HAL
open science

Reconnaissance d'objets 3D par points d'intérêt

Ayet Shaiek

To cite this version:

Ayet Shaiek. Reconnaissance d'objets 3D par points d'intérêt. Autre [cs.OH]. Ecole Nationale Supérieure des Mines de Paris, 2013. Français. NNT : 2013ENMP0011 . pastel-00871080

HAL Id: pastel-00871080

https://pastel.hal.science/pastel-00871080v1

Submitted on 8 Oct 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY

MINES ParisTech

École doctorale n° 432 : Sciences des Métiers de l'Ingénieur

Doctorat ParisTech

T H È S E

pour obtenir le grade de docteur délivré par

l'École nationale supérieure des mines de Paris

Spécialité " Informatique temps réel – Robotique - Automatique "

présentée et soutenue publiquement par

Ayet SHAIEK

21 /03/2013

Reconnaissance d'objets 3D par points d'intérêt

Directeur de thèse : **Arnaud DE LA FORTELLE**
Maître de thèse : **Fabien MOUTARDE**

Jury

M. Mohamed DAOUDI	LIFL, Telecom Lille1	Rapporteur
Mme. Anne VERROUST-BLONDET	IMEDIA, Inria Rocquencourt	Rapporteur
M. Jean-Luc Dugelay	EURECOM	Examinateur
M. Fabien MOUTARDE	Centre de robotique, Mines ParisTech	Maître de thèse
M. Arnaud DE LA FORTELLE	Centre de robotique, Mines ParisTech	Directeur de thèse

T H È S E

MINES ParisTech
Centre de robotique
60 Boulevard Saint Michel 75006 Paris

A mes parents

Remerciements

Je remercie l'ensemble des membres du jury pour l'intérêt qu'ils ont manifesté à évaluer et juger mon travail de recherche. Je remercie mes rapporteurs Madame Anne VERROUST-BLONDET et Monsieur Mohamed DAOUDI pour leurs questions et remarques judicieuses. Je remercie également Monsieur Jean-Luc DUGELAY pour avoir accepté la présidence de mon jury de thèse.

Je remercie Monsieur Arnaud DE LA FORTELLE de m'avoir accueilli au sein du CAOR et d'avoir accepté de diriger mes travaux de thèse.

Mes plus sincères remerciements vont à mon maître de thèse, Monsieur Fabien MOUTARDE qui m'a aidé, conseillé et soutenu tout au long de ces années.

Je remercie également l'ensemble des doctorants du laboratoire, en particulier Raoul DE CHARETTE, Jean Emmanuel DESCHAUD et Taha RIDEN pour leur soutien moral et sincère amitié. Je remercie aussi les permanents du laboratoire, trop nombreux pour être cités nommément, grâce à qui j'ai travaillé dans un environnement scientifique de qualité et une ambiance amicale.

Enfin, merci à l'ensemble de mes proches en Tunisie et en France pour le soutien qu'ils m'ont apporté durant cette thèse.

Résumé

Soutenue par les progrès récents et rapides des techniques d'acquisition 3D, la reconnaissance d'objets 3D a suscité de nombreux efforts de recherche durant ces dernières années. Cependant, il reste à résoudre dans ce domaine plusieurs problématiques liées à la grande quantité d'information, à l'invariance à l'échelle et à l'angle de vue, aux occultations et à la robustesse au bruit.

Dans ce contexte, notre objectif est de reconnaitre un objet 3D isolé donné dans une vue requête, à partir d'une base d'apprentissage contenant quelques vues de cet objet. Notre idée est de formuler une méthodologie locale qui combine des aspects d'approches existantes et apporte une amélioration sur la performance de la reconnaissance.

Nous avons opté pour une méthode par points d'intérêt (PIs) fondée sur des mesures de la variation locale de la forme. Notre sélection de points saillants est basée sur la combinaison de deux espaces de classification de surfaces : l'espace SC (indice de forme- intensité de courbure), et l'espace HK (courbure moyenne-courbure gaussienne).

Dans la phase de description de l'ensemble des points extraits, nous proposons une signature d'histogrammes, qui joint une information sur la relation entre la normale du point référence et les normales des points voisins, avec une information sur les valeurs de l'indice de forme de ce voisinage. Les expérimentations menées ont permis d'évaluer quantitativement la stabilité et la robustesse de ces nouveaux détecteurs et descripteurs.

