarsest resolution, lead at the

Quantization step	$w1(a_{(0,0)})$	$w2(dv_{(0,0)})$	$w3(a_{(1,0)})$	$w4(dh_{(1,0)})$	$w5(dd_{(1,0)})$
1	0.6047	0.7975	0.4113	1.0166	1.0038
5	0.6047	0.7976	0.4113	1.0160	1.0049
10	0.6047	0.7982	0.4113	1.0142	1.0029
20	0.6044	0.7980	0.4116	1.0120	0.9957
30	0.6044	0.7905	0.4116	1.0102	0.9901
50	0.6034	0.7750	0.4124	1.0041	0.9665
60	0.6034	0.7777	0.4124	1.0072	0.9610
70	0.6024	0.7783	0.4127	0.9819	0.9403

TABLE 3.9: The evolution of the computed least squares weights on each subband for the image "Mandrill" affected by 4% of random losses

Remark

Looking again at Figure 3.11 an important question arises: if the selective-weights for each subband give quasi-optimal results, then how can one take advantage of the convex optimization any more? The answer to that question resides in the interest in using MDC techniques instead of SDC ones. The on/off functioning for the transmission channel is not the most realistic scenario. Usually independent losses occur in the bitstream in addition to channel failure (such as congestion for instance).

Therefore in the second part of this case study we are presenting the random losses scenario affecting each description at coefficient level individually.

3.3.2 Random losses in each description

As we have seen in the previous section, a judicious choice of an initial reconstruction given the two descriptions can lead to optimal solutions that eliminate the need for the convex optimization algorithm. This was however a somewhat un-realistic scenario since in practice the channels may be disrupted by noise, thus provoking individual losses inside descriptions. We can model this kind of functioning by randomly setting to zero some of the received wavelet coefficients in each description.

In this situation a first observation is that the proposed fixed weights in Section 3.2.4.4, which are close to those computed in an optimal manner in the absence of losses at the encoder end, are no longer accurate. Indeed if we could calculate these weights at the decoder end, knowing the loss pattern that affected the signal, we would see that considerable differences occur.

For instance, in the case of our example image "Mandrill", one can observe the obtained Least Squares weights for each subband in the chosen invertible combination for the shifting scheme having as parameters $(s, s') \in \{(0,0),(1,0)\}$ in Table 3.9. These are given for one realization of losses that affect approximatively 4% of the wavelet coefficients.

The corresponding optimal weights computed at the encoder in the absence of losses can be observed in Table 3.10. According to this set of weights we had chosen a fixed set of five values which approximate those obtained for a quantization step equal to 1 in each description. Thus,

central decoder, to either 3 whole subbands distributed into the two descriptions in a quincunx manner or to 2 whole subbands plus 2 halves of subband - which gives a total of 3 half-subbands in each description.

Quantization step	$w1(a_{(0,0)})$	$w2(dv_{(0,0)})$	$w3(a_{(1,0)})$	$w4(dh_{(1,0)})$	$w5(dd_{(1,0)})$
1	1.0000	1.0001	0.0000	1.0000	0.9998
5	0.9999	1.0004	0.0001	0.9995	1.0007
10	0.9986	0.9992	0.0014	0.9981	0.9991
20	0.9924	0.9924	0.0077	0.9961	0.9911
30	0.9907	0.9873	0.0094	0.9947	0.9839
50	0.9483	0.9556	0.0513	0.9874	0.9626
60	0.9431	0.9537	0.0569	0.9883	0.9485
70	0.9264	0.9418	0.0726	0.9660	0.9357

TABLE 3.10: The evolution of the least squares weights computed on each wavelet subband for the image "Mandrill" (simulation at the encoder for different quantization steps)

for this particular shifting scheme the proposed set of fixed weights would be of $[1,\ 1,\ 0,\ 1,\ 1]$, respectively (with the same subband order as in Table 3.10). This is obviously an inadequate choice (at least for this particular noise realization). Note that even if several realizations of noise are performed for this loss percentage, in order to get a statistically accurate evaluation, the subband weights might vary a lot with different percentages of losses and therefore fixing a set of weights that would comply with the real ones obtained for known losses patterns is quickly becoming impractical if not impossible.

In Figure 3.12 we present a comparison between the different initialization points at the central decoder. Note that the simplest initialization from the two lossy descriptions yields quite good results, therefore a trade-off between complexity and quality might be in order. As expected, the fixed weights solution is very far from the optimal one at this stage. However, as it can be seen in Figure 3.13, where we present the results obtained after 30 iterations of quadratic optimization corresponding to the initialization curves we just mentioned, the iterative algorithm compensates the performances of this solution at least at small to medium bitrates. The quality is degrading at higher rates where the influence of the reconstruction weights is greater, since less drowned in the transmission noise.

The visual quality corresponding to the point of approximately 1.25 bpp in the two PSNR vs. rate graphs are given in Figure 3.14 for the initialization stage and in Figure 3.15 after 30 iterations of quadratic optimization. We also present the visual quality in the Single Description case, when a complete set of critically sampled wavelet coefficients is affected by the same 4% loss pattern as the first description.

Concluding remarks

To sum up this case study we are presenting a couple of conclusions regarding the decoding with multiple descriptions.

- If no individual losses occur on any of the transmission channels but instead some of them might be in impossibility of functioning (for instance in the case of bottlenecks) then a reconstruction from an average combination of received subbands is possible such that no iterative procedure is needed in order to have good performances.
- If individual losses might occur on some of the channels then the convex optimization algorithm presented in Section 3.2.3 leads to high and quick increases in the reconstruc-

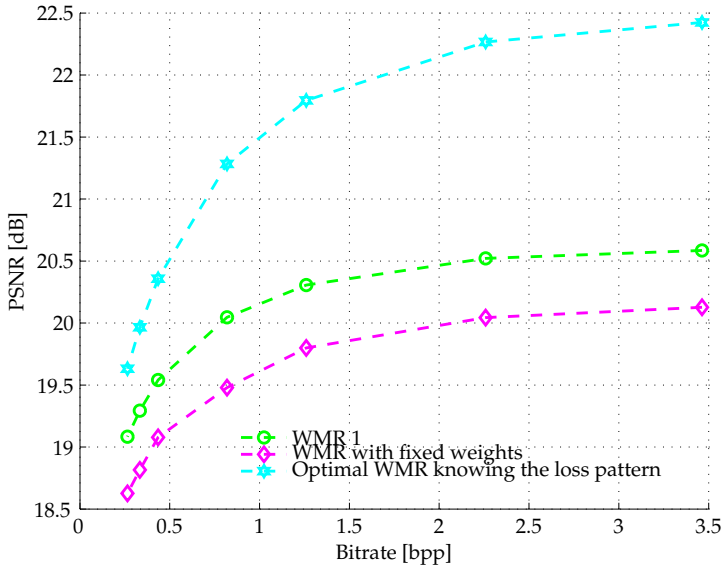


FIGURE 3.12: Comparison between the different initialization strategies for the image "Mandrill" affected by 4% of random losses (we have denoted by WMR 1 the simplest initialization which performs the weighted mean reconstruction with energy-given weights as in Section3.2.4.1).

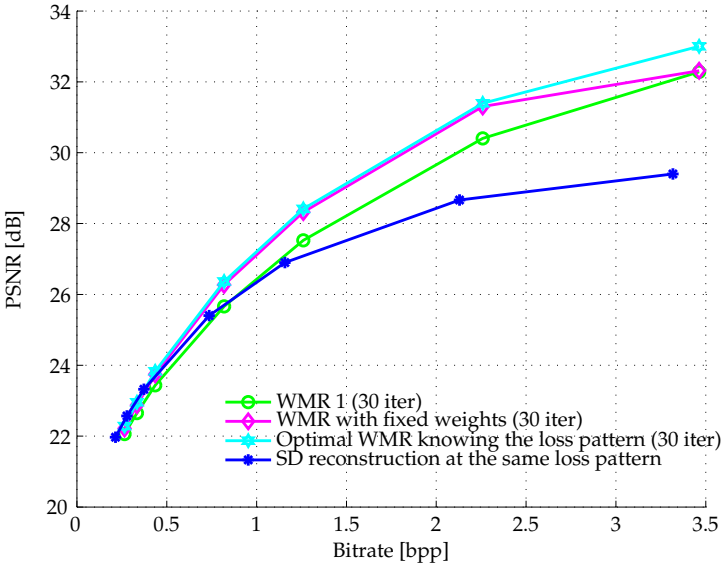
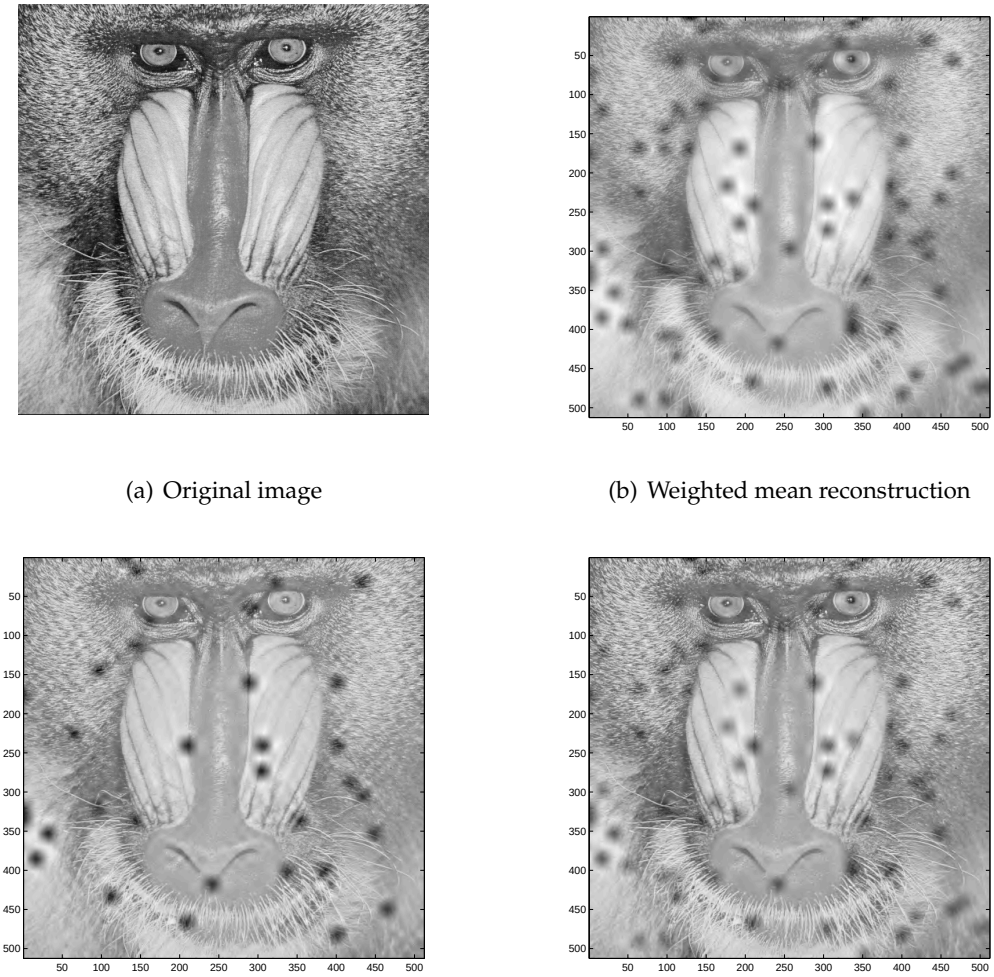


FIGURE 3.13: Comparison between the different iterative reconstructions for the image "Mandrill" affected by 4% of random losses starting from the initialization points given in Figure 3.12.

tion quality. Convergence speed curves are presented later on in the simulation results section. In this situation the initialization of the convex algorithm is important in order to reach the theoretical bounds that are expected of the proposed schemes (at very high bitrates), as well as in order to increase as much as possible the reconstruction at low and medium bitrates. We have distinguished among two strategies:



(a) Original image

(b) Weighted mean reconstruction

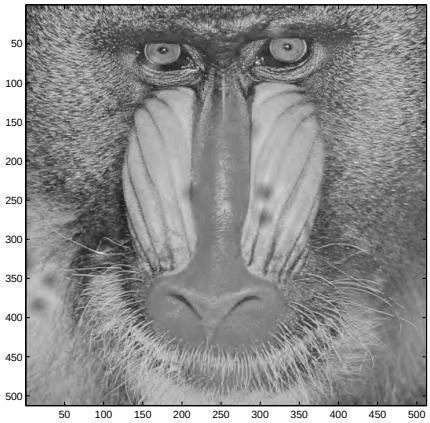
(c) Fixed weights

(d) Optimal LS weights for known losses

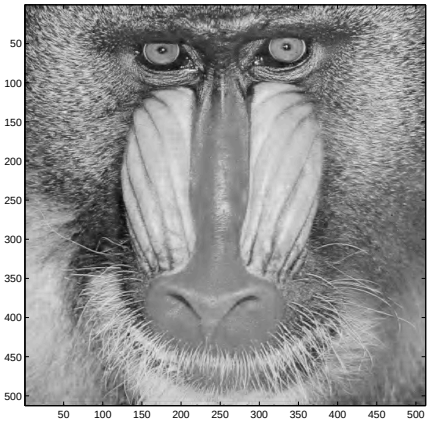
FIGURE 3.14: Different initializations for the 4% random losses scenario and reconstruction at 1.25 bpp.

- The simple average initialization - gives fairly good performances over the single description reconstruction with losses
- The fixed weights by subband initialization - slightly more complex but yielding better performances especially at medium bitrates (at least 0.5dB over the first initialization and some 1.5dB over the SD situation, at approximately 1.5bpp).

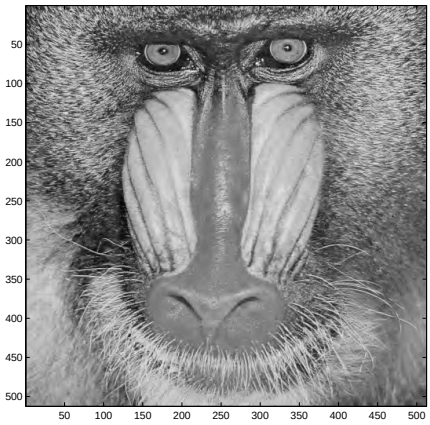
According to the transmission scenario as well as the desired quality of reconstruction, a trade-off must be considered in order to choose one of those techniques over the other.



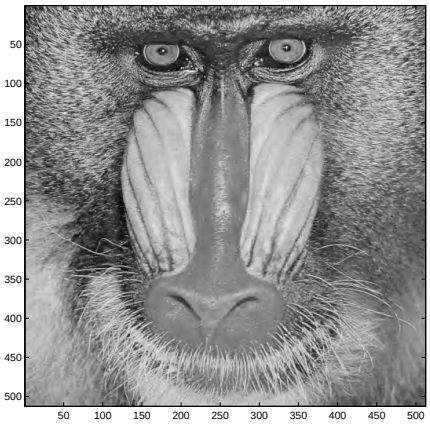
Single Description



Weighted mean reconstruction



Fixed weights



Optimal LS weights for known losses

FIGURE 3.15: The different reconstructions after 30 iterations of quadratic optimization for the 4% random losses scenario and reconstruction at 1.25 bpp.

3.4 Further extensions of the MDC scheme

The performances of our 2-description schemes might be enhanced by adding more diversity between the most energetic subbands, mainly the approximation ones. These remain highly correlated after the frame transform, therefore the central decoding cannot fully exploit this redundancy.

With this in mind we propose to perform a basic multiple description scalar quantization between the two approximation subbands, consisting in shifting the quantization indices by one half. The reception of the two descriptions leads to a finer quantized reconstruction of the approximation subbands.

By combining this with the chosen weighting strategy for the subbands, this should lead to good decoding results.

Another extension that can be brought to the initialization strategy is a finer choice of the fixed weighting coefficients of each received subbands according to the quantization step, in-

stead of approximating the values corresponding to the integer truncation quantization step, equal to 1.

3.5 Simulation results

In this section we are discussing in more detail the simulation framework that we have considered in order to validate the proposed MDC strategy.

3.5.1 Choice between the schemes

The first set of simulations aims at giving a choice criterion among the different invertible MDC schemes, since one can expect comparable performances at the central decoder (and the same quantization step) from several combinations of subbands as we have seen in Section 3.1.2.2.

Recall thus that we have four combinations of subbands at the coarsest resolution level in the case of the two schemes with shift parameters $(s, s') \in \{(0,0), (1,0), (0,1)\}$ (as presented in Tables 3.2, 3.4), that we denote by $S_{(1,0)}$ in the first case and by $S_{(0,1)}$ in the second case, that lead to better MSE performances than the non-redundant decomposition[11].

The only other possible shifting scheme in the impulse responses of the filters, denoted analogously by $S_{(1,1)}$ has two combinations that provide the best central reconstruction, as we have seen before in Table 3.6. Moreover, as we have previously seen, these combinations are merely the whole non-redundant basis plus the approximation subband from the other non-redundant representation.

While all these ten possible schemes have comparable central performances, as we can see from Figure 3.16, in the case of the following combination $S_{(1,0)}$: $\{a_{(0,0)}, a_{(1,0)}, dv_{(0,0)}, dh_{(1,0)}, dd_{(1,0)}\}$ called "Scheme 1" in Figure 3.16 and the combination containing the whole coefficient set, $C_{(0,0)}$ plus the approximation subband $a_{(1,1)}$ which is called "Scheme 2". The image considered here is the standard "Lena" image of 512×512 pixels.

In this situation, no indication for the choice of one scheme over the other can be given, therefore we propose to look at the side decoding reconstruction, in the basic functioning mode of MDC (on/off channels) without additional random losses. The PSNR-rate performances for the two schemes mentioned above are given in Figure 3.17.