Finalement nous évaluons, sur plusieurs bases publiques d'objets 3D, le taux de reconnaissance atteint par notre méthode, qui montre des performances supérieures aux techniques existantes.

Mots clés: Reconnaissance 3D, Points d'intérêt, Images de Profondeurs, Nuages de Points 3D, Détecteurs, Descripteurs, Objet, Courbures, Normales.

Abstract

There has been strong research interest in 3D object recognition over the last decade, due to the promising reliability of the 3D acquisition techniques. 3D recognition, however, conveys several issues related to the amount of information, to scales and viewpoints variation, to occultations and to noise.

In this context, our objective is to recognize an isolated object given in a request view, from a training database containing some views of this object. Our idea is to propose a local method that combines some existent approaches in order to improve recognition performance.

We opted for an interest points (IPs) method based on local shape variation measures. Our selection of salient points is done by the combination of two surface classification spaces: the SC space (Shape Index-Curvedness), and the HK space (Mean curvature- Gaussian curvature).

In description phase of the extracted set of points, we propose a histogram based signature, in which we join information about the relationship between the reference point normal and normals of its neighbors, with information about the shape index values of this neighborhood. Performed experiments allowed us to evaluate quantitatively the stability and the robustness of the new proposed detectors and descriptors.

Finally we evaluate, on several public 3D objects databases, the recognition rate attained by our method, which outperforms existing techniques on same databases.

Keywords: 3D Recognition, Interest Points, Range Images, 3D Point Clouds, Detectors, Descriptors, Object, Curvatures, Normals.

Table des matières

REMERCIEMENT **3**

RESUME **5**

ABSTRACT **7**

TABLE DES MATIERES **9**

LISTE DES FIGURES **12**

LISTE DES TABLEAUX **19**

ACRONYMES **21**

INTRODUCTION **23**

CHAPITRE 1 : TERMINOLOGIE ET NOTIONS DE BASE SUR LA 3D **27**

1.1. TYPES DE DONNEES 3D 27
1.1.1 Nuage de points *27*
1.1.2 Image de profondeurs *27*
1.1.3 Données polygonales *29*
1.1.4 Prétraitements des données *29*
1.2. MESURES QUANTITATIVES DU DEGRE DE COURBURE 31
1.2.1 Normale et Courbure *31*
1.2.2 Courbures principales *34*
1.2.3 Courbure gaussienne (K) et moyenne (H) *35*
1.2.4 Espace HK *35*
1.2.5 Indice de forme (S) *37*
1.2.6 Intensité de courbure (C) *38*
1.2.7 Espace SC *38*
1.2.8 Comparaison HK vs SC *39*

CHAPITRE 2 : ETAT DE L'ART DES TECHNIQUES DE RECONNAISSANCE 3D **41**

2.1. TYPOLOGIE DES APPROCHES DE RECONNAISSANCE 3D 41
2.1.1 Approches 2D/3D *42*
2.1.1.1 Approches 2D/3D globales 42
2.1.1.2 Approches 2D/3D locales 43
2.1.2 Approches 3D *46*
2.1.2.1 Approches 3D globales 46
2.1.2.2 Approches 3D locales 48
2.1.3 Conclusions et remarques *52*
2.2. DETECTION ET DESCRIPTION DES POINTS D'INTERET EN 2D 55
2.2.1 Détection *56*
2.2.2 Descripteurs locaux *59*
2.3. DETECTION ET DESCRIPTION DES POINTS D'INTERET (PI) EN 3D 60
2.3.1 Détecteurs *60*
2.3.1.1 Détecteurs à échelle fixe 60

2.3.1.2 Détecteurs invariants à l'échelle 64
2.3.1.3 Mesures de la performance de détection des Points d'Intérêt 3D 71
2.3.1.3.1 Facteur de qualité des PIs 72
2.3.1.3.2 Répétabilité 73
2.3.2 Descripteurs de points d'intérêt 3D 73
2.4. MESURES DE LA PERFORMANCE 77
2.5. CONCLUSIONS 77