We can see from these curves that the two schemes exhibit very different performances in terms of side decoding. Thus, the Scheme $S_{(1,0)}$ (Scheme 1) performs as a balanced one, whereas the Scheme $S_{(1,1)}$ (Scheme 2) has very different side decoders. According to the transmission scenario to which one is confronted, one of these two schemes can be chosen. More precisely, if unequal error protection is available a good choice would be given by the Scheme $S_{(1,1)}$ since we could favor the transmission of the first description over the second. Conversely, for equally error-subjected networks a better choice is $S_{(1,0)}$. In these PSNR-rate curves we are illustrating the inherent performances of the quadratic optimization algorithm even in the situation of a poor initialization.

These preliminary PSNR-rate results have been obtained with an $EZBC$ lossless encoder used on uniformly quantized coefficients. In the next set of simulations we are considering

[11] We are omitting the combination $(s, s') = (0,0)$ in our scheme notation, since it is common to all the proposed schemes, and corresponds to the first wavelet decomposition in these redundant representations.

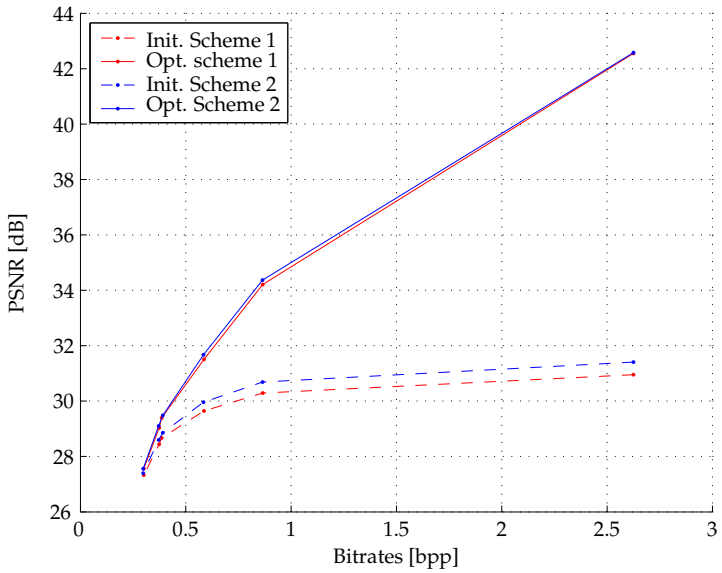


FIGURE 3.16: PSNR-rate evaluation of two MDC schemes. Dashed lines for the curves obtained before the optimization algorithm (initialization with the WMR1 method), and full lines for the curves after the quadratic optimization algorithm.

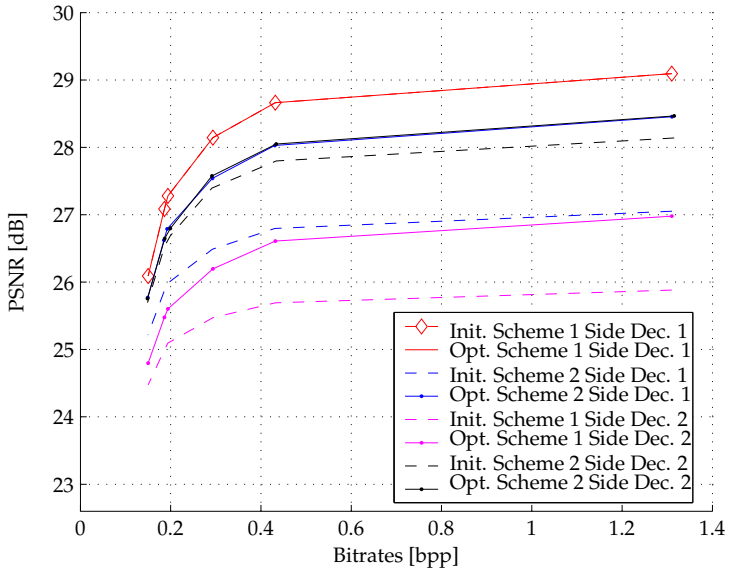


FIGURE 3.17: PSNR-rate evaluation of two schemes for the side decoders (dashed lines correspond to values before convex optimization (WMR1 initialization) and full lines to the values obtained after 30 iterations of the optimization algorithm).

only Scheme 1 because it works in a more general transmission setup.

Fig. 3.18 gives an idea on the convergence speed for the iterative algorithm for the two central decoders and the four side decoders corresponding to the two considered schemes. In the bottom graph the fact that the second scheme is balanced can be easily noticed. Also, one should observe that since in Scheme 1 the first side decoder contains only coefficients from one of the basis, no optimization needs to be performed. The characteristics of the convergence

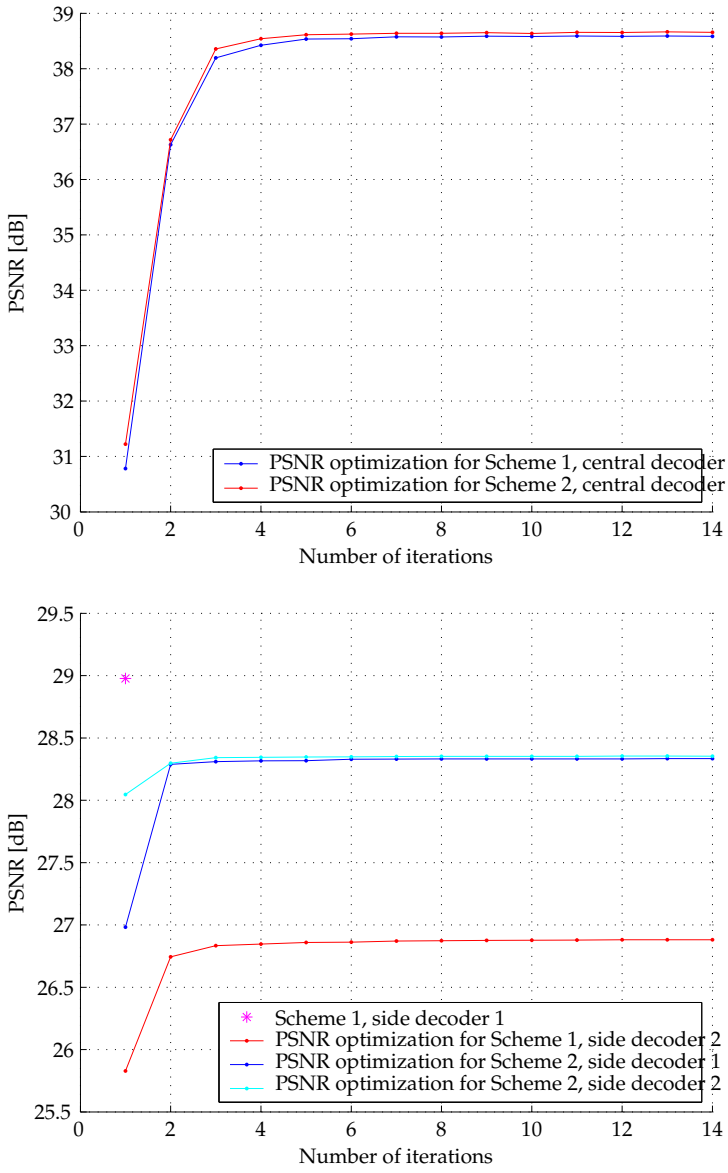


FIGURE 3.18: Convergence speed for the proposed optimization algorithm: central (top graph) at 1.7 bpp and side (bottom graph) at 0.8 bpp decoders.

curves are similar for all bitrates, the quickest convergence being obtained at low bitrates (almost 90% of gain in 2 iterations) and the slowest at high bitrates (Fig.3.18).

3.5.2 Scheme performance versus a critically sampled decomposition

In order to verify the theoretical performances in terms of the MSE of the reconstruction for a quantization step: $\delta = 1$ as presented in Section 3.1.2.2, we are going to compare the obtained PSNR of the MDC scheme $S_{(1,0)}$ mentioned earlier with that of a critically sampled decomposition uniformly quantized with the same step of 1.

We recall here the bounds for the MSE in these two cases. Thus we have obtained the

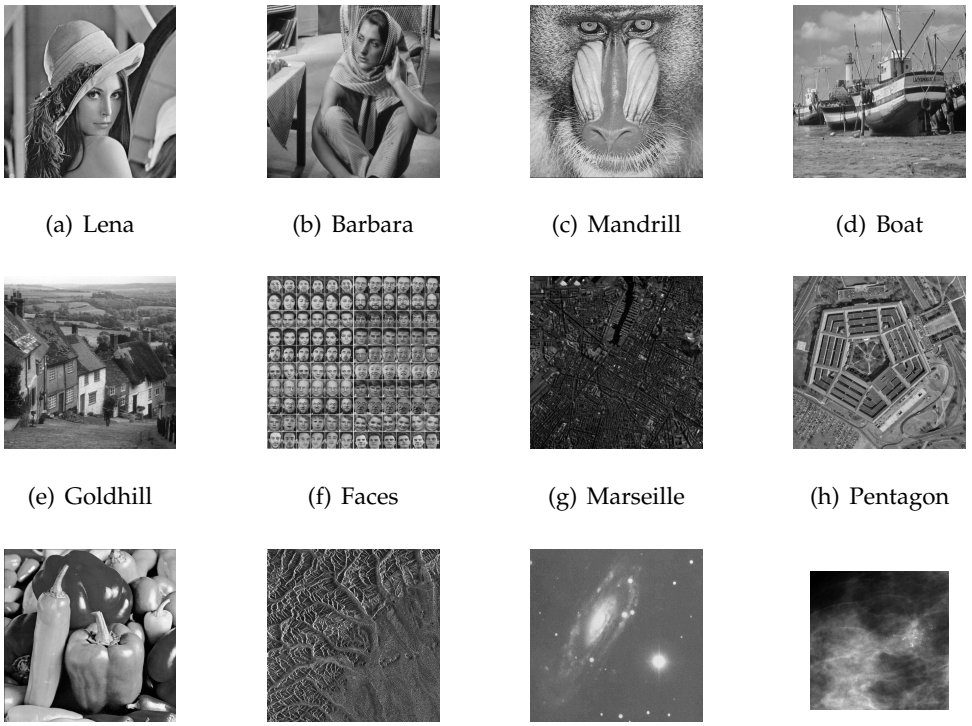
MSE-s: $e_{\mathcal{M}^{\#}_{(0,0)}} = 8.1882$ at $\delta = 1$ for the critically sampled decomposition corresponding to Eq. (3.1) and $e_{\bar{\mathcal{M}}^{\#}} = 6.9801$ for the $S_{(1,0)}$ scheme given by the union of the following subbands $\{a_{(0,0)},\ a_{(1,0)},\ dh_{(0,0)},\ dd_{(0,0)},\ dv_{(1,0)}\}$ at the last level of decomposition and the same detail subbands as the critically sampled decomposition at higher resolutions.

Thus a bound for the performance gain to be expected between the MDC scheme and the SDC one is given by:

$$G_{\text{MD vs SD}} = 10 \log_{10} \frac{e_{\mathcal{M}^{\#}_{(0,0)}}}{e_{\bar{\mathcal{M}}^{\#}}} \approx 0.69$$

Note however that the evaluations of the reconstruction error are given for the quincunx transfer matrices corresponding to the coarsest resolution level, since this is where the diversity with respect to a critically sampled decomposition is introduced with our MDC schemes. Therefore we are evaluating the performances at the central decoder and in the absence of additional losses, except those introduced by quantization, on one level of wavelet decomposition. Thus the $S_{(1,0)}$ scheme is formed only with the following subbands: $\{a_{(0,0)},\ a_{(1,0)},\ dh_{(0,0)},\ dd_{(0,0)},\ dv_{(1,0)}\}$ whereas the SD scheme is given by: $\{a_{(0,0)},\ dh_{(0,0)},\ dd_{(0,0)},\ dv_{(0,0)}\}$.

Moreover since we do not compute the pseudoinverse of the quincunx transfer matrix in our scheme, we are going to compare the performances of the reference image obtained with the selective WMR with least squares weights for each individual wavelet subband in the MD scheme (as in Section 3.2.4.3) with the ones for the direct inversion of the SD scheme. We are presenting these results for several test images, given in Figure 3.19 in Table 3.11.



(a) Lena (b) Barbara (c) Mandrill (d) Boat

(e) Goldhill (f) Faces (g) Marseille (h) Pentagon

(i) Peppers (j) SAR (k) Astro (l) Medical

FIGURE 3.19: Test images for MDC vs SDC comparison.

Image	PSNR [dB]			Gain	
	SD	MDC	dual frame	G_{MDC}	G_{dual}
Lena	58.8283	59.2483	59.5265	0.42	0.6982
Barbara	58.8188	59.2220	59.5173	0.40	0.6985
Mandrill	58.8130	59.2000	59.5017	0.39	0.6888
Goldhill	58.8177	59.2270	59.5158	0.40	0.6981
Boat	58.8237	59.2140	59.5010	0.39	0.6773
Faces	58.8145	59.2910	59.4929	0.48	0.6784
Marseille	58.8198	59.2319	59.5076	0.41	0.6878
Pentagon	58.8168	59.2029	59.4989	0.39	0.6820
Peppers	58.8068	59.2174	59.5089	0.41	0.7021
SAR	58.8106	59.1979	59.5030	0.38	0.6881
Astro	58.8212	59.2039	59.4983	0.38	0.6771
Medical	58.8249	59.2196	59.4887	0.39	0.6638

TABLE 3.11: Comparison between the MDC scheme $S_{(1,0)}$ and the SDC scheme at quantization step of 1, 1 level of wavelet decomposition and 10 iterations of convex optimization (for the MDC scheme). PSNR MDC gives the value after the convex optimization algorithm starting with the reference image given by the selective WMR with least squares weights for each individual subbands. The gains are computed as the difference between PSNR MDC and PSNR SDC. The theoretical gain is obtained by computing the reconstruction from the dual frame operator (last column, denoted here by G_{dual})

From the results obtained in Table 3.11 we can see that the convex optimization algorithm for the decomposition on one level is not too far from the theoretical expected gain. For comparisons purposes we have equally implemented the reconstruction by the dual frame operator and we are also giving the scheme gain in this situation in Table 3.11.

A second remark for the results obtained in Table 3.11 concerns the types of the test images. We can see that the values that are the closest of the theoretical gain of the MDC scheme are natural images, on which is safer to assume that the quantization noise for the wavelet coefficients is uniform and i.i.d.

3.5.3 Random losses scenario

We are now considering the more general transmission scenario in which random losses can occur into each of the transmitted descriptions.

In the following we are presenting the simulation results obtained for a test image that does not belong to the original image database that we have considered in order to compute the LS reconstruction weights for the initialization of the convex optimization algorithm using the combined reconstruction from individual subbands. This test image is given in Figure 3.20.

The wavelet coefficients are obtained with biorthogonal $9/7$ wavelets and the encoding is done using JPEG2000. We are obtaining the bitrates by imposing a quantization step to the JPEG2000 encoder and we are using the dead-zone quantization option.

We are going to test the robustness to losses for the Scheme $S_{(1,0)}$ for different percentages of individual losses in each description.

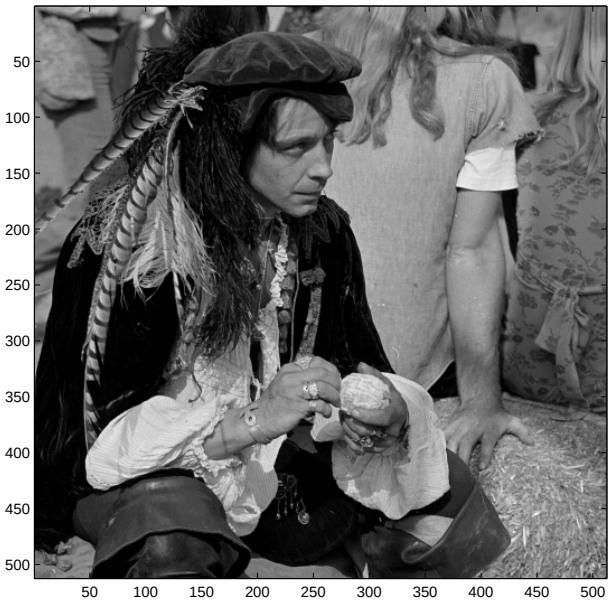


FIGURE 3.20: Test image: "Man.pgm".

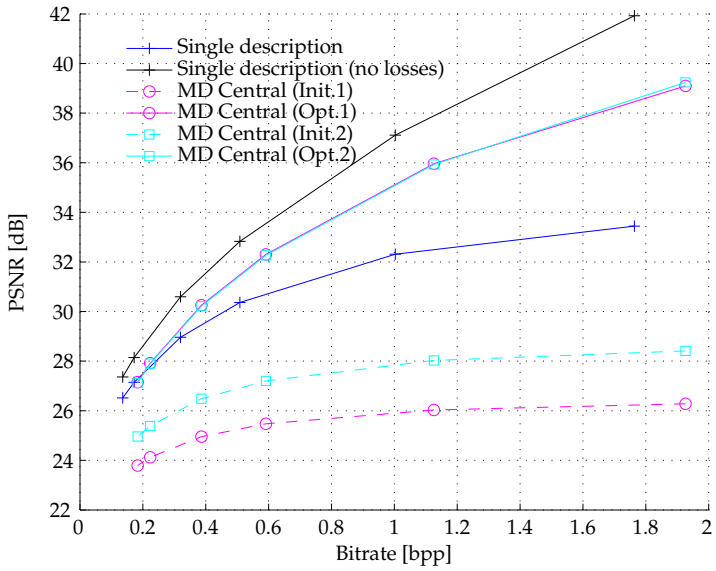


FIGURE 3.21: PSNR vs. bitrate. Comparison between the central decoder of Scheme $S_{(1,0)}$ at 1% random losses in each description and the single description case affected by the same loss pattern as the first description.

In Figure 3.21 we can see a comparison with the single description scheme (given by the first non-redundant decomposition used to form our MDC schemes, a classical biorthogonal $9/7$ decomposition) of our proposed MDC scheme. We are considering the fixed-weights initialization for the optimization algorithm, with the fixed-weights computed on the considered image database in the no-loss situation. In addition we compute the LS weights for each sub-

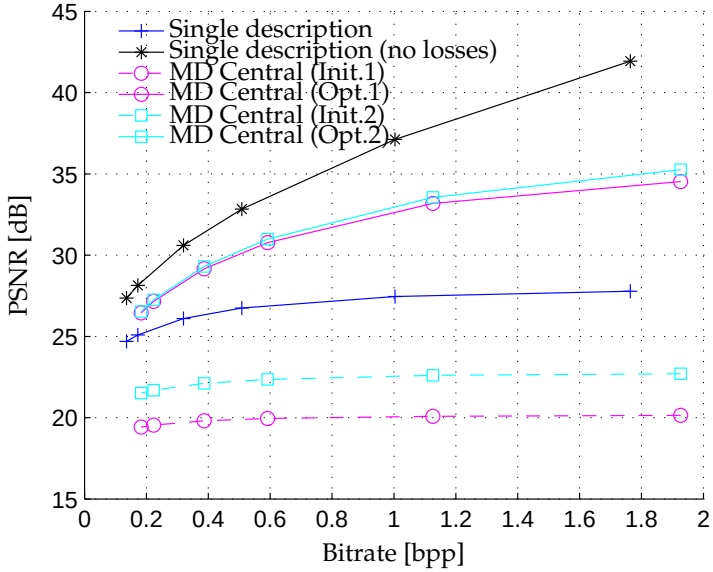


FIGURE 3.22: PSNR vs. bitrate. Comparison between the central decoder of Scheme $S_{(1,0)}$ at 4% random losses in each description and the single description case affected by the same loss pattern as the first description.

Central Bitrate [bpp]	1.9265	1.1260	0.5917	0.3866	0.2233	0.1834
PSNR Central Init. 1 [dB]	20.1451	20.0878	19.9473	19.8026	19.5424	19.4265
PSNR Central Opt. 1 [dB]	34.5244	33.1773	30.7742	29.1586	27.1567	26.4473
PSNR Central Init. 2 [dB]	22.7032	22.6016	22.3583	22.1170	21.6958	21.5137
PSNR Central Opt. 2 [dB]	35.2535	33.5549	30.9885	29.3071	27.2528	26.5337
SD Bitrate [bpp]	1.7640	1.0035	0.5087	0.3201	0.1723	0.1360
PSNR SD with losses [dB]	27.7820	27.4607	26.7470	26.0965	25.0876	24.6953
PSNR SD without losses [dB]	41.9294	37.1143	32.8276	30.5977	28.1441	27.3631

TABLE 3.12: PSNR vs Bitrate comparison between the MDC scheme $S_{(1,0)}$ and the SD scheme affected by 4% of random losses (the values are averaged over 100 runs of random losses in each description)

band knowing the loss pattern, in order to have an idea of the influence of the approximation by fixed-weights on the overall reconstruction performances. All the following tests are averaged over 100 runs of random losses with different percentages in each description. We also give the performances of the single description scheme when no losses had occurred.

Figure 3.22 presents the same results as before for 4% of losses in each description. We can observe that the MDC scheme greatly outperforms this baseline reconstruction of the single description strategy. Naturally, error concealment could be in addition applied to the single description scheme in order to enhance its performances.

In Figure 3.23 we can observe the performances of the side decoders before the optimization algorithm and after 30 iterations of convex optimization. One can observe that a gain of about 1dB can be obtained for the side decoders, too.

The exact numerical values corresponding to the curves in Figure 3.22 are given in Table 3.12

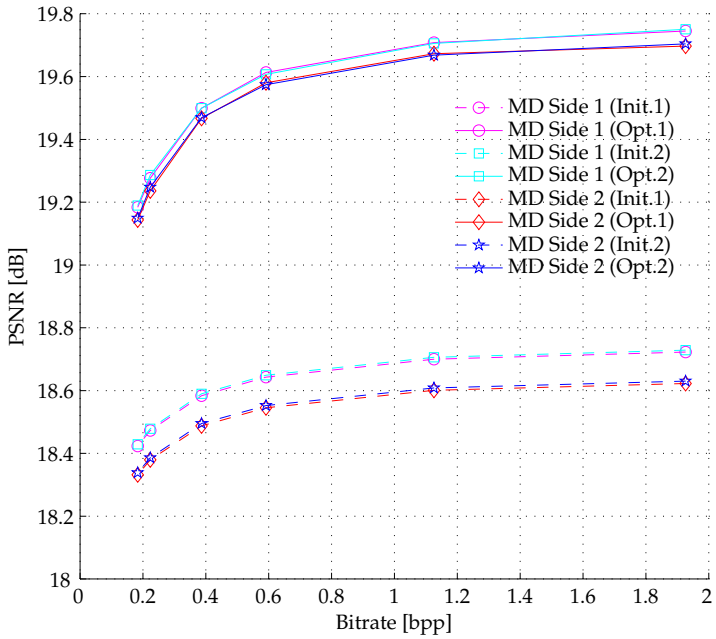


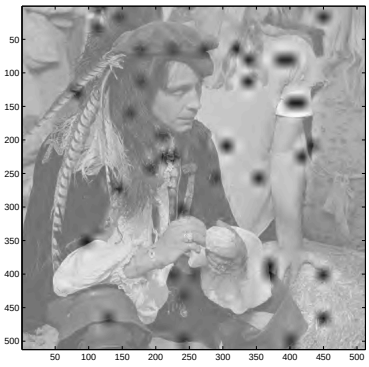
FIGURE 3.23: PSNR vs. bitrate performances for the side decoders at 4% of random losses in each description (values averaged over 100 runs of random losses).

and the reconstructed images for the third point of the curves are presented in Figure 3.24.

Increasing the loss percentage in each description to 6% leads to the curves presented in Figure 3.25. As expected, the greater the loss percentage, the greater the gain between the MDC and SDC can be expected.

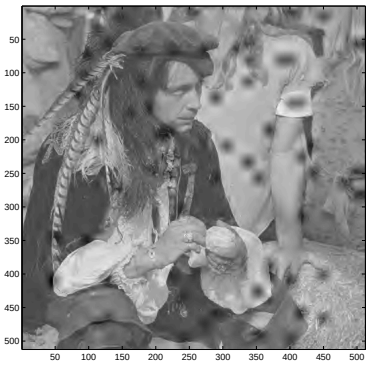
(a) SD reconstruction (no losses), 32.8276 dB

(b) SD reconstruction with 4% of losses, 26.7470dB



(c) MDC fixed weights (before), 19.9473 dB

(d) MDC fixed weights (after), 30.7742 dB



(e) MDC LS weights (before), 22.3583 dB

(f) MDC LS weights (after), 30.9885 dB

FIGURE 3.24: The different reconstructions after 30 iterations of quadratic optimization for the 4% random losses scenario and reconstruction at 0.39 bpp for the multiple description scheme and at 0.32 for the single description scheme (the same quantization step in both cases).

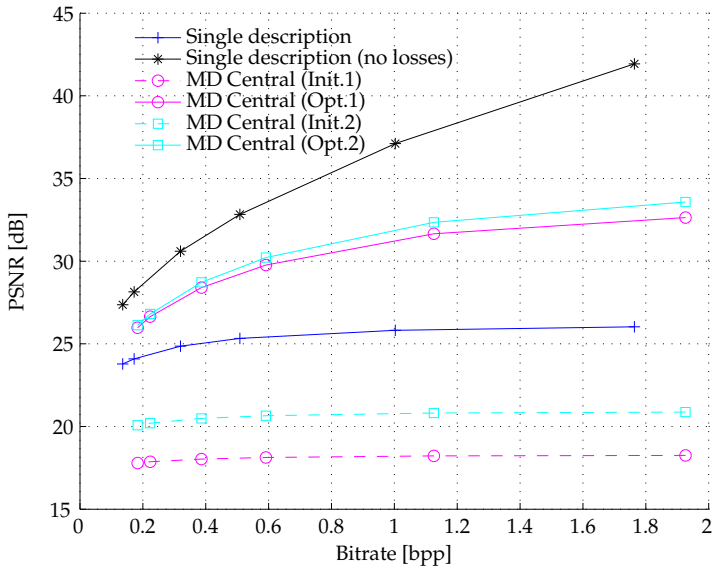


FIGURE 3.25: PSNR vs. bitrate. Comparison between the central decoder of Scheme $S_{(1,0)}$ at 6% random losses in each description and the single description case affected by the loss pattern as the first description (values averaged over 100 runs).

3.6 Conclusions

In this chapter we have considered low redundancy MDC schemes for images, based on redundant wavelet frame expansions. We have shown that by an appropriate design of this decomposition the perfect reconstruction property is satisfied.

The decoding performances of the multiple descriptions schemes are improved by an iterative quadratic optimization algorithm. We have discussed the importance of choosing some schemes over the others among the possible combinations and we have equally discussed methods of improving the final reconstruction by improving the initialization point of the optimization algorithm.

We have considered a lossy scenario and tested the performances of our schemes over the single description case, affected by the same losses as one of the two descriptions. We have shown experimentally that by initializing the iterative optimization algorithm with the combination of the received individually recovered subbands from each description we can obtain satisfactory performances. This combination, which is in fact a weighted average, utilizes a fixed set of weights that is chosen as an approximation of the ones which are optimal in a least squares sense when no losses have occurred on the network.

We have computed these optimal weights for several schemes that are giving the best MSE errors in each MDC scheme that can be conceived for the proposed frame decomposition, on a database of diverse images (natural, synthetic, satellite, biomedical etc.). It has thus been shown experimentally, that for each shifting scheme and subband combination in order to form two descriptions, there is a unique set of optimal weights which presents negligible variations of values among the images.

The problem with computing the optimal reconstruction set of weights for each MDC scheme is that it would incur the transmission of this set along with the descriptions, and, therefore, the MDC technique would greatly loose in flexibility. We have then studied comparatively the performances obtained if we approximate this optimal weights by a fixed set and found out that the results can be considered as satisfactory.

In the lossy scenario we have also computed the optimal reconstruction weights when the pattern of losses is known in each description, in order to compare the performances obtained with the fixed weights with the maximal achievable ones. This could also be useful if the functioning of the transmission network can be considered deterministic.

Several extensions of the MDC strategies are possible. First, the coding efficiency might be improved if the two approximation subbands in each description would be passed by a decorrelating transform for instance at the central decoder. Secondly, the optimal weights involved in the initialization of the optimization algorithm are varying across the bitrates, or in other words with the quantization step. A further optimization in order to find the best set of subband weights with the quantization step might also be considered.

Chapter 4

A complementary approach exploiting sparsity

In this chapter we present a complementary viewpoint inspired from the recent Compressed Sensing (CS) theory which gives remarkable results for analyzing signals having a sparse representation in some frame. These results basically state that such a signal can be perfectly recovered from a reduced number of arbitrary projections, provided that it is compressible. Being compressible means that it exists a sparse representation of this signal in some basis, and in this case the "sparsity" of the signal is given by the number of non-zero samples. This signal recovery problem can also be found under the name of sparse approximation and it has mainly been addressed by Matching Pursuit [DDWB06] or Basis Pursuit techniques [Don06]. Matching Pursuit has also been used in a first stage of generating multiple balanced descriptions (see [RF06] and references therein) for still images, in a lossy network scenario.

In our MDC framework we borrow from the CS field the idea that the encoder should determine a reduced number of components of the image in a frame representation from the observation of its pixel values. The choice of the components is grounded on a rate-distortion formulation of the MDC problem, which after some simplifications is re-expressed as a convex optimization problem.

Some of the results in this chapter are the outcome of a joint work with A. Fraysse and they were published in [PPPP07] and [PFPPP07].

4.1 Analysis vs. synthesis frames

Let us first reformulate the frame-based multiple description coding problem in a more general framework than in the previous chapters. We assume that the signal to be encoded belongs to a real Hilbert space $\mathcal{H}$ endowed with an inner product $\langle .,. \rangle$ and the associated norm $\|.\|$ and we do not restrict the MDC application to the two-description case.

Thus, in a general Multiple Description scenario with D descriptions, each description $i \in \{1, \ldots, D\}$ is obtained from a vector family $(e_{i,k})_{k \in \mathbb{K}_i}$ in $\mathcal{H}$ with $\mathbb{K}_i \subseteq \mathbb{N}$. The union of these families is assumed to form a redundant frame of $\mathcal{H}$.

The associated decomposition operators of this frame are considered as follows. For all

$i \in \{1, \dots, D\}$,

$$\begin{aligned} L_i \ : \ \mathcal{H} &\to \ell^2(\mathbb{K}_i) \\ x &\mapsto (\langle x, e_{i,k} \rangle)_{k \in \mathbb{K}_i}. \end{aligned} \tag{4.1}$$

And their adjoint operators are given by:

$$\begin{aligned} L_i^* \ : \ \ell^2(\mathbb{K}_i) &\to \mathcal{H} \\ (\xi_k)_{k \in \mathbb{K}_i} &\mapsto \sum_{k \in \mathbb{K}_i} \xi_k e_{i,k}. \end{aligned} \tag{4.2}$$

Based on these frame operators, two different points of view can be adopted in the design of an MDC scheme, and we are distinguishing them by the names of "analysis" and "synthesis" frame paradigm. The former corresponds to the conventional philosophy of an MDC scheme, that has been approached in the literature so far, whereas the latter expresses a complementary approach in some sense, that will be more apparent a little further.

In the *analysis* frame paradigm, a signal $x \in \mathcal{H}$ is decomposed by the linear operators L_1 and L_2, as given before, so as to provide D descriptions

$$c_i = L_i x, \qquad i \in \{1, \dots, D\}, \tag{4.3}$$

which are quantized and transmitted separately. At the decoder side, only a subset of descriptions may be available and the problem that needs to be solved is to reconstruct a signal $\hat{x}$ as close as possible to x. In order to improve the quality of the reconstruction, the convex constraints induced by the quantization rules can be addressed within a convex optimization approach as we have seen in Chapter 3. Solving this problem leads to a nonlinear, hence computationally complex, reconstruction.

A block-diagram representation for these considerations in the two-description scenario is given in Figure 4.1. Here, we have highlighted the fact that, once the strategy for creating the descriptions is chosen, the remaining difficulty of the scheme lies at the decoder end and, consequently, the improvement in terms of quality of reconstruction will be solely steered by the chosen decoding method. Thus, we point out the fact that such an application is mostly suitable for networks providing scalable[1] decoders, or more generally for the situation in which computationally complex decoders are available and only simple encoders are used.

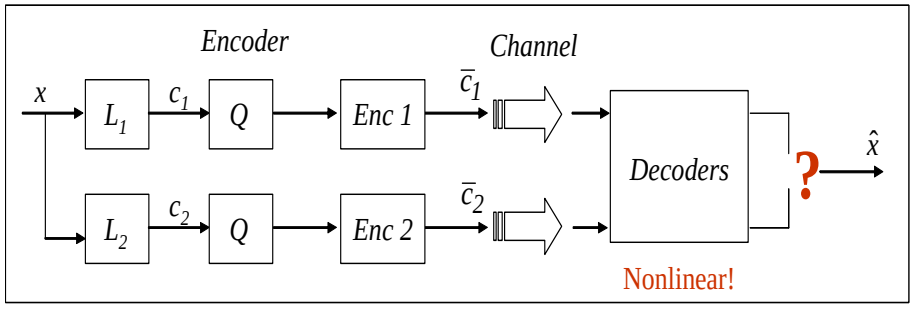


FIGURE 4.1: The analysis frame paradigm.

[1]Here we employ the term of *scalable* in the sense of adaptability of the network with respect to the user demands or the receivers characteristics

A second point of view concentrates more on the efficient design of the MD encoders and on imposing very simple decoders (for instance linear ones). This framework is more adequate for low-delay applications such as broadcasting. While this strategy seems similar to the rate-distortion trade-off sought by standard MDC methods, the two differ by the decomposition operators that are used. Hence, in the so-called *synthesis* frame paradigm, the adjoint operators L_1^* and L_2^* are used at the decoder side and they serve in reconstructing the signal from its sequences of quantized values $\bar{c}_i$ with $i \in \{1, \dots, D\}$ that correspond to the transmitted descriptions, as follows:

$$\hat{x} = \sum_{i \in \{1,\dots,D\}} L_i^* \bar{c}_i \tag{4.4}$$

for the central decoder, and

$$\hat{x} = \widetilde{L}_i^* \bar{c}_i, \qquad i \in \mathbb{I}. \tag{4.5}$$

where $\widetilde{L}_i^* : \ell^2(\mathbb{K}_i) \to \mathcal{H}$ is a given reconstruction operator and $\mathbb{I} \subset \{1, \dots, D\}$ is the set of received descriptions, for the side decoders.

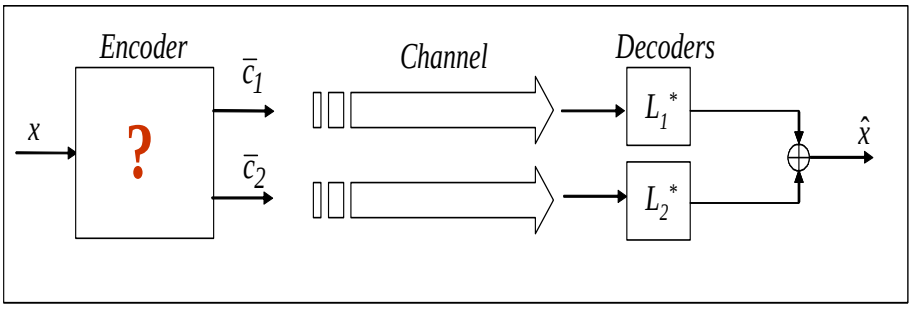


FIGURE 4.2: The synthesis-frame paradigm.