CHAPITRE 3: DETECTION 81

3.1. PRINCIPES DES DETECTEURS TESTES 81
3.1.1. Seuillage des valeurs de SI 81
3.1.2 Sélection des Points Harris 3D 82
3.1.3 SURF 3D 83
3.1.4 Facteur de qualité 84
3.1.5 La classification HK et SC 85
3.1.6 Combinaison des critères 87
3.2. ELEMENTS IMPACTANT SUR LES MESURES DE COURBURE 91
3.2.1 Artefacts et bruit 91
3.2.2 Invariance aux transformations, échelle et résolution 93
3.2.3 Maillage et voisinage 93
3.3. RESULTATS EXPERIMENTAUX 94
3.3.1 Outils de développements et bases de données 94
3.3.2 Evaluation des détecteurs 100
3.3.2.1 Evaluation visuelle 100
3.3.2.2 Stabilité aux changements d'angles 107
3.3.2.3 Stabilité à l'échelle 109
3.3.2.4 Stabilité au bruit 113
3.3.2.5 Temps de calcul 115
3.4. SYNTHESE 117

CHAPITRE 4 : DESCRIPTEURS 119

4.1. PRESENTATION DES DESCRIPTEURS TESTES 119
4.1.1 Descripteur Spin Image 119
4.1.2 Descripteur Thrift 122
4.1.3 Local surface patches (LSP) 123
4.1.4 IndThrift 124
4.1.5 Descripteur SHOT 125
4.1.6 CSHOT 127
4.1.7 IndSHOT 127
4.2. EVALUATION EXPERIMENTALE 129
4.2.1 Paramètres 129
4.2.1.1 Choix du nombre de Bins 130
4.2.1.2 Choix du rayon du support 132
4.2.1.3 Choix de l'angle du support 134
4.2.2 Comparaison des descripteurs 135
4.2.2.1 Invariance aux rotations 136
4.2.2.2 Invariance à l'échelle et à la densité de points 139
4.2.2.3 Robustesse au bruit 142
4.2.2.4 Robustesse à un lissage 143
4.2.2.5 Pouvoir descriptif et stabilité pour un décalage de la position physique 144
4.2.2.6 Temps de calcul et outils de développement 147
4.3. CONCLUSIONS 149

CHAPITRE 5 : RECONNAISSANCE 151

5.1. TECHNIQUES D'APPRENTISSAGE STATISTIQUE SUPERVISE POUR LA CLASSIFICATION.....151
5.1.1 Classificateur par le plus proche voisin.....151
5.1.1.1 Métriques d'appariement.....152
5.1.1.2 Kd-Tree.....153
5.1.1.3 Hachage géométrique.....155
5.1.2 Séparateur à vaste marge.....157
5.1.3 Adaboost (Adaptive Boosting).....157
5.1.4 Réseau de neurones.....158
5.1.5 Algorithme de vote des PIs.....158
5.1.6 Conclusion.....159
5.2. EVALUATIONS DES METHODES.....159
5.2.1 Paramètres.....159
5.2.2 Evaluation de l'invariance et de la robustesse sur la base Minolta.....161
5.2.2.1 Invariance à la rotation.....162
5.2.2.2 Variation d'échelles.....166
5.2.2.3 Robustesse au bruit.....170
5.2.3 Base Stuttgart.....175
5.2.4 Base Carotte.....181
5.2.5 Base RGBD.....183
5.2.6 Conclusions.....184

CONCLUSIONS ET PERSPECTIVES.....187

LISTE DES PUBLICATIONS.....189

BIBLIOGRAPHIE.....191

Liste des Figures

Figure 1-1. Nuage de point d'un modèle 3D de voiture 27
Figure 1-2. L'image RGB d'une scène réelle (à gauche) et l'image recueillie par la caméra 3D Kinect™ (à droite).....28
Figure 1-3. -a- Six images de profondeur associées aux six faces de la boîte de projection (Zaharia, et al., 2002) ; -b- les 20 angles de vues associés aux sommets du dodécaèdre (Chaouch, et al., 2007).....28
Figure 1-4. Exemple de simplification avec différentes valeurs de subdivision régulière.....29
Figure 1-5. Trois représentations du visage: image 2D en niveau de gris, image de profondeur, et modèle surfacique 30
Figure 1-6. a- Point du bord (définition et conditions); b-Variation de la normale pour l'estimation de la courbure (Kalogerakis, et al., 2009).....31
Figure 1-7. Mesure du caractère « plus ou moins courbé » de l'objet : courbure (à gauche), et normales (à droite). 31
Figure 1-8. a- Le plan de la normale avec la courbure maximale. -b- Le plan de la normale avec la courbure minimale. -c- Exemple de valeurs de courbures principales 34
Figure 1-9. Classification des surfaces selon les valeurs de H et K (Akagündüz, 2011).....35
Figure 1-10. Classification HK dans le plan (k_1, k_2) (Akagündüz, 2011).....37
Figure 1-11. Positions des 9 types de surfaces par la classification sur l'échelle des valeurs de l'indice de forme SI (Dorai & Jain, 1997) 38
Figure 2-1. Etapes d'un système de reconnaissance 41
Figure 2-2. Représentation des courbes faciales. En haut : six expressions faciales de la même personne. En bas: la même expression faciale de six personnes différentes (Samir, et al., 2006).....43
Figure 2-3. Exemples de filtres rectangulaires (Jones, et al., 2003) 45
Figure 2-4. Représentation des normales en coordonnées sphériques (Hetzel, et al., 2001) 45
Figure 2-5. A gauche-Exemples d'objets avec le même EGI-A droite- un objet avec la sphère de Gauss (Horn, 1984) 47
Figure 2-6. Etapes de calcul des harmoniques sphériques (Kazhdan, et al., 2002) 48
Figure 2-7. (a)- Les mesures géométriques caractérisant la relation entre deux facettes (b)- L'histogramme de paire géométrique représentant cette relation 49