We see that one of the advantages of this approach is that the decoders take a simple linear form. At the encoder, the problem is however to generate the sequences $\bar{c}_1$ and $\bar{c}_2$ in the best way in a rate-distortion sense, taking into account the channel characteristics. This situation is represented by the block diagram in Figure 4.2.

The design of the encoding rule yields a nonlinear optimization problem which is formulated in the next section.

4.2 Rate-distortion problem for the synthesis frame approach

With little loss of generality for practical purposes, we will subsequently assume that a finite number of frame coefficients is considered, that is $\mathbb{K}_i = \{1, \dots, K_i\}$ (which implies that $\mathcal{H}$ is finite dimensional).

Let $R(\bar{c}_i)$, with $i \in \{1, \dots, D\}$, denote the number of bits required to transmit the sequence of quantized values $\bar{c}_i$. We aim at minimizing the global bitrate

$$R_{\text{global}} = \sum_{i=1}^{D} R(\bar{c}_i) \tag{4.6}$$

under a distortion constraint which is expressed as

$$D = \sum_{\mathbb{I}\in\mathcal{P}} \alpha_{\mathbb{I}} \|x - \sum_{i\in\mathbb{I}} \tilde{L}^*_{\mathbb{I},i}\bar{c}_i\|^2 \leq D_{\max}. \tag{4.7}$$

where $\mathcal{P}$ is the set of non-void parts of $\{1,\ldots,D\}$. Hereabove, $D_{\max}$ is the maximum distortion allowed whereas $\alpha_{\mathbb{I}}$ are, for instance, the probabilities of receiving the descriptions in the different scenarios, but other criteria can equally be considered (perceptual quality, etc).

Finding the sequences $\bar{c}_i$ which minimize (4.6) subject to the constraint (4.7) is a difficult global non-convex optimization problem. However, under some hypotheses on the coefficients and on the quantization noise some simplifications can be brought to the problem, thus casting it as a convex optimization one, which can be dealt with using recent algorithms, [CCPW06]. Note also that the upper bound, $D_{\max}$, should be chosen large enough to guarantee the existence of a solution to the optimization problem.

More precisely, the coefficients $(\bar{c}_{i,k})_{1\leq k\leq K_i}$ with $i \in \{1,\ldots,D\}$ can be viewed as the outputs of a uniform quantizer of step $q > 0$ driven with real-valued coefficients $c_{i,k}$. In the same time, they can also be viewed as a realization of a random vector $\overline{C}_i = (\overline{C}_{i,k})_{1\leq k\leq K_i}$ taking its values in $\{\ldots -2q, -q, 0, q, 2q, \ldots\}^{K_i}$.

Since for memoryless sources, the entropy provides a lower-bound for the rate, we are going to minimize the global entropy:

$$H_{\text{global}} = \sum_{i=1}^{D}\sum_{k=1}^{K_i} H(\overline{C}_{i,k}), \tag{4.8}$$

where $H(\overline{C}_{i,k})$ is the discrete entropy of $\overline{C}_{i,k}$, defined by:

$$H(\overline{C}_{i,k}) = -\sum_{n\in\mathbb{Z}} P(\overline{C}_{i,k} = nq)\log_2\big(P(\overline{C}_{i,k} = nq)\big). \tag{4.9}$$

The initial (unquantized) coefficients $c_{i,k}$ can be equally considered as being realizations of independent, real, random variables $C_{i,k}$ whose probability densities, $p_{i,k}$, can be modelled by generalized Gaussian laws. Their probability density is then given by:

$$\forall \xi \in \mathbb{R}, \qquad p_{i,k}(\xi) = \frac{\beta_{i,k}\omega_{i,k}^{1/\beta_{i,k}}}{2\Gamma(1/\beta_{i,k})} e^{-\omega_{i,k}|\xi|^{\beta_{i,k}}} \tag{4.10}$$

where $\beta_{i,k} \geq 1$ et $\omega_{i,k} > 0$.

Let us set $N_i \in \mathbb{N}^*$ for every $i \in \{1,\ldots,D\}$. We then define the following set:

$$\mathcal{S}_i := \{(\beta_i^{(l)}, \omega_i^{(l)}),\ 1 \leq l \leq N_i\}. \tag{4.11}$$

And we suppose that:

$$\forall K_i \in \mathbb{N}^*\ \ \forall k \in \{1,\ldots,K_i\} \quad (\beta_{i,k}, \omega_{i,k}) \in \mathcal{S}_i. \tag{4.12}$$

In other words, we assume that, for all $i \in \{1,\ldots,D\}$ and all K_i, the parameters corresponding to the random variable $C_{i,k}$ law must take a precise finite number of values. In the case of wavelet coefficients, this hypothesis can be translated by the fact that the coefficients are identically distributed in each subband.

Moreover, according to [GP68], at high-resolution (i.e. small quantization steps), the entropy of $\overline{c}_{i,k}$ and the differential entropy of $C_{i,k}$ are related by:

$$H(\overline{C}_{i,k}) \approx h(C_{i,k}) - \log_2(q) \tag{4.13}$$

where

$$\begin{aligned} h(C_{i,k}) &= -\int p_{i,k}(x)\log_2(p_{i,k}(x))dx \\ &= -\mathbb{E}(\log_2(p_{i,k}(C_{i,k}))). \end{aligned} \tag{4.14}$$

For every $i \in \{1,\dots,D\}$ and every $l \in \{1,\dots,N_i\}$ we build the set:

$$\mathcal{S}_i^{(l)} = \{k \in \{1,\dots,K_i\} \ \mid \ (\beta_{i,k},\omega_{i,k}) = (\beta_i^{(l)},\omega_i^{(l)})\}, \tag{4.15}$$

where $(\beta_i^{(l)},\omega_i^{(l)}) \in \mathcal{S}_i$. In this case, for each $i \in \{1,\dots,D\}$ and $l \in \{1,\dots,N_i\}$, the sequences $(C_{i,k})_{k\in\mathcal{S}_i^{(l)}}$ are independent and identically distributed. In addition to this, we assume that for every $i \in \{1,\dots,D\}$ and every $l \in \{1,\dots,N_i\}$, the cardinality of $\mathcal{S}_i^{(l)}$ is proportional to K_i. And since the random variables $(\log_2(p_{i,k}(C_{i,k})))_{1\le i\le D;\ k\in\mathcal{S}_i^{(l)}}$ have finite variance, we can invoke the strong law of large numbers, allowing to deduce that, when $K_i \to \infty$,

$$-\frac{1}{\mathrm{Card}(\mathcal{S}_i^{(l)})}\sum_{k\in\mathcal{S}_i^{(l)}}\log_2(p_{i,k}(C_{i,k})) \xrightarrow{\text{a.s.}} h(C_i^{(l)}), \tag{4.16}$$

where $C_i^{(l)}$ is a random variable with probability law given by a generalized Gaussian function of parameters $(\beta_i^{(l)},\omega_i^{(l)})$.

In this manner the differential entropy is approximated by an empirical mean.

By using (4.16) in (4.8), the minimization of H_{global} is cast as the minimization of

$$J(c_1,\dots,c_D) = \sum_{i=1}^{D}\sum_{k=1}^{K_i}\omega_{i,k}|c_{i,k}|^{\beta_{i,k}}, \tag{4.17}$$

Note that the parameters $\omega_{i,k}$ and $\beta_{i,k}$ can be estimated by Maximum Likelihood.

Let us now investigate the approximation of the distortion by a convex function of c_i. We define, for each $i \in \{1,\dots,D\}$, the quantization error vector $\varepsilon_i = (\varepsilon_{i,k})_{1\le k\le K_i}$ which is reflected to the quantized coefficients in an additive manner: $\overline{c}_i = c_i + \varepsilon_i$.

The global distortion can thus be rewritten as follows:

$$D = \sum_{\mathbb{I}\in\mathcal{P}}\alpha_{\mathbb{I}}\|x - \sum_{i\in\mathbb{I}}\tilde{L}^*_{\mathbb{I},i}(c_i+\varepsilon_i)\|^2.$$

By using the Hilbert structure of $\mathcal{H}$ and the linearity of the operators $\tilde{L}^*_{\mathbb{I},i}$, we can express each term of the preceding summation under the form:

$$\begin{aligned} \|x - \sum_{i\in\mathbb{I}}\tilde{L}^*_{\mathbb{I},i}(c_i+\varepsilon_i)\|^2 &= \|x - \sum_{i\in\mathbb{I}}\tilde{L}^*_{\mathbb{I},i}c_i\|^2 + \|\sum_{i\in\mathbb{I}}\tilde{L}^*_{\mathbb{I},i}\varepsilon_i\|^2 \\ &\quad + 2\sum_{j\in\mathbb{I}}\langle\tilde{L}_{\mathbb{I},j}(x - \sum_{i\in\mathbb{I}}\tilde{L}^*_{\mathbb{I},i}(c_i)),\varepsilon_j\rangle. \end{aligned} \tag{4.18}$$

We now assume that ε_j et $\tilde{L}_{\mathbb{I},j}(x - \sum_{i\in\mathbb{I}} \tilde{L}^*_{\mathbb{I},i}(c_i))$ are realizations of the random vectors $E_j = (E_{j,k})_{1\leq k\leq K_j}$ and $Z_j = (Z_{j,k})_{1\leq k\leq K_j}$. For a fine enough quantization step, one can assume that the $E_{j,k}$ are centered i.i.d and equally independent of $Z_{j,k}$. In this case, the variables $(E_{j,k}Z_{j,k})_{1\leq k\leq K_j}$ are centered and decorrelated.

Moreover, the condition (4.12) implies that the variances of the random variables $Z_{i,k}$ are uniformly bounded in (i,k). By letting K_j tend to infinity we obtain

$$K_j^{-1}\sum_{k=1}^{K_j} Z_{j,k}E_{j,k} \xrightarrow{a.s.} 0. \tag{4.19}$$

And thus, for large values of $(K_j)_{1\leq j\leq D}$ the scalar products in (4.18) can be neglected. Similarly, the term corresponding to the error can be put under the form,

$$D_{\varepsilon,\mathbb{I}} = \|\sum_{i\in\mathbb{I}} \tilde{L}^*_{\mathbb{I},i}(\varepsilon_i)\|^2 = \sum_{i\in\mathbb{I}}\sum_{j\in\mathbb{I}}\sum_{k=1}^{K_i}\sum_{l=1}^{K_j} E_{i,k}E_{j,l}\langle u_{\mathbb{I},i,k}, u_{\mathbb{I},j,l}\rangle, \tag{4.20}$$

where, for each $i \in \{1,\dots,D\}$, $\{u_{\mathbb{I},i,k}\}_{1\leq k\leq K_i}$ is the family of vectors in $\mathcal{H}$ associated to the linear synthesis operator $\tilde{L}^*_{\mathbb{I},i}$.

This term is independent of the coefficients $c_{i,k}$ and can be estimated from the variance of $E_{1,1}$. In some cases, especially when the union of $\tilde{L}^*_{\mathbb{I},i}$ forms a basis and for large values of K_i, the law of large numbers gives:

$$D_\varepsilon = \sum_{\mathbb{I}\in\mathcal{P}} \alpha_{\mathbb{I}} D_{\varepsilon,\mathbb{I}} \sim \mathbb{E}(E_{1,1}^2)\sum_{\mathbb{I}\in\mathcal{P}}\alpha_{\mathbb{I}}\sum_{i\in\mathbb{I}} K_i. \tag{4.21}$$

Thus, the distortion constraint (4.7) is brought to a quadratic constraint on the coefficients c_i, given by:

$$G(c_1,\dots,c_D) \leq G_{\max} = D_{\max} - D_\varepsilon. \tag{4.22}$$

with $G(c_1,\dots,c_D) = \sum_{\mathbb{I}\in\mathcal{P}} \alpha_{\mathbb{I}}\|x - \sum_{i\in\mathbb{I}} \tilde{L}^*_{\mathbb{I},i}c_i\|^2$.

Finally, we can say that the initial optimization problem is reformulated to minimizing J under a convex quadratic constraint. Note that, for the case in which $\beta_{i,k} \equiv 1$ and $\omega_{i,k} \equiv 1$, this optimization problem is quite similar to the one addressed in compressed sensing.

4.3 Convex optimization

Let us now address the convex optimization problem previously defined. Classically, this problem is solved by finding the critical points of the Lagrangian. One must thus solve:

$$\max_{\mu\geq 0}\ \min_{(c_1,\dots,c_D)} \left(J(c_1,\dots,c_D) + \mu(G(c_1,\dots,c_D) - G_{\max}\right). \tag{4.23}$$

The main difficulty in (4.23) is at the minimization stage. This problem is generally tackled by iterative algorithms. We suggest using the algorithm given in [CCPW07], which is itself an extension of the methods in [DDM04]. To this end, we set the initial values $(c_1^{(0)},\dots,c_D^{(0)})$

and we generate the sequence $(c_1^{(n)}, \dots, c_D^{(n)})_{n\geq 1}$ which converges toward the solution of the optimization problem. At the n^{th} iteration, we compute

$$\pi_{i,k}^{(n)} = \text{prox}_{\gamma\omega_{i,k}|\,.\,|^{\beta_{i,k}}}(c_{i,k}^{(n)} - \gamma g_{i,k}^{(n)}) \tag{4.24}$$

$$c_{i,k}^{(n+1)} = c_{i,k}^{(n)} + \lambda(\pi_{i,k}^{(n)} - c_{i,k}^{(n)}) \tag{4.25}$$

for all $i \in \{1, \dots, D\}$ et $k \in \{1, \dots, K_i\}$.

Here, $\gamma \in [0, \gamma_{\max}]$ is the algorithm step and λ is a relaxation parameter. The vectors $(g_{i,k}^{(n)})_{i=1,\dots,D;\ 1\leq k\leq K_i}$ are given by:

$$(g_{i,k}^{(n)})_{1\leq k\leq K_i} = 2\sum_{\mathbb{I}\in\mathcal{P}} \mathbb{1}_{\{i\in\mathbb{I}\}}\alpha_{\mathbb{I}}\tilde{L}_{\mathbb{I},i}\Big(\sum_{i\in\mathbb{I}}\tilde{L}_{\mathbb{I},i}^{*}c_i^{(n)} - x\Big). \tag{4.26}$$

Recall that $\text{prox}_{\gamma\omega_{i,k}|\,.\,|^{\beta_{i,k}}}$ is the proximal operator of the function $\gamma\omega_{i,k}|\,.\,|^{\beta_{i,k}}$. The proximal operator of a convex function, $f : \mathbb{R} \to \mathbb{R}$, is defined by

$$\text{prox}_f : u \mapsto \text{argmin}_v \frac{1}{2}(v-u)^2 + f(v).$$

According to [CCPW07], for the specific case of our functions, this operator can be explicitly computed for particular values of $\beta_{i,k}$, and easily computed numerically, otherwise.

4.3.1 Proof of theoretical results

Let us first develop the transition from ((4.8)) to (4.17).

Lemma 2. *For all $i = 1, \dots, D$ and large values of K_i the discrete entropy of $\overline{C}_{i,k}$ is equivalent to:*

$$J(c_1, \dots, c_D) = \sum_{i=1}^{D}\sum_{k=1}^{K_i}\omega_{i,k}|c_{i,k}|^{\beta_{i,k}}. \tag{4.27}$$

Proof. As seen in (4.16), the weak law of large numbers serves in obtaining the proof of this Lemma. Let us first verify that this theorem can be applied. For each $C_{i,k}$, we define the random variable $Y_{i,k}$ of probability law given by $\log_2(C_{i,k})$. We first verify the independence of the random variables $(Y_{i,k})_{i=1,\dots,D,1\leq k\leq K_i}$.

More easily, if $(C_i)_{1\leq i\leq n}$ are independent then, for any function f_i with $i \in \{1, \dots, n\}$, $f_i(C_i)$ are independent. This can be viewed as the fact that the components of a random vector $(X_k)_{1\leq k\leq N}$ are independent if the probability density, f, of the N-tuple verifies that N functions, f_i, exist such that:

$$f(x_1, \dots, x_N) = f_1(x_1)\dots f_N(x_N).$$

Let $p : \mathbb{R}^N \to \mathbb{R}$ be the probability density of the N-tuple $Y_{i,k}$, where $N = K_1 + \dots + K_D$ and the variables are reordered. We have, by definition:

$$p(Y_1, \dots, Y_N) = e^{y_1+\dots+y_N}p_C(e^{y_1}, \dots, e^{y_N}).$$

where p_C is the density of the N-tuple $(C_{i,k})_{i=1,\dots,D,1\leq k\leq K_i}$. But since $C_{i,k}$ are considered to be independent we have:

$$p(Y_1, \dots, Y_N) = e^{y_1}p_{1,1}(e^{y_1})\dots e^{y_N}p_{D,K_D}(e^{y_N}).$$

In other words, the random variables $(Y_{i,k})_{i=1,\dots,D,1\leq k\leq K_i}$ are independent.

Set $i \in \{1,\dots,D\}$ and $l \in \{1,\dots,N_i\}$. By the definition of the set $\mathcal{S}_{i,l}$, the random variables $(C_{i,k})_{k\in\mathcal{S}_{i,l}}$, and thus the random variables $(\log_2(C_{i,k}))_{k\in\mathcal{S}_{i,l}}$, are identically distributed. Moreover, we assume that when K_i tends toward infinity, $\mathrm{Card}\,\mathcal{S}_{i,l}$ also tends toward infinity. In this case one can apply the law of large number and this gives:

$$-\frac{1}{\mathrm{Card}(\mathcal{S}_{i,l})}\sum_{k\in\mathcal{S}_{i,l}}\log_2(p_{i,k}(C_{i,k}))\xrightarrow{\text{a.s.}}h(C_i^{(l)}).$$

Let us now look at H_{global}. This function is defined by:

$$H_{\text{global}}=\sum_{i=1}^{D}\sum_{k=1}^{K_i}H(\overline{C}_{i,k}).$$

And we have seen that $H(\overline{C}_{i,k})\sim h(C_{i,k})-\log_2(q)$ where $q>0$ is the quantization step. Then

$$H_{\text{global}}\sim\sum_{i=1}^{D}\sum_{k=1}^{K_i}h(C_{i,k})-\log_2(q).$$

But, as seen above we can group the $C_{i,k}$ into packets. In other words:

$$H_{\text{global}}\sim\sum_{i=1}^{D}\sum_{l=1}^{N_i}\mathrm{Card}(\mathcal{S}_{i,l}h(C_i^{(l)})-\log_2(q)).$$

And from (4.16), when $\mathrm{Card}(\mathcal{S}_{i,l})$ is large,

$$h(C_i^{(l)})\sim-\frac{1}{\mathrm{Card}(\mathcal{S}_{i,l})}\sum_{k\in\mathcal{S}_{i,l}}\log_2(p_{i,k}(C_{i,k}))$$

Therefore,

$$\begin{aligned}H_{\text{global}}&\sim\sum_{i=1}^{D}\sum_{l=1}^{N_i}\sum_{k\in\mathcal{S}_{i,l}}-\log_2(p_{i,k}(C_{i,k}))-\log_2(q)\\&\sim\sum_{i=1}^{D}\sum_{k=1}^{K_i}(\log_2(p_{i,k}(C_{i,k}))-\log_2(q)).\end{aligned}$$

□

The second point that we are going to prove is taking the limit (4.19), which allows for neglecting the inner products when computing the distortion.

Proof. A faster way for this proof can be given by invoking the Theorem 19.4, p. 295 of Davidson [Dav94] which, in addition, ensures the almost sure convergence.

Let us adopt the notation: $S_{K_j}=\sum_{k=1}^{K_j}E_{j,k}Z_{j,k}$.

Since the $E_{j,k}$ are centered and both mutually independent and from $Z_{j,k}$, we have:

$$\mathbb{E}(S_{K_j}) = \sum_{k=1}^{K_j} \mathbb{E}(E_{j,k}Z_{j,k}) = \sum_{k=1}^{K_j} \mathbb{E}(E_{j,k})\mathbb{E}(Z_{j,k}) = 0,$$

and

$$\begin{aligned}
\mathbb{E}(S_{K_j}^2) &= \mathbb{E}(\sum_{k=1}^{K_j}(E_{j,k}Z_{j,k}))^2 \\
&= \sum_{k=1}^{K_j} \mathbb{E}(E_{j,k}^2 Z_{j,k}^2) + \sum_{k=1}^{K_j}\sum_{k'\neq k} \mathbb{E}(E_{j,k}Z_{j,k}E_{j,k'}Z_{j,k'}) \\
&= \sum_{k=1}^{K_j} \mathbb{E}(E_{j,k}^2)\mathbb{E}(Z_{j,k}^2) \\
&+ \sum_{k=1}^{K_j}\sum_{k'\neq k} \mathbb{E}(E_{j,k})\mathbb{E}(E_{j,k'})\mathbb{E}(Z_{j,k}Z_{j,k'}) \\
&= \mathbb{E}(E_{1,1}^2)\sum_{k=1}^{K_j} \mathbb{E}(Z_{j,k}^2).
\end{aligned}$$

By applying the Bienaymé-Tchebychev inequality to the random variable S_{K_j} we obtain that $\forall t > 0$,

$$\mathbb{P}(|S_{K_j}| \geq t) \leq \frac{\mathbb{E}(S_{K_j}^2)}{t^2}.$$

Then we can set $t = K_j\eta$ for all $\eta > 0$ and

$$\mathbb{P}(|S_{K_j}| \geq K_j\eta) \leq \frac{\mathbb{E}(S_{K_j}^2)}{(K_j\eta)^2}.$$

Therefore,

$$\mathbb{P}(\frac{|S_{K_j}|}{K_j} \geq \eta) \leq \frac{\mathbb{E}(E_{1,1}^2)\sum_{k=1}^{K_j}\mathbb{E}(Z_{j,k}^2)}{(K_j\eta)^2}.$$

Moreover, since the probability densities $p_{i,k}$ related to the law of the $Z_{i,k}$ are parametered by: $(\beta_{i,k}, \omega_{i,k}) \in \mathcal{S}$, we can write that:

$$\begin{aligned}
\sum_{k=1}^{K_j} \mathbb{E}(Z_{j,k}^2) &= \sum_{l=1}^{N_j} \operatorname{Card}\mathcal{S}_{i,l}\mathbb{E}(Z_i^{(l)\,2}) \\
&\leq \sup_{l=1,\ldots,N_j} \mathbb{E}(Z_i^{(l)\,2}) \sum_{l=1}^{N_j} \operatorname{Card}\mathcal{S}_{i,l} = \sup_{l=1,\ldots,N_j} \mathbb{E}(Z_i^{(l)\,2})K_j,
\end{aligned}$$

where $Z_i^{(l)}$ is the random variable indexed by $(\beta_i^{(l)}, \omega_i^{(l)}) \in \mathcal{S}$.

Finally, we obtain that for all $\eta > 0$,

$$\mathbb{P}(\frac{|S_{K_j}|}{K_j} \geq \eta) \leq \frac{K_j \mathbb{E}(E_{1,1}^2)c}{(K_j\eta)^2} = \frac{\mathbb{E}(E_{1,1}^2)c}{K_j\eta^2}.$$

which tends to zero when K_j tends to infinity. □

We are now concentrating on estimating the quantization error in the distortion, or in other words the transition to (4.21).

Proof. $\tilde{L}^*_{\mathbb{I},i}$ are square summable linear operators in $\ell^2(\mathbb{K}_i)$ taking their values in $\mathcal{H}$. Thus, for each $i \in \{1, \dots, D\}$, it exists a vector $(u_{i,k})_{1\leq k\leq K_i}$ such that for all $k = 1, \dots, K_i$, $u_{i,k} \in \mathcal{H}$. From (4.20), the error obtained in the expression of the distortion for each configuration $\mathbb{I} \in \mathcal{P}$ is:

$$\begin{aligned} D_{\varepsilon,\mathbb{I}} &= \| \sum_{i\in\mathbb{I}} \tilde{L}^*_{\mathbb{I},i}(\varepsilon_i)\|^2 = \sum_{i\in\mathbb{I}} \sum_{j\in\mathbb{I}} \langle \tilde{L}^*_{\mathbb{I},i}(E_i), \tilde{L}^*_{\mathbb{I},j}(E_j)\rangle \\ &= \sum_{i\in\mathbb{I}} \sum_{j\in\mathbb{I}} \langle \tilde{L}_{\mathbb{I},j}(\tilde{L}^*_{\mathbb{I},i}(E_i)), E_j\rangle \end{aligned}$$

But, by definition,

$$\begin{aligned} \tilde{L}_{\mathbb{I},j}(\tilde{L}^*_{\mathbb{I},i}(E_i)) &= (\langle \tilde{L}^*_{\mathbb{I},i}(E_i), u_{j,l}\rangle)_{1\leq l\leq K_j} \\ &= (\sum_{k=1}^{K_i} E_{i,k}\langle u_{i,k}, u_{j,l}\rangle)_{1\leq l\leq K_j}. \end{aligned}$$

□

Finally, with the necessary replacements we obtain (4.20). For practical reasons we reorder $i \in \mathbb{I}$, such that $K_1 \leq K_2 .... \leq K_{\mathrm{Card}\,\mathbb{I}}$. We can thus rewrite:

$$\begin{aligned} D_{\varepsilon,\mathbb{I}} = &\sum_{i\in\mathbb{I}} \sum_{j>i} K_i K_j \frac{1}{K_j} \sum_{l=1}^{K_j} \frac{1}{K_i} \sum_{k=1}^{K_i} E_{j,l}E_{i,k}\langle u_{i,k}, u_{j,l}\rangle \\ &+ \sum_{i\in\mathbb{I}} \sum_{j\leq i} K_i K_j \frac{1}{K_i} \sum_{l=1}^{K_i} \frac{1}{K_j} \sum_{k=1}^{K_j} E_{j,l}E_{i,k}\langle u_{i,k}, u_{j,l}\rangle \end{aligned}$$

Let us first look at the first term of the summation. In this case we are considering the convergence in mean of the random variable $S_{i,j}$ defined by

$$S_{i,j} = \sum_{l=1}^{K_j} \sum_{k=1}^{K_i} E_{j,l}E_{i,k}\langle u_{i,k}, u_{j,l}\rangle. \tag{4.28}$$

If we compute

$$\begin{aligned} \mathbb{E}(|S_{i,j}|) &\leq \sum_{l=1}^{K_j} \sum_{k=1}^{K_i} \mathbb{E}(|E_{j,l}E_{i,k}|)|\langle u_{i,k}, u_{j,l}\rangle| \\ &\leq \mathbb{E}(|E_{1,1}|^2) \sum_{l=1}^{K_j} \sum_{k=1}^{K_i} |\langle u_{i,k}, u_{j,l}\rangle|. \end{aligned} \tag{4.29}$$

Therefore, if $(u_{i,k})_{i=1,\dots,D;\ 1\leq k\leq K_i}$ form a basis, the scalar products are nulls and we have convergence in mean to zero since $i \neq j$ and thus $E_{i,k}$ and $E_{j,l}$ are independent for all $k = 1,\dots,K_i$ and all $l = 1,\dots,K_j$, respectively. Moreover, using the independence of the random variables $(E_{i,k})_{i=1,\dots,D;\ 1\leq k\leq K_i}$,

$$\begin{aligned}\mathrm{Var}(S_{i,j}) &= \sum_{l=1}^{K_j}\sum_{k=1}^{K_i}\sum_{l'=1}^{K_j}\sum_{k'=1}^{K_i}\mathbb{E}(E_{j,l}E_{i,k}E_{j,l'}E_{i,k'})\langle u_{i,k},u_{j,l}\rangle\langle u_{i,k'},u_{j,l'}\rangle\\ &= \sum_{l=1}^{K_j}\sum_{k=1}^{K_i}\mathbb{E}(E_{j,l}^2)\mathbb{E}(E_{i,k}^2)\langle u_{i,k},u_{j,l}\rangle^2\\ &= \mathbb{E}(E_{1,1}^2)^2\sum_{l=1}^{K_j}\sum_{k=1}^{K_i}\langle u_{i,k},u_{j,l}\rangle^2.\end{aligned}$$

Since $(u_{i,k})_{i=1,\dots,D;\ 1\leq k\leq K_i}$ are unitary vectors, and with the Cauchy-Schwarz inequality, we obtain

$$\mathrm{Var}(S_{i,j}) = \mathbb{E}(S_{i,j}) \leq \mathbb{E}(E_{1,1}^2)^2 K_i K_j. \tag{4.30}$$

We then apply the Bienaymé-Tchebychev inequality to the random variable $S_{i,j}$ and we obtain that $\forall t > 0$,

$$\mathbb{P}(|S_{i,j}| \geq t) \leq \frac{\mathbb{E}(S_{i,j}^2)}{t^2}.$$

Then we set $t = K_jK_i\eta$, for all $\eta > 0$ and

$$\mathbb{P}(|S_{i,j}| \geq K_iK_j\eta) \leq \frac{\mathbb{E}(S_{i,j}^2)}{(K_jK_i\eta)^2}.$$

Thus, from (4.30)

$$\mathbb{P}(|\frac{S_{i,j}}{K_iK_j}| \geq \eta) \leq \frac{\mathbb{E}(E_{1,1}^2)^2}{K_jK_i\eta^2}.$$

And this tends to zero when K_i and K_j tend toward infinity. In the second term of this summation we similarly obtain that the corresponding sums tend to zero for $i \neq j$. Therefore, for large K_i, and since it is assumed that $\|u_{k,i}\| = 1$,

$$D_{\varepsilon,\mathbb{I}} \sim \sum_{i\in\mathbb{I}}\sum_{k=1}^{K_i} E_{i,k}^2.$$

By applying once more the law of large numbers we can write that:

$$\frac{1}{K_i}\sum_{k=1}^{K_i} E_{i,k}^2 \xrightarrow{a.s.} \mathbb{E}(E_{1,1}^2). \tag{4.31}$$

And this leads to (4.21).

4.4 Example and numerical results

Let us consider a simple example inspired from the JPEG2000 standard. For a two description scenario we can build a frame-based scheme by using the reconstruction operator L_1^* associated to a 9-7 biorthogonal wavelet basis and the reconstruction operator L_2^*, associated to the same wavelet basis functions but shifted by 1 pixel in each spatial direction. In this case, a natural choice for the side decoders is $\widetilde{L}_i^* = 2L_i^*$, $i \in \{1, 2\}$. A 3-resolution level dyadic filter bank structure is applied for the MDC encoding of the 512×512 standard Lena image. The weighting factors in the distortion constraint have been chosen here as $\alpha_{1,2} = 0.8$, $\alpha_1 = \alpha_2 = 0.1$. The frame coefficients are synthesized by the optimization approach described in the previous section where the parameters of the generalized Gaussian model have been estimated by an iterative Maximum Likelihood method. The quantization step, q, has been optimized for each rate, and the JPEG2000 algorithm has been employed to encode the two quantized descriptions. Fig. 4.3 shows the evolution of the PSNR w.r.t. the global bitrate for the central and side decoders. For comparison, the results corresponding to the direct application of the JPEG2000 encoder at half the bitrate are provided. As expected, the proposed scheme provides better results for the central decoder while showing a good performance for the two side decoders. It is worth noticing that better results could be expected by using more sophisticated frames.

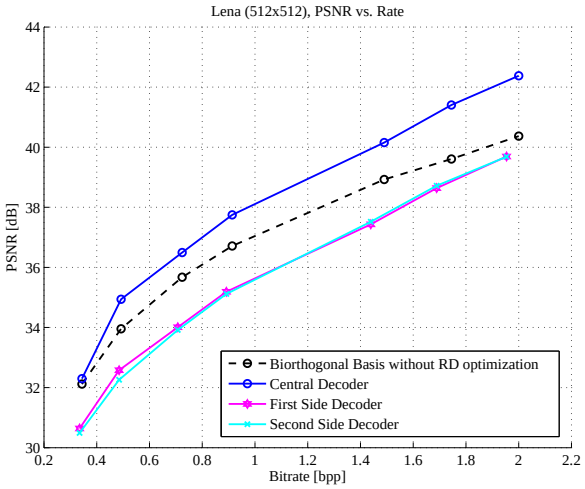


FIGURE 4.3: Rate-distortion performance of the two-description scheme for Lena.

Let us now look at a similar scenario having three descriptions this time, where we have chosen the synthesis operators L_1^*, L_2^* and L_3^* corresponding as before to a biorthogonal 9/7 wavelet basis and two shifted versions (the $(1,1)$-shift for L_2^* and the $(1,0)$-shift for L_3^*). These correspond to the three descriptions denoted D_1, D_2, and D_3, respectively, in Figure 4.4.

Now we have two types of side decoders: one which receives only one out of three descriptions and the other which receives two out of the three. Therefore a natural choice for $\tilde{L}_{\mathbb{I},i}^*$ is the following:

$$\tilde{L}_{\mathbb{I},i}^* = \begin{cases} \frac{3}{2}L_i^*, & \text{if } \mathrm{Card}(\mathbb{I}) = 2 \\ 3L_i^*, & \text{if } \mathrm{Card}(\mathbb{I}) = 1. \end{cases} \tag{4.32}$$

The weights involved in the distortion constraint take the following values: $\alpha_{\{1,2,3\}} = 0.8$,

$\alpha_{\mathbb{I}} = 0.0618$ when $\mathrm{Card}(\mathbb{I}) = 2$ and $\alpha_{\mathbb{I}} = 0.0048$, when $\mathrm{Card}(\mathbb{I}) = 1$.

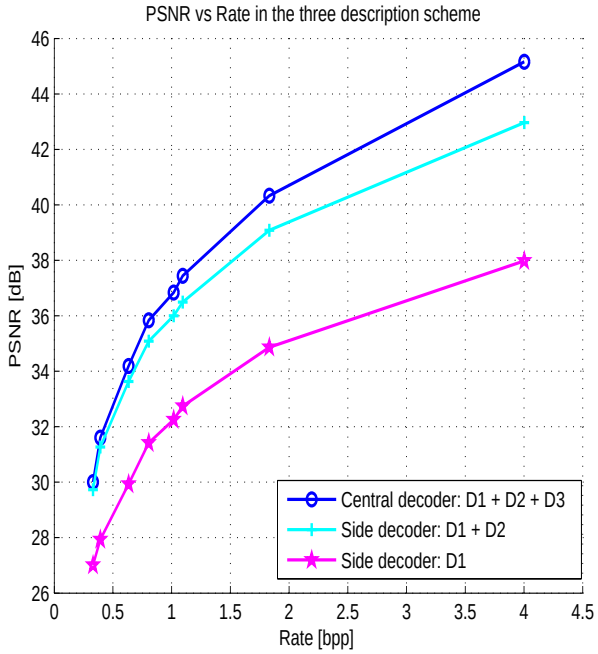


FIGURE 4.4: Rate-distortion performance of the three-description scheme for Lena.

For perceptual quality assessment we equally give the reconstructed images corresponding to the central rate of 0.8 bpp.