Figure 2-8. Un graphe Reeb Multi-résolution utilisant la fonction de hauteur (Hilaga, et al., 2001) 50
Figure 2-9. Point v (en noir) et ses anneaux de voisinage. Le premier anneau est composé par les sommets verts. L'anneau 2 est composé par les sommets verts et bleus 50
Figure 2-10. Extraction des crêtes (en bleu) et des ravines (rouge) sur différents modèles (Song, et al., 2005) 51
Figure 2-11. Distributions des types de région en fonction des deux plus grandes valeurs propres de la matrice de Harris (Harris, et al., 1988). 56
Figure 2-12. Processus de production des images de différence de gaussienne (Lowe, 2004) 58
Figure 2-13. De gauche à droite : La gaussienne dérivée partielle discrétisée de second ordre dans la direction y et la direction xy, respectivement. L'approximation de la gaussienne dérivée partielle de second ordre dans la direction y et la direction xy, respectivement. 58
Figure 2-14. Histogramme SIFT (Lowe, 2004) 59
Figure 2-15. Descripteurs SURF calculés: sur une région homogène à gauche, en présence de fréquence sur l'axe x au milieu et sur une région d'intensité dégradée à droite (Bay, et al., 2006) 60
Figure 2-16 Typographie des surfaces selon la valeur de l'indice de forme 61
Figure 2-17. A gauche-Etendu des 14 composantes connexes détectées-A droite- les centres de ces composantes. 61
Figure 2-18. a-Modèles de points saillants sur un visage 3D- b- La carte 2D longitude-latitude 62
Figure 2-19. Points détectés sur une surface symétrique (gaussienne 2D) et sur le modèle d'un coq (Mian, et al., 2009) 63
Figure 2-20. Points Harris 3D sur des modèles synthétiques (Sipiran, et al., 2010) 64
Figure 2-21 . Exemples d'images 2,5D frontales générées par la rotation du même nuage de point 3D autour de l'axe x avec un angle de (a) -10°, (b) 0° et (c) +10° (Mayo, et al., 2009) 65
Figure 2-22. Schéma de l'application du SIFT sur une image SI 66
Figure 2-23. Illustration sur des images SI, des mises en correspondances entre les paires de PIs obtenus par la méthode SIFT 66
Figure 2-24. Les dix primitives les plus larges extraites pour trois objets; 67
Figure 2-25. Mise en correspondance des paires de points détectées pour : deux niveaux d'échelles différents (à gauche) et deux vues différentes (à droite) du même objet. 68
Figure 2-26. Algorithme multi-échelle basée sur l'intensité de courbure C (Ho, et al., 2009) 69
Figure 2-27. Schéma de la phase inter-octave et intra-octave pour extraire les PIs 70
Figure 2-28. A gauche- Pourcentage des PIs en fonction de la qualité Qk -A droite- Répétabilité de l'identification des PIs en fonction de la qualité Qk (Mian, et al., 2009) 72