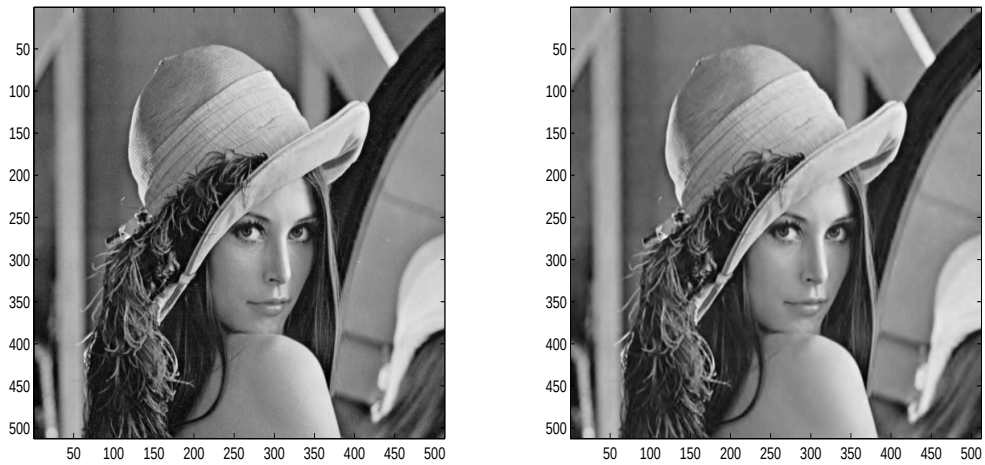


FIGURE 4.5: Reconstruction: Original image (left), Central decoder $D_1 + D_2 + D_3$ (right) at 0.8 bpp.

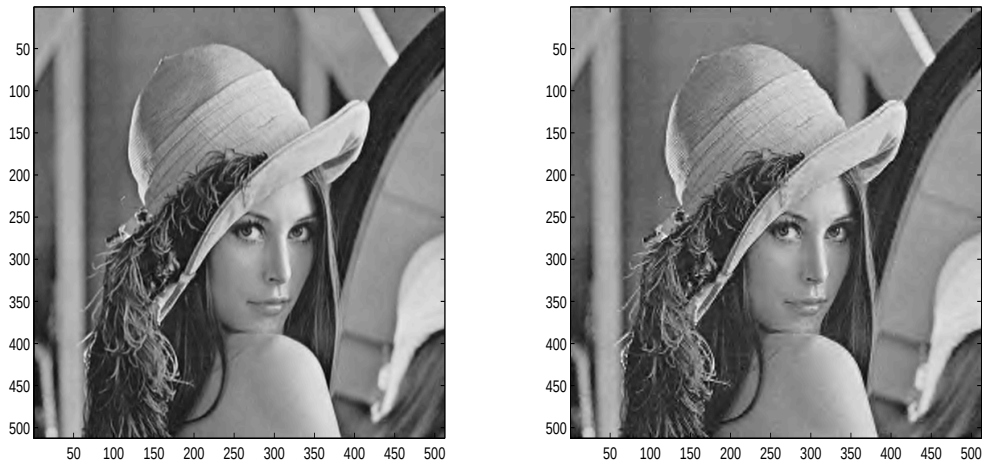


FIGURE 4.6: Reconstruction: Side decoder $D_1 + D_2$ (left), Side decoder D_1 (right).

4.5 Synthesis frame approach versus the classical MDC approach

This chapter gave a complementary approach to the previously presented wavelet multiple description schemes. Recall that this allowed us to shift the computational complexity to the encoder side leaving the decoder fast and linear. The shift in complexity however is only a side effect and not a purpose per se. The main idea of the synthesis frame approach was to better take into account the joint source-channel coding aspect by choosing the "good" quantization steps that would lead to a given maximum distortion.

In the previous chapter the problem was reversed. The encoder had only the purpose of minimizing the redundancy while preserving the perfect reconstruction, whereas the decoder had the task of combining the correctly received data in an efficient manner by a post-processing optimization step.

Even though the approach in Chapter 3 and the current one are complementary in some sense, it would be useful to see how far apart in performance these two are, without taking into consideration the fact that one is preferable in a given scenario (for instance the classical approach could work well for simple encoders and complex decoders) and the other in an orthogonal scenario (for instance a broadcast application in which the decoders need to be fast and simple).

Before presenting such a comparison we need to do some simplifying hypothesis however, because the MDC approach chosen in this chapter in order to validate the theoretical concept is very basic in terms of the chosen frame and the amount of introduced redundancy. In this context, we shall compare the synthesis frame approach with a classical two-description approach based on the same analysis frame in which the quantization step is chosen to yield the same bitrates. This is more of a preliminary result that could be useful in continuing the research in the synthesis frame approach.

The outcome of this comparison for the Barbara image (512×512 pixels) is given in Figure 4.7. On the left-hand side we present the central decoder obtained when both descriptions have been received, as well as the critically sampled biorthogonal $9/7$ wavelet decomposition

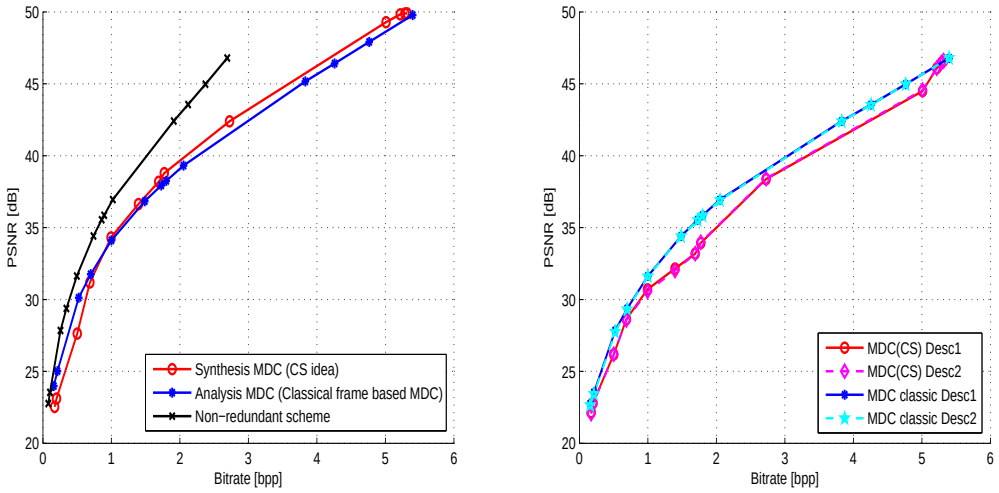


FIGURE 4.7: Comparison between the synthesis frame approach and a classical MDC approach using an equivalent redundant wavelet frame, for Barbara and a two description scenario: Central decoders (left), Side decoders(right). For the central decoder we also give the PSNR vs rate curve for a critically sampled biorthogonal 9/7 wavelet decomposition.

for performance boundary purposes. On the right-hand side we can observe the rate-distortion performances of the side decoders for the two MDC schemes. These results exhibit comparable performances between the central decoders, with the novel scheme performing slightly better at high bitrates. This confirms the fine quantization approximation and opens new perspectives for the synthesis frame scheme. If more refined approximations are to be done, the performances of the scheme could enhance drastically. When compared to the single description situation (i.e. the critically sampled decomposition), these results are quite bad, but this is hardly a surprise considering the great amount of redundancy (double) at the central decoders.

The side decoders results slightly contradict the central decoder ones which could indicate that refinements in the used hypothesis in the synthesis frame approach are in order.

4.6 Conclusions

In this chapter we have presented a novel approach to Multiple Description Coding in which the focus is on efficiently generating sequences of quantized coefficients at the encoder, in a frame-based scenario. We have imposed linearity at the decoder and we have proposed a convex formulation of the rate-distortion optimization problem to be solved at the encoder. By optimizing the rate at the encoder the smallest number of coefficients is selected thus leading to a sparse representation of the source.

Chapter 5

Conclusion & Future Work

In this thesis we have addressed the problem of data transmission over error-prone networks, by employing a technique called Multiple Description Coding (MDC). Such methods, which define a joint source-channel coding strategy, have gained popularity as an alternative to error correcting codes because they handle efficiently much larger data-chunk losses. In addition, they are much more flexible to variable network conditions such as changing bandwidth, different size displays, multiple available transmission channels etc.

We have chosen the transform-based approach to MDC, since this enables the use of inherently redundant wavelet transforms for signal encoding. In this manner we introduce information diversity in the transform domain of the signal and simultaneously benefit from the scalability induced by these kind of multiresolution transforms. This thesis focused on building reduced-redundancy schemes, and, more concretely, we have studied schemes in which the redundancy is tunable with the size of an approximation subband in a classical wavelet decomposition.

In the following, we summarize the contributions of this thesis and address several perspectives for future work.

5.1 Overview of thesis contributions

The first direction, considered in Chapter 2, consisted in building redundant schemes for one-dimensional signals. In practice, we have applied the theoretical framework to the temporal signal in a $t + 2D$-encoded video sequence. The redundancy was achieved by using a wavelet frame approach which is equivalent to an oversampled filter bank (OFB) applied to the signal. Through this transform we have created two descriptions of the source signal, each containing a set of wavelet coefficients given by an approximation and three detail coefficients subsets. Each set was to be transmitted over its own independent channel. However, the union of these sets yields a redundancy of a factor of 2, which would result in a highly inefficient coding scheme if the absence of losses. Therefore, we have introduced an additional subsampling of all the detail subbands while keeping the approximation subbands entirely. This allows for the tuning of the overall redundancy of the MDC scheme to the size of an approximation subband in a critically sampled wavelet decomposition.

Next, we addressed the perfect reconstruction issue raised by this additional subsampling.

We have proven the perfect reconstruction for certain schemes and we have established choice criteria among them based on the minimization of the quantization noise. We have compared the performances of several schemes among the efficient ones in a scalable video coding context provided by the Motion Compensated - Embedded Zero-trees Block Coding (MC-EZBC) codec. We have implemented two scenarios: in the first one a whole description is lost at a time (this corresponding to a network packet per description), and in the second one we have simulated the loss of several packet into each wavelet description.

A second direction that we have explored in this thesis referred to the multiple description coding of still images, which is viewed as an extension of the temporal schemes developed in Chapter 2. The same idea of reducing the redundancy by an additional subsampling of the wavelet subbands was employed (the subbands which contain the core of the signal energy were again preserved). In the two-dimensional schemes presented in Chapter 3, the problem of structure inversion is not trivial, and therefore an exhaustive study has been conducted in order to select the efficient schemes among all possible combinations based on the proposed subsampling strategies.

Moreover, we have explored the possibility of improving the decoding by a post-processing stage based on *a priori* information on the system. This information is given by the quantization steps which can be viewed as convex constraints. The reconstruction problem has thus been formulated as the optimization of a quadratic function under convex constraints and the decoded image gains several dB in terms of Peak Signal to Noise Ratio (PSNR) both when a whole description is lost and when random pixels in each description are destroyed.

Finally, we have approached the Multiple Description problem from a different angle by considering the problem as a rate-distortion optimization in which a certain maximal distortion is allowed and the best transmission rate (linked to the quantization step) is sought. The philosophy of this approach is related to the newly expanding theory called "Compressed Sensing". This theory aims at re-thinking signal acquisition by exploiting the fact that many signals have a sparse representation in some basis. Thus, acquiring all signal samples just to discard most of them in the next step of the transmission chain might not be the most efficient way to proceed. To this end a signal recovery problem needs to be solved. Researchers have thus formalized a framework in which a number slightly bigger than the sparsity[1] of the signal suffices to recover it perfectly with very high probability.

Our multiple description approach, which rejoins this philosophy, aimed at selecting a small number of quantized coefficients at the encoder, such that the best rate-distortion trade-off was to be attained at the decoder. Some approximations have been formulated in order to be able to solve this problem as a convex optimization, and the scheme has been generalized to an arbitrary number of descriptions. By doing this, the complexity is shifted to the encoder whereas the decoding becomes a simple linear process. We have tested this framework for still images encoded with the EZBC and JPEG2000 coders. Preliminary results have been compared to a classical MDC approach for the same level of redundancy.

5.2 Perspectives for future work

The work presented in this thesis leads toward several possible extensions. In the following we identify some of them.

[1] i.e. the number of non-zero elements in the signal, either in its time domain or in some basis.

Hybrid MDC schemes for video

For the case of temporal descriptions our schemes may be further combined with other spatially redundant strategies that have already been proposed in the literature. The increased flexibility thus achieved may be exploited to better adapt packet-forming to different situations of network losses and also to improve the reconstruction at different levels.

Enhanced spatial MDC schemes

In the spatial case, the performances of our 2-description schemes might be enhanced by adding more diversity between the most energetic subbands, mainly the approximation ones. These remain highly correlated after the decomposition, therefore the central decoding cannot fully exploit this redundancy. With this in mind we propose to perform a basic multiple description scalar quantization between the two approximation subbands, consisting in shifting the quantization indices by one half - in the Vaishampayan [Vai93b] style. The reception of the two descriptions would lead to a finer quantized reconstruction of the approximation subbands. By combining this with the chosen weighting strategy for the subbands, better decoding results might be expected.

Another extension that can be brought to the spatial MDC schemes, mainly in the initialization strategy for the iterative decoding, is a finer choice of the fixed weighting coefficients of each received subband. Indeed, instead of approximating the values corresponding to the integer truncation to 1 of the quantization step, an adaptive quantization step might be chosen for each bitrate.

Sparsity approach with finer redundancy frames

The last direction explored in this thesis is incipient. Many possible extensions can be envisaged. An immediate one is to employ less redundant frames, such as those explored in the previous chapters. Also finer approximations for the entropy, better suited for the low-bitrate regime based on the works of Fraysse et al. [FPPP09] could be a possible extension.

Different transforms and/or performance measures

An important issue in any MDC approach is the quality of the side reconstruction. In our work we only considered the "usual" performance evaluation criteria, based on PSNR. Two directions are possible from this point.

The first one would be to use other transforms which enhance the subjective quality perception at comparable PSNR-s. In this sense, a great deal of effort has been spent in the literature in order to enhance the subjective quality perception based on geometry preserving transforms. We could investigate replacing the proposed filter banks by more advanced wavelet-like techniques such as: bandlets [Pen02], curvelets [CDDY06], contourlets [SW01, Chap. 4], the adaptive lifting scheme [PH02], etc. . See [SW01] and [Mal09] for surveys on other interesting wavelet-like ("x-let") transforms. We could envisage replacing our frame decomposition built as a union of bases with some of these transforms. A beneficial side effect to this would also be a sparser representation of the signal and thus much better compression performances.

The second direction to explore when evaluating the side decoding performance would be to consider other criteria, based on perceptual quality measures, for instance [vdBLV96], [dFZRS03].

Cross-domain applications

Recently MDC methods have been evaluated for the development of 3D stereoscopic video/television [NAB+06], [KHWK08]. In [KHWK08] a simple MDC method based on odd/even temporal splitting is employed starting from the works of Apostolopoulos [Apo99] for closed-loop video coders. We could imagine a somewhat similar approach using our reduced redundancy methods. In this context, MDC makes a very good candidate for the base layer encoding of the stereoscopic signal.

Finally, on a more general note, the MDC philosophy is clearly not restricted to signal processing, in its commonly accepted definition. The explosion of multi- and many-core computing architectures brings with it a stringent need for fault tolerance at different system levels: hardware, operating system, communication buses etc. In this context, the analogy between a multi-core architecture and a packet-based network is not far-fetched, especially in the case of reconfigurable architectures. The idea that some redundancy is introduced in order to cope with possible failures (either hardware or software) is very much in-line with the MDC framework.

Moreover, an important ingredient in such architectures is parallelism - yet another similarity with multiple description coding. Exploiting parallelism efficiently is one of the open problems in present day computer science community. Being applied to all pixels in an image simultaneously, the algorithms presented in this dissertation are "embarrassingly" parallel [ABC+06]. A next step would thus be to transpose the schemes presented in Chapter 3 to a parallel architecture in order to evaluate their computational performances in a more real-life scenario.

List of Figures

Bibliography

[ABC+06] K. Asanovic, R. Bodik, B. C. Catanzaro, J. J. Gebis, P. Husbands, K. Keutzer, D. A. Patterson, W. L. Plishker, J. Shalf, S. W. Williams, and K. A. Yelick. The landscape of parallel computing research: A view from berkeley. Technical report, Electrical Engineering and Computer Sciences, University of California at Berkeley, December 2006. 188

[ABE+96] J. Albanese, A. Blomer, J. Edmonds, M. Luby, and M. Sudan. Priority encoded transmission. *IEEE Transactions on Information Theory*, 42(6):1737–1744, November 1996. 56

[Agm54] S. Agmon. The relaxation method for linear inequalities. *Canadian Journal of Math*, 6(3):382–392, 1954. 142

[Ahl85] R. Ahlswede. The rate distortion region for multiple descriptions without excess rate. *IEEE Transactions on Information Theory*, 31(6):721–726, November 1985. 43

[Apo99] J.G. Apostolopoulos. Error-resilient video compression via multiple state streams. In *International Workshop on Very Low Bitrate Video Coding (VLBV)*, 1999. 34, 64, 188

[Apo01] J. Apostolopoulos. Reliable video communication over lossy packet networks using multiple state encoding and path diversity. In *Visual Communications and Image Processing (VCIP)*, San Jose, California, USA, 2001. 64

[BD86] J. P. Boyle and R. L. Dykstra. A method of finding projections onto the intersection of convex sets in hilbert spaces. *Lecture Notes in Statistics*, 37:28–47, 1986. 142

[BDR+04] R. Bernardini, M. Durigon, R. Rinaldo, L. Celetto, and A. Vitali. Polyphase spatial subsampling multiple description coding of video streams with H264. In *IEEE Int. Conf. on Image Processing*, Singapore, October 2004. 63

[BDV00] R. Balan, I. Daubechies, and V. Vaishampayan. The analysis and design of windowed Fourier frame based multiple description source coding schemes. *IEEE Transactions on Information Theory*, 46(7):2491–2536, November 2000. 52

[Ber71] T. Berger. *Rate Distortion theory*. Prentice-Hall, Englewood Cliffs, NJ, 1971. 53

[BHG02] J. Barros, J. Hagenauer, and N. Gortz. Turbo cross decoding of multiple descriptions. In *IEEE Int. Conf. on Communications*, New York, USA, 2002. 57

[BHH98] H. Bolcskei, F. Hlawatsch, and H.G. Feichtinger. Frame-theoretic analysis of oversampled filter banks. *IEEE Transactions on Signal Processing*, 46(12):3256–3268, December 1998. 52, 82

[BR05] R. Bernardini and R. Rinaldo. Efficient reconstruction from frame-based Multiple Descriptions. *IEEE Transactions on Signal Processing*, 53(8):3282–3296, August 2005. 60, 125

[BV04] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004. 139

[BW03] I. V. Bajic and J. W. Woods. Domain-based Multiple Description Coding of images and video. *IEEE Transactions on Image Processing*, 12(18):1211–1225, October 2003. 60

[BZKS06] N. V. Boulgouris, K. E. Zachariadis, A. Kanlis, and M. G. Strintzis. Multiple description wavelet coding of layered video using optimal redundancy allocation. *EURASIP Journal on Applied Signal Processing*, (2006):Article ID 83542, 19 pages, 2006. 62

[Cas00] P.G. Casazza. The art of frame theory. *Taiwanese Journal of Math*, 4(2):129–202, 2000. 50

[CCPW06] C. Chaux, P. L. Combettes, J.-C. Pesquet, and V. R. Wajs. Iterative image deconvolution using overcomplete representations. In *European Signal Processing Conference (EUSIPCO 2006)*, Sept 2006. 29, 172

[CCPW07] C. Chaux, P. L. Combettes, J.-C. Pesquet, and V. R. Wajs. A variational formulation for frame-based inverse problems. *Inverse Problems*, 23:1495–1518, 2007. 174, 175

[CDDY06] Emmanuel Candès, Laurent Demanet, David Donoho, and Lexing Ying. Fast discrete curvelet transforms. *Multiscale Modeling & Simulation*, 5(3):861–899, 2006. 34, 187

[CFH05] H.-T. Chan, C.-M. Fu, and C.-L. Huang. A new error resilient video coding using Matching Pursuit and Multiple Description Coding. *IEEE Transactions on Circuits and Systems for Video Technology*, 15(8):1047–1052, August 2005. 62

[CGPPT07] O. Crave, C. Guillemot, B. Pesquet-Popescu, and C. Tillier. Robust video transmission based on distributed multiple description coding. In *European Signal Processing Conference (EUSIPCO), Poznan, Poland*, September, 2007. 61

[CHG03] J. Chakareski, S. Han, and B. Girod. Layered coding vs. multiple descriptions for video streaming over multiple paths. In *MULTIMEDIA '03: Proceedings of the eleventh ACM international conference on Multimedia*, pages 422–431, New York, NY, USA, 2003. ACM. 41

[CKS01] H. Coward, R. Knopp, and S. D. Servetto. On the performance of multiple description codes over bit error channels. In *Proc. IEEE Int. Symp. on Information Theory*, Washington, DC, USA, June 2001. 57

[CLJB05] S. S. Channappayya, J. Lee, R. W. Heath Jr., and A. C. Bovik. Frame based Multiple Description coding in the wavelet domain. In *IEEE Int. Conf. on Image Processing*, Genoa, Italy, September 2005. 60

[CM06] P. A. Chou and Z. Miao. Rate-distortion optimized streaming of packetized media. *IEEE Transactions on Multimedia*, 8(2):390–404, April 2006. 57

[CMW99] P.A. Chou, S. Mehrotra, and A. Wang. Multiple Description decoding of overcomplete expansions using Projections onto Convex Sets. In *Data Compression Conference*, pages 72–81, Snowbird, UT, USA, March 1999. 28, 60, 123, 145

[Com93] P. L. Combettes. The foundations of set theoretic estimation. *Proceedings of the IEEE*, 81(2):182–208, February 1993. 140, 141, 142

[Com97] P. L. Combettes. Convex set theoretic image recovery by extrapolated iterations of parallel subgradient projections. *IEEE Transactions on Image Processing*, 6(4):493–506, 1997. 140, 141, 145

[Com03] P. L. Combettes. A block-iterative surrogate constraint splitting method for quadratic signal recovery. *IEEE Transactions on Signal Processing*, 51:1771–1782, July 2003. 123, 142, 143, 145

[CSOM03] D. Comas, R. Singh, A. Ortega, and F. Marques. Unbalanced multiple-description video coding with rate-distortion optimization. *EURASIP Journal on Applied Signal Processing*, 2003(1):81–90, 2003. 64

[CV98] Z. Cvetkovic and M. Vetterli. Oversampled filter banks. *IEEE Transactions on Signal Processing*, 46(5):1245–1255, May 1998. 52, 82

[Cve95] Z. Cvetkovic. Oversampled modulated filter banks and tight gabor frames in l^2(z). In *IEEE Int. Conf. on Acoustics, Speech and Signal Proc.*, volume 2, pages 1456–1459 vol.2, May 1995. 52

[CW99a] S.J. Choi and J.W. Woods. Motion-compensated 3-D subband coding of video. *IEEE Transactions on Image Processing*, 8:155–167, 1999. 103, 111

[CW99b] D.-M. Chung and Y. Wang. Multiple description image coding using signal decomposition and reconstruction based on lapped orthogonal transforms. *IEEE Transactions on Circuits and Systems for Video Technology*, 9(6):895–908, September 1999. 60

[CWP03] P. Chou, H. Wang, and V. Padmanabhan. Layered multiple description coding. In *Packet Video Workshop*, Nantes, France, April 2003. 56

[Dau88] I. Daubechies. Orthonormal bases of compactly supported wavelets. *Commun. on Pure and Appl. Math.*, 41:909–996, 1988. 52

[Dau92] I. Daubechies. *Ten lectures on wavelets*. Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, 1992. 74

[Dav94] James Davidson. *Stochastic Limit Theory - Advanced Texts in econometrics*. Oxford University Press, November 1994. 176

[DDM04] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Comm. Pure Appl. Math.*, 57:1413–1457, 2004. 174

[DDWB06] M. F. Duarte, M. A. Davenport, M. B. Wakin, and R.G. Baraniuk. Sparse signal detection from incoherent projections. In *IEEE Int. Conf. on Acoustics, Speech and Signal Proc.*, 2006. 169

[dFZRS03] R. de Freitas Zampolo, R., and Seara. A measure for perceptual image quality assessment. In *Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on*, volume 1, pages I–433–6 vol.1, Sept. 2003. 34, 188

[Don06] D.L. Donoho. Compressed Sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006. 169

[DS52] R.J. Duffin and A.C. Schaeffer. A class of nonharmonic Fourier series. *Transactions of the American Mathematical Society*, 72(2):341–366, March 1952. 50

[DS98] I. Daubechies and W. Sweldens. Factoring wavelet transforms into lifting steps. *Journal of Fourier Analysis and Applications*, 4(3):245–267, 1998. 18, 49, 96, 137

[DSV02a] S.N. Diggavi, N.J.A. Sloane, and V.A. Vaishampayan. Asymmetric multiple description lattice vector quantizers. *IEEE Transactions on Information Theory*, 48(1):174–191, January 2002. 47

[DSV02b] P. L. Dragotti, S. D. Servetto, and M. Vetterli. Optimal filter banks for multiple description coding: analysis and synthesis. *IEEE Transactions on Information Theory*, 48(7):2036–2052, July 2002. 54

[ezb] 3D MC-EZBC Software package. available on the MPEG CVS repository. 111

[FFLT05] N. Franchi, RM. Fumagalli, R. Lancini, and S. Tubaro. Multiple description video coding for scalable and robust transmission overIP. *IEEE Transactions on Circuits and Systems for Video Technology*, 15(3):321–334, March 2005. 61, 63

[Fow05] J. E. Fowler. The redundant discrete wavelet transform and additive noise. *IEEE Signal Processing Letters*, 12:629–632, 2005. 70

[FPP07] J. E. Fowler and B. Pesquet-Popescu. Wavelets in source coding, communications, and networks: An overview. *EURASIP Journal on Image and Video Processing*, 2007:27 pages, 2007. Article ID 60539. 14, 67

[FPPP09] A. Fraysse, B. Pesquet-Popescu, and J.-C. Pesquet. On the uniform quantization of a class of sparse sources. *IEEE Transactions on Information Theory*, 55(7):3243–3263, July 2009. 34, 187

[FSF06] S.E. Flierl, T. Sikora, and P. Frossard. Unbalanced quantized multiple state video coding. *EURASIP Journal on Applied Signal Processing*, 2006:1–10, 2006. 64

[FV97] E. Fornasini and M. E. Valcher. nD Polynomial Matrices with Applications to Multidimensional Signal Analysis. *Kluwer Academic Publishing, Multidimensional Systems and Signal Processing*, 8(4):387–408, October 1997. 82

[Gas66] N. Gastinel. *Analyse numérique linéaire*. Paris: Hermann, 1966. 142

[GC82] A. Gamal and T. Cover. Achievable rates for multiple descriptions. *IEEE Transactions on Information Theory*, 28(6):851–857, November 1982. 42

[GG92] Allen Gersho and Robert M. Gray. *Vector quantization and signal compression*. Kluwer Academic Publishers, 1992. 45, 46

[GGF02] T. Guionnet, C. Guillemot, and E. Fabre. Soft decoding of Multiple Descriptions. In *IEEE International Conference on Multimedia and Expo*, Lausanne, Switzerland, August 2002. 47

[GGM06] T. Gan, L. Gan, and K.-K. Ma. Reducing video-quality fluctuations for streaming scalable video using unequal error protection, retransmission, and interleaving. *IEEE Transactions on Image Processing*, 15(4):819–832, 2006. 63

[GK98] V.K. Goyal and J. Kovacevic. Optimal multiple description transform coding of Gaussian vectors. *Proc. IEEE Data Compression Conference*, pages 388–397, March 30- April 1 1998. Snowbird, Utah, USA. 44, 49

[GK01] V.K. Goyal and J. Kovacevic. Generalized multiple description coding with correlating transforms. *IEEE Transactions on Information Theory*, 47(6):2199 – 2224, September 2001. 43

[GKAV98] V.K. Goyal, J. Kovacevic, R. Arean, and M. Vetterli. Multiple description transform coding of images. In *IEEE Int. Conf. on Image Processing*, volume I, pages 674–678, Chicago, IL, October 1998. 58, 145

[GKK02] V. K. Goyal, J. A. Kelner, and J. Kovacevic. Multiple Description Vector Quantization with a coarse lattice. *IEEE Transactions on Information Theory*, 48(3):781–788, 2002. 48, 190

[GKV99] V.K. Goyal, J. Kovacevic, and M. Vetterli. Quantized frame expansions as source-channel codes for erasure channels. In *Proc. IEEE Data Compression Conference*, pages 326–335, Snowbird, UT, USA, March 1999. 13, 50, 51, 52, 125, 145

[GL99] Alexei Gorokhov and Philippe Loubaton. Blind identification of MIMO-FIR systems: a generalized linear prediction approach. *Signal Process.*, 73(1-2):105–124, 1999. 82, 83

[GN98] R.M. Gray and D.L. Neuhoff. Quantization. *IEEE Transactions on Information Theory*, 44(6):2325–2383, October 1998. 46

[Goy98] V.K. Goyal. *Beyond traditional transform coding*. PhD thesis, Univ. California, Berkeley, 1998. 13, 50

[Goy00] V. K. Goyal. Transform coding with integer-to-integer transforms. *IEEE Transactions on Information Theory*, 46(2):465–473, March 2000. 54

[Goy01] V. K. Goyal. Multiple Description Coding: Compression meets the network. *IEEE Signal Processing Magazine*, 18(5):74–93, September 2001. 44, 190

[GP68] H. Gish and J.N. Pierce. Asymptotically efficient quantizing. *IEEE Transactions on Information Theory*, IT-14(5):676–683, 1968. 173

[GSK01] M. Gallant, S. Shirani, and F. Kossentini. Standard-compliant multiple description video coding. In *IEEE Int. Conf. on Image Processing*, Thessaloniki, Greece, October 2001. 62

[GVT98] V.K. Goyal, M. Vetterli, and N.T. Thao. Quantized overcomplete expansions in $\mathbb{R}^N$: analysis, synthesis and algorithms. *IEEE Transactions on Information Theory*, 44(1):16–31, January 1998. 51, 52

[HAL06] B. A. Heng, J.G. Apostolopoulos, and J.S. Lim. End-to-end rate-distortion optimized MD mode selection for Multiple Description video coding. *EURASIP Journal on Applied Signal Processing*, 2006:Article ID 32592, 12 pages, 2006. 63

[HP96] F. J. Hampson and J.-C. Pesquet. A nonlinear subband decomposition with perfect reconstruction. In *IEEE Int. Conf. on Acoustics, Speech and Signal Proc.*, pages 1523–1526, Atlanta, Georgia, May 7-10 1996. 18, 95

[HP98] F. J. Hampson and J.-C. Pesquet. M-Band nonlinear subband decompositions with perfect reconstruction. *IEEE Transactions on Image Processing*, 7(11):1547–1560, November 1998. 18, 70, 95

[HW00] Shih-Ta Hsiang and J.W. Woods. Embedded image coding using zeroblocks of subband/wavelet coefficients and context modeling. In *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS 2000)*, volume 3, pages 662–665, Geneva, May 2000. 56, 102

[HW01] S.-T. Hsiang and J. W. Woods. Embedded video coding using invertible motion compensated 3-d subband/wavelet filter bank. *Signal Processing: Image Communication*, 16(8):705–724, May 2001. 102

[Jay81] N.S. Jayant. Subsampling of a DPCM speech channel to provide two "self-contained" half-rate channels. *Bell Syst. Tech. Journal*, 60(4):501–509, 1981. 55

[JO99] W. Jiang and A. Ortega. Multiple description coding via polyphase transform and selective quantization. In *Proc. SPIE : Visual Communications and Image Processing*, San Jose, CA, USA, January 1999. 55, 58, 59, 190

[JZY+04] G. Jiang, W. Zhang, M. Yu, T.-Y. Choi, and Y.-D. Kim. New multiple description scalable video coding based on redundant wavelet. In *IEEE TENCON*, Bangkok, Thailand, November 2004. 63

[KA05] M. Kang and M.-S. Alouini. Transmission of Multiple Description codes over wireless channels using channel balancing. *IEEE Transactions on Wireless Communications*, 4(5):2070–2075, 2005. 57

[Kai80] T. Kailath. *Linear systems*. Prentice-Hall, Clifs, NJ, 1980. 83

[KC07a] J. Kovacevic and A. Chebira. Life beyond bases: the advent of frames (part i). *IEEE Signal Processing Magazine*, 24(4):86–104, July 2007. 50

[KC07b] J. Kovacevic and A. Chebira. Life beyond bases: the advent of frames (part ii). *IEEE Signal Processing Magazine*, 24(5):115–125, September 2007. 50

[KDG02] J. Kovacevic, P. L. Dragotti, and V. K Goyal. Filter bank frame expansions with erasures. *IEEE Transactions on Information Theory*, 48(6):1439–1450, June 2002. 50, 70

[KFR04] M. Koca, E. Fabre, and A. Roumy. Turbo decoding of multiple descriptions with quantized frame expansions. In *Proc. of International Symposium on Images and Video Coding*, 2004. 57

[KGK00] J. Kelner, V.K. Goyal, and J. Kovacevic. Multiple description lattice vector quantization: Variations and extensions. In *Proc. of Data Compression Conference*, pages 480–489, Snowbird, Utah, USA, March 2000. 47

[KHWK08] H.A. Karim, C. Hewage, S. Worrall, and A. Kondoz. Scalable multiple description video coding for stereoscopic 3d. *Consumer Electronics, IEEE Transactions on*, 54(2):745–752, May 2008. 34, 188

[KL01] C-S. Kim and S-U. Lee. Multiple Description coding of motion fields for robust video transmission. *IEEE Transactions on Circuits and Systems for Video Technology*, 11(9):999–1010, September 2001. 61

[KMA05] J. Kim, R. M. Mersereau, and Y. Altunbasak. Distributed video streaming using multiple description coding and unequal error protection. *IEEE Transactions on Image Processing*, 14(7):849 – 861, 2005. 63

[Kon05] L. P. Kondi. A rate-distortion optimal hybrid scalable/Multiple-Description video codec. *IEEE Transactions on Circuits and Systems for Video Technology*, 15(7):921–927, July 2005. 64

[KZC+94] M. Khansari, A. Zakauddin, Wai-Yip Chan, E. Dubois, and P. Mermelstein. Approaches to layered coding for dual-rate wireless video transmission. In *Image Processing, 1994. Proceedings. ICIP-94., IEEE International Conference*, volume 1, pages 258–262 vol.1, Nov 1994. 41

[LAM05] Y.-C. Lee, Y. Altunbasak, and R. M. Mersereau. Coordinated application of Multiple Description Scalar Quantization and error concealment for error-resilient MPEG video streaming. *IEEE Transactions on Circuits and Systems for Video Technology*, 15(4):457–468, April 2005. 64

[Lar95] J. Laroche. *Traitement des signaux audio-fréquences*. ENST, Février 1995. Available at: `http://www.tsi.enst.fr/ prado/enseignement/enseignement.html`. 52

[LMWA05] J. N. Laneman, E. Martinian, G. W. Wornell, and J. G. Apostolopoulos. Source-channel diversity for parallel channels. *IEEE Transactions on Information Theory*, 51(10):3518–3539, October 2005. 57

[LPFA00] W. S. Lee, M. R. Pickering, M. R. Frater, and J. F. Arnold. A robust codec for transmission of very low bit-rate video over channels with bursty errors. *IEEE Transactions on Circuits and Systems*, 10(8):1403–1412, December 2000. 61

[Mal89] S. G. Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11:674–693, 1989. 14, 52, 67

[Mal98] S. Mallat. *A wavelet tour of signal processing*. Academic Press, second edition, 1998. 14, 54, 60, 67, 68, 70, 76

[Mal09] Stephane Mallat. *A Wavelet Tour of Signal Processing, The Sparse Way*. Academic Press (Elsevier), 3rd edition edition, 2009. 34, 187