Figure 2-29. a- Invariance au changement d'échelle-b- Invariance au changement d'échelle dans une scène avec des occultations (Mian, et al., 2009) 73
Figure 2-30. L'image gauche : le descripteur SIFT 2D normal. Milieu : le descripteur SIFT 2D multiple utilisé sur une vidéo sans modification de la méthode originale. Droite : le descripteur SIFT 3D avec ses sous-volumes 3D, chaque sous-volume est accumulé dans son propre sous-histogramme. Ces histogrammes forment le descripteur final. (Scovanner, et al., 2007) 74
Figure 2-31. Description des points d'intérêt : à gauche, les positions de deux PIs en noir avec leur chemin spiral. En haut à droite, un zoom sur le point v_i dont le chemin spiral est limité par le rayon géodésique r. En bas à droite, la construction d'une spirale en présence d'un trou dans la maille. 76
Figure 2-32. Courbes présentant: 1-précision-rappel- 2- ROC -3- Taux de reconnaissance 78
Figure 3-1-Gauche-Image de profondeur du modèle d'un ange - Milieu- Image de l'indice de forme en niveau de gris–Droite- Les PIs détectés. 82
Figure 3-2 Algorithme de regroupement des points d'intérêt (clustering) 83
Figure 3-3- A gauche -Sélection des points avec la plus grande réponse d'Harris et - A droite - Sélection des points Harris avec regroupement. 83
Figure 3-4 Illustration des PIs détectés par le SURF 3D. La forme (a) est voxélisée en une grille cubique (b). Les PIs détectés (c), et représentés par des sphères dont le rayon illustre l'échelle. (Knopp, et al., 2010) 84
Figure 3-5-A gauche-Image de profondeur d'un modèle d'ange –Au milieu- L'image correspondant aux valeurs de la mesure facteur de qualité, en niveau de gris- A droite- Les PIs détectés. 85
Figure 3-6 Images correspondant à -en haut- mesures de SI et -en bas- mesures de C 86
Figure 3-7 Classification SC dans le plan (κ_1, κ_2) (Akagündüz, 2011) 86
Figure 3-8 Formulation mathématique de la classification du couple HK&SC (Akagündüz, 2011) 87
Figure 3-9 L'espace couplé HK&SC de classification dans le plan (k_1, k_2) (Akagündüz, 2011) 88
Figure 3-10 Résultat du processus de croisement de deux espaces SC et HK sur deux modèles d'objets. 88
Figure 3-11 Formule de l'indice de confiance γ basée sur l'intensité de courbure C (Ho, et al., 2009) 89
Figure 3-12 Illustration du processus d'extraction des PIs en utilisant le détecteur SC_HK_C_Con. 90

Figure 3-13 a) Surface 3D artificielle sans bruit b) Directions des normales sur la surface non bruitée c) Surface 3D artificielle avec bruit. d) Directions des normales sur la surface bruitée (Bozkurt, et al., 2009) 91
Figure 3-14 De gauche à droite : Modèle Original – Ajout de bruit Artificiel - Modèle lissé par un lissage Laplacien – Modèle dé-bruité par un filtrage bilatéral 92
Figure 3-15 Pseudo-code de l'algorithme de filtrage bilatéral (Fleishman, et al., 2003) 92
Figure 3-16 Images de l'indice de forme pour différentes résolutions de la forme initiale 93
Figure 3-17 Les trois bases orthonormées utilisées pour l'estimation des courbures 95
Figure 3-18 A gauche- Les 42 objets de la base de Stuttgart –A droite- L'ensemble des images de la base d'apprentissage pour l'objet « machine » 96
Figure 3-19 En haut- 20 objets de la base Minolta, en bas-Images de profondeurs des 9 objets de la base Minolta utilisés pour notre évaluation (Chen, et al., 2007) 97
Figure 3-20 Les 7 objets présentant une variation d'échelle. Sur chaque ligne, nous illustrons les vues 0° de chaque échelle 98
Figure 3-21 Les 46 exemples d'objets segmentés de la base RGB-D (RGB12) 99
Figure 3-22 Deux vues du même objet pour la base RGB-D Object 99
Figure 3-23 Les 20 objets de la base carotte du laboratoire 100
Figure 3-24 Résultat d'extraction des PIs pour le processus de combinaison 101
Figure 3-25 Positions des PIs détectés par respectivement de haut vers les bas : Harris_fract, Harris_clust, SURF, SI et SC_HK_FQ pour les angles (0°, 20°, 40° et 80°) 103
Figure 3-26 Résultat de détection des PIs sur les objets de la base Carotte avec respectivement de gauche à droite: les détecteurs SC_HK_FQ (poubelle), SC_HK_C (bidon) et SC_HK_Conf (ventilateur) 104
Figure 3-27 Positions des PIs détectés par le détecteur SC_HK_C sur des objets de type mécanique pour les modèles agfa, machine et auto de la base Stuttgart 104
Figure 3-28 Les PIs détectés sur les objets: oignon, orange et tomate de la base RGB-D montrant la forte similarité de la forme. Une couleur de la surface est rajoutée pour différencier ces objets. 105
Figure 3-29 Résultat de la détection en échelle fixe (première ligne) et en multi échelles (deuxième ligne) pour le détecteur SC_HK_Con. La première colonne représente l'étendu des composantes connexes et la deuxième les positions des PIs finaux. 105
Figure 3-30 Invariance aux angles de vues pour l'approche SC_HK_Con multi-échelle. Les nombres représentent respectivement : l'ordre de la vue, le numéro de la vue dans la base Minolta et le nombre des PIs 106
Figure 3-31 18 Résultats de comparaison de l'approche SC_HK_Con muti-échelle par rapport à l'approche de (Castellani, et al., 2008) 106