[MC99] S. Mehrotra and P. A. Chou. On optimal frame expansions for multiple description quantization. In *Proc. of the Int. Symposium on Information Theory*, 1999. 70

[Mey90] Y. Meyer. *Ondelettes et Opérateurs*, volume 1. 1990. 14, 67

[MG04a] R. Motwani and C. Guillemot. A para-pseudo inverse based method for reconstruction of filter bank frame-expanded signals from erasures. In *IEEE Int. Conf. on Image Processing*, Singapore, October 2004. 60

[MG04b] R. Motwani and C. Guillemot. Tree-structured oversampled filterbanks as joint source-channel codes: application to image transmission over erasure channels. *IEEE Transactions on Signal Processing*, 52(9):2584 – 2599, September 2004. 77, 126

[MLP+03] S. Mao, S. Lin, S.S. Panwar, Y. Wang, and E. Celebi. Video transport over Ad Hoc networks: Multistream coding with multipath transport. *IEEE Journal on Selected Areas in Communications*, 21(10):1721–1737, 2003. 64

[MMR99] A. Miguel, A. Mohr, and E. Riskin. SPIHT for generalized multiple description coding. In *IEEE Int. Conf. on Image Processing*, Kobe, Japan, October 1999. 59

[Mor00] N. Moreau. *Compression pour le multimédia, Outils théoriques*. ENST, Décembre 2000. 54

[MRL99] A. Mohr, E. Riskin, and R. Ladner. Generalized multiple description coding through unequal forward error correction. In *IEEE Int. Conf. on Image Processing*, Kobe, Japan, October 1999. 56

[MS54] T. S. Motzkin and I. J. Schoenberg. The relaxation method for linear inequalities. *Canadian Journal of Math*, 6(3):393–404, 1954. 142

[MS05] A. Mavlankar and E. Steinbach. Multiple Description video coding using motion-compensated lifted 3D wavelet decomposition. In *IEEE Int. Conf. on Acoustics, Speech and Signal Proc.*, Philadelphia, PA, USA, March 2005. 63

[NAB+06] A. Norkin, A. Aksay, C. Bilen, G.B. Akar, A. Gotchev, and J. Astola. Schemes for multiple description coding of stereoscopic video. In Lecture Notes in Computer Science, editor, *Proceedings of International Workshop on Multimedia Content Representation, Classification and Security*, volume 4105/2006, pages 730–737, September 2006. 34, 188

[OWVR97] M.T. Orchard, Y. Wang, V. Vaishampayan, and A.R. Reibman. Redundancy rate-distortion analysis of multiple description coding using pairwise correlating transforms. In *IEEE Int. Conf. on Image Processing*, volume 1, pages 608–611, Washington, DC, USA, October 1997. 54

[Oza80] L. Ozarow. On a source-coding problem with two channels and three receivers. *Bell Syst. Tech. J.*, 59(10):1909–1921, December 1980. 42

[PAB02a] M. Pereira, M. Antonini, and M. Barlaud. Channel adapted scan-based Multiple Description video coding. *Proc. of the IEEE Int. Conf. on Multimedia and Expo*, 2:609–612, August 2002. 60

[PAB02b] M. U. Pereira, M. Antonini, and M. Barlaud. Channel adapted multiple description coding scheme using wavelet transform. In *IEEE Int. Conf. on Image Processing*, volume 2, pages 197–200, Rochester, New York, USA, 2002. 60

[PAB03a] M. Pereira, M. Antonini, and M. Barlaud. Multiple Description coding for internet video streaming. *IEEE Int. Conf. on Image Processing*, 3:281–284, September 2003. 60

[PAB03b] M. Pereira, M. Antonini, and M. Barlaud. Multiple description image and video coding for wireless channels. *EURASIP: Image Communication Special Issue on Recent Advances in Wireless Video*, 18(10):925–945, November 2003. 60

[Pen02] Erwan Le Pennec. *Bandelettes et représentations géometriques des images*. PhD thesis, Ecole Polytechnique, 2002. 34, 187

[Per04] M. Pereira. *Multiple Description Image and Video coding for noisy channels*. PhD thesis, I3S Laboratory of CNRS, France, 2004. 60

[PFPPP07] T. Petrisor, A. Fraysse, B. Pesquet-Popescu, and J.-C. Pesquet. Une nouvelle approche du codage par descriptions multiples, utilisant des representations linéaires creuses. In *Proc. of GRETSI, Traitement du Signal et des Images*, Troyes, France, September 2007. 169

[PH02] G. Piella and H. J. A. M. Heijmans. Adaptive lifting schemes with perfect reconstruction. *IEEE Transactions on Signal Processing*, 7:1620–1630, 2002. 34, 187

[PKC96] J.-C. Pesquet, H. Krim, and H. Carfantan. Time invariant orthonormal wavelet representations. *IEEE Transactions on Signal Processing*, 44(8):1964–1970, August 1996. 125

[PLRB01] R. Puri, S-U. Lee, K. Ramchandran, and V. Bharghavan. An integrated source transcoding and congestion control paradigm for video streaming in the Internet. *IEEE Transactions on Multimedia*, 3(1):18–32, March 2001. 61

[PPB01] B. Pesquet-Popescu and V. Bottreau. Three-dimensional lifting schemes for motion compensated video compression. In *IEEE Int. Conf. on Acoustics, Speech and Signal Proc.*, Salt Lake City, UT, May 2001. 63, 102, 110

[PPPP05a] T. Petrisor, B. Pesquet-Popescu, and J.-C. Pesquet. Perfect reconstruction in reduced redundancy wavelet-based multiple description coding of images. In *Proc. of the European Signal Processing Conference (EUSIPCO)*, Antalya, Turkey, September 2005. 123

[PPPP05b] T. Petrisor, B. Pesquet-Popescu, and J.-C. Pesquet. Wavelet-based multiple description coding of images with iterative convex optimization techniques. In *IEEE Int. Conf. on Image Processing*, volume III, pages 924–927, Genoa, Italy, September 2005. 28, 123

[PPPP07] T. Petrisor, B. Pesquet-Popescu, and J.-C. Pesquet. A compressed sensing approach to frame-based multiple description coding. In *IEEE Int. Conf. on Acoustics, Speech and Signal Proc.*, Honolulu, USA, April 2007. 169

[PR99] R. Puri and K. Ramchandran. Multiple description source coding through forward error correction codes. In *Proc. Asilomar Conf. on Signals, Systems and Computers*, Asilomar, CA, October 1999. 56

[PR00] S. S. Pradhan and K. Ramchandran. On the optimality of block orthogonal transforms for multiple description coding ofGaussian vector sources. *IEEE Signal Processing Letters*, 7(4):76–78, April 2000. 54

[Pro01] John G. Proakis. *Digital Communications*. McGraw-Hill, 2001. 45

[PTPPH04] G. Pau, C. Tillier, B. Pesquet-Popescu, and H. Heijmans. Motion compensation and scalability in lifting-based video coding. *Signal Processing: Image Communication, special issue on Wavelet Video Coding Elsevier/EURASIP*, 19:577–600, 2004. 102

[PTPPP04] T. Petrisor, C. Tillier, B. Pesquet-Popescu, and J.-C. Pesquet. Redundant multiresolution analysis for multiple description video coding. In *IEEE International Workshop on Multimedia Signal Processing*, Siena, Italy, 2004. 67

[PTPPP05] T. Petrisor, C. Tillier, B. Pesquet-Popescu, and J.-C. Pesquet. Comparison of redundant wavelet schemes for multiple description coding of video sequences. In *IEEE Int. Conf. on Acoustics, Speech and Signal Proc.*, volume 5, pages 913– 916, Philadelphia, PA, USA, March 2005. 67, 70

[Pur02] R. Puri. *Robust multimedia coding: Information Theory and practical architectures*. PhD thesis, 2002. 61

[RC98] J. K. Rogers and P. C. Cosman. Robust wavelet zerotree image compression with fixed-length packetization. In *Proc. of Data Compression Conference*, 1998. 59

[RF06] I. Radulovic and P. Frossard. Multiple description image coding with block-coherent redundant dictionaries. In *Picture Coding Symposium, China, 6pg.*, 2006. 169

[RF07] I. Radulovic and P. Frossard. Multiple description coding with redundant expansions and application to image communications. *EURASIP Journal on Image and Video Processing*, 2007:Article ID 24863, 15 pages, 2007. 60

[RJW+99] A.R. Reibman, H. Jafarkhani, Y. Wang, M.T. Orchard, and R. Puri. Multiple description coding for video using motion compensated prediction. In *Proc. IEEE Int. Conf. Image Processing*, Kobe, Japan, October 1999. 61, 62

[RJW+02] A. R. Reibman, H. Jafarkhani, Y. Wang, M. T. Orchard, and R. Puri. Multiple-Description video coding using Motion-Compensated temporal prediction. *IEEE Transactions on Circuits and Systems*, 12(3):193–204, March 2002. 62, 63

[Roc70] R.T. Rockafellar. *Convex Analysis*. Princeton University Press, 1970. 123, 139, 141

[RRP05] G. Romano, P. S. Rossi, and F. Palmieri. Optimal correlating transform for erasure channels. *IEEE Signal Processing Letters*, 12(10):677–680, October 2005. 49

[RSO06] T. Russert, M. Spiertz, and J.-R. Ohm. H.264/AVC compatible scalable Multiple Description video coding with RD optimization. In *Proc. of the IEEE ISPACS*, Tottori, Japan, December 2006. 64

[RWG00] J. Ridge, F. W. Ware, and J.D. Gibson. Permuted smoothed descriptions and refinement coding for images. *IEEE Journal on Selected Areas in Communications*, 18(6):915–926, June 2000. 60

[RWW+07] Ivana Radulovic, Ye-Kui Wang, Stephan Wenger, Antti Hallapuro, Miska M. Hannuksela, and Pascal Frossard. Multiple description h.264 video coding with redundant pictures. In *ACM Mobile Video*, pages 37–42, Ausburg, Germany, September 2007. ACM. 64

[Sha48] C.E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423 and 623–656, July and October 1948. 42

[Sha93] J.M. Shapiro. Embedded image coding using zerotrees of wavelet coefficients. *IEEE Transactions on Signal Processing*, 41:3445–3462, December 1993. 59

[Shi06] S. Shirani. Content-based Multiple Description image coding. *IEEE Transactions on Multimedia*, 8(2):411–419, 2006. 60

[SKK05] Y. Sriraja, S. Krishnamurthy, and T. Karp. Multiple description wavelet image coding for reliable delivery and QoS maintenance. In *Int. Conf. on Multimedia Services Access Networks (MSAN)*, Orlando, FL, USA, June 2005. 59, 60, 190

[SM02] B. M. R Shankar and A. Makur. Allpass delay chain-based IIR PR filterbank and its application to Multiple Description subband coding. *IEEE Transactions on Signal Processing*, 50(4):814–823, 2002. 55

[SP96] A. Said and W. A. Pearlman. A new, fast, and efficient image codec based on set partitioning in hierarchical trees. *IEEE Transactions on Circuits and Systems for Video Technology*, 6(3):243–250, June 1996. 56, 59

[SRVN00] S. Servetto, K. Ramchandran, V. Vaishampayan, and K. Nahrstedt. Multiple description wavelet based image coding. *IEEE Transactions on Image Processing*, 9(5):813–826, 2000. 58, 60

[SS04] K. F. Sadri and S. Shirani. Multiple Description coding of images using phase scrambling. In *IEEE Int. Conf. on Acoustics, Speech and Signal Proc.*, 2004. 60

[SW01] J. Stockler and G. V. Welland, editors. *Beyond wavelets.* Academic Press (Elsevier), 2001. 34, 187

[Tau98] David S. Taubman. Embedded block coding with optimized truncation. ISO/IEC JTC1/SC29/WG1 N1020R, October 1998. 56

[TCPPG07] C. Tillier, O. Crave, B. Pesquet-Popescu, and C. Guillemot. A comparison of four video multiple description coding schemes. *In proceedings of EUSIPCO 2007, 15th European Signal Processing Conference, Poznan, Poland*, September, 2007. 61

[TGO07] T. Tillo, M. Grangetto, and G. Olmo. Multiple Description image coding based on Lagrangian rate allocation. *IEEE Transactions on Image Processing*, 16(3):673–683, 2007. 60

[TO04] T. Tillo and G. Olmo. A novel Multiple Description coding scheme compatible with the JPEG 2000 decoder. *IEEE Signal Processing Letters*, 11(11):908–911, November 2004. 60

[TPL04] J. R. Taal, J. A. Pouwelse, and R. L. Lagendijk. Scalable Multiple Description coding for video distribution in P2P networks. In *Picture Coding Symposium*, San Francisco, CA, USA, December 2004. 64

[TPP03] C. Tillier and B. Pesquet-Popescu. 3D, 3-band, 3-tap temporal lifting for scalable video coding. In *Proceedings of the IEEE International Conference on Image Processing*, Barcelona, Spain, Sept. 2003. 102

[TPPPP05] C. Tillier, T. Petrisor, B. Pesquet-Popescu, and J.-C. Pesquet. Codage par descriptions multiples pour la transmission vidéo. In *Proc. of GRETSI, Traitement du Signal et des Images*, Louvain-la-Neuve, Belgium, September 2005. 67

[TPPPP07] C. Tillier, T. Petrisor, B. Pesquet-Popescu, and J.-C. Pesquet. A Motion-Compensated Overcomplete Temporal Decomposition for Multiple Description Scalable Video Coding. *EURASIP Journal on Image and Video Processing*, (2007):12 page, 2007. 67

[TPPvdS05] C. Tillier, B. Pesquet-Popescu, and M. van der Schaar. Improved update operators for lifting-based motion-compensated temporal filtering. *IEEE Signal Processing Letters*, 12(2):146–149, 2005. 103

[TV94] N.T. Thao and M. Vetterli. Deterministic analysis of oversampled A/D conversion and decoding improvement based on consistent estimates. *IEEE Transactions on Signal Processing*, 42(3):519–531, March 1994. 52

[TZ02] X. Tang and A. Zakhor. Matching Pursuits Multiple Description Coding for wireless video. *IEEE Transactions on Circuits and Systems*, 12(6):566–575, June 2002. 61, 62

[Vai93a] P. Vaidyanathan. *Multirate systems and filter banks*. Prentice-Hall, Englewood Cliffs, NJ, 1993. 52

[Vai93b] V. Vaishampayan. Design of multiple description scalar quantizers. *IEEE Transactions on Information Theory*, 39(3):821–834, May 1993. 13, 33, 44, 45, 46, 47, 187

[VB98] V. Vaishampayan and J.C. Batllo. Asymptotic analysis of multiple description quantizers. *IEEE Transactions on Information Theory*, 44(1):278–284, January 1998. 46, 47

[VD94] V. Vaishampayan and J. Domaszewicz. Design of entropy-constrainted multiple description scalar quantizers. *IEEE Transactions on Information Theory*, 40(1):245–250, January 1994. 46, 47

[vdBLV96] C. van den Branden Lambrecht and O. Verscheure. Perceptual Quality Measure using a Spatio-Temporal Model of the Human Visual System. In *IS&T/SPIE*, San Jose, CA, USA, 1996. 34, 188

[VDMdW07] K. Vermeirsch, Y. Dhondt, S. Mys, and R. Van de Walle. Low complexity multiple description coding for h.264/avc. In *In Proc. of the Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS)*, pages 61–61, Santorini, Greece, June 2007. 64

[vdST03] M. van der Schaar and D. S. Turaga. Multiple description scalable coding using wavelet-based motion compensated temporal filtering. In *IEEE Int. Conf. on Image Processing*, Barcelona, Spain, September 2003. 64

[Vet86] M. Vetterli. Filter banks allowing perfect reconstruction. *Signal Processing*, 10(3):219–244, April 1986. 52

[VK95] M. Vetterli and J. Kovacevic. *Wavelets and subband coding*. Prentice Hall, 1995. 52, 53, 76, 90

[VKG03] R. Venkataramani, G. Kramer, and V. K. Goyal. Multiple Description coding with many channels. *IEEE Transactions on Information Theory*, 49(9):2106–2114, September 2003. 44

[VSS01] V. Vaishampayan, N. Sloane, and S. Servetto. Multiple description vector quantization with lattice codebooks: design and analysis. *IEEE Transactions on Information Theory*, 47(5):1718–1734, July 2001. 47

[WOR97] Y. Wang, M.T. Orchard, and A.R. Reibman. Multiple description image coding for noisy channels by pairing transform coefficients. *Proc. IEEE Workshop on Multimedia Signal Processing*, pages 419–424, June 23-25 1997. Princeton, New Jersey, USA. 13, 44, 48, 58

[WOR98] Y. Wang, M.T. Orchard, and A. Reibman. Optimal pairwise correlating transforms for multiple description coding. In *IEEE Int. Conf. on Image Processing*, volume 1, pages 679–683, Chicaco, IL, USA, October 1998. 48

[WRL05] Y. Wang, A. R. Reibman, and S. Lin. Multiple Description coding for video delivery. *Proceedings of the IEEE*, 93(1):57–70, 2005. 61

[WRTJ02] Y. Wang, A. R. Reibman, Orchard M. T., and H. Jafarkhani. An improvement to Multiple Description Transform Coding. *IEEE Transactions on Signal Processing*, 50(11):2843–2854, 2002. 57

[WV07] H. Wang and P. Viswanath. Vector Gaussian Multiple Description with individual an central receivers. *IEEE Transactions on Information Theory*, 53(6):2133–2153, June 2007. 44

[YR98] X. Yang and K. Ramchandran. Optimal multiple description subband coding. In *IEEE Int. Conf. on Image Processing*, Chicago, IL, USA, October 1998. 53, 54

[ZB95] Z. Zhang and T. Berger. Multiple description source coding with no excess marginal rate. *IEEE Transactions on Information Theory*, 41(2):349–357, March 1995. 43