Figure 3-32 Représentation du résultat de détection en utilisant trois niveaux de voisinages sur les images d'indice de forme. 107
Figure 3-33 Répétabilté des PIs pour les 10 détecteurs : SC_HK_FQ, SC_HK_C, SC_HK_Conf, SC_HK_C_con, Harris_fract, Harris_clust, SI, SC, HK et SURF sur les 9 objets de la base de Minolta. 108
Figure 3-34 Exemple de décimation avec un facteur f = 50% -à gauche le modèle original et à droite le modèle après décimation 110
Figure 3-35 Répétabilité des 9 détecteurs pour la transformation d'échelle Ech_1- A droite- Zoom sur la partie gauche de la première courbe 111
Figure 3-36 Répétabilité des 9 détecteurs pour la transformation d'échelle Ech_2 -A droite- Zoom sur la partie gauche de la première courbe 111
Figure 3-37 Répétabilité des 9 détecteurs SC_HK_FQ, SC_HK_C, SC_HK_Conf, SI, SC, HK, Harris_fract, Harris_clust et SURF entre sur 11 vues de deux échelles réelles 112
Figure 3-38 Répétabilité des PIs extraits par les détecteurs Harris_fract, Harris_clust, SC_HK_C et Surf dans l'ordre de gauche à droite, pour la même vue sur deux échelles : échelle 1 en première ligne et échelle 2 en deuxième ligne. 113
Figure 3-39 Courbes de répétabilité des 9 détecteurs testés sur les 9 objets Minolta et pour un bruit de variance 0.2 MR. 114
Figure 3-40 Répétabilité des PIs entre les deux vues 0° et 20° initiales (première ligne) et leur version bruitée (deuxième ligne) pour les détecteurs Surf (à gauche) et SC_HK_FQ (à droite) 115
Figure 4-1 Formation du système de coordonnées cylindriques pour un point orienté (Johnson, et al., 1998) 121
Figure 4-2 Processus de création de la représentation 2D du spin image (Johnson, et al., 1998) 121
Figure 4-3 Représentation des Spin-images pour deux valeurs de support pour trois points orientés de la surface du modèle de Duck. Dans l'accumulateur, les zones foncées correspondent à plus de points projetés dans les cases (Johnson, et al., 1998) 122
Figure 4-4 Deux plans par moindre carrés P_1 et P_2 et leurs normales (Flint, et al., 2007) 123
Figure 4-5 Histogramme 2D des indices de formes vs le produit scalaire entre la normale du point référence et ceux de ses voisins dans la représentation LSP (Chen, et al., 2007) 124
Figure 4-6 Structure spatiale du descripteur SHOT: 8 divisions en azimut, 2 divisions en élévation et 2 divisions en radial (par soucis de clarté seulement 4 divisions azimut sont représentées) (Tombari, et al., 2010) 127
Figure 4-7 Représentation du descripteur IndSHOT 128
Figure 4-8 Courbe rappel et courbe précision en fonction du nombre de bins 132
Figure 4-9 Courbe précision et courbe rappel en fonction du rayon du support 133
Figure 4-10 courbe temps de calcul en fonction du rayon du support 134

Figure 4-11 Résultat de la mise en correspondance après rotation des PIs entre la vue initiale à 100° et la vue à 120° des descripteurs IndSHOT 137
Figure 4-12 Courbes Rappel vs 1-Précision pour un changement d'angles entre les modèles Minolta 137
Figure 4-13 Représentations, au même point physique situés sur deux vues (100° et 120°), des descripteurs: IndThrift, LSP, Spin, IndSHOT, SHOT et Thrift, dans l'ordre de haut vers le bas 138
Figure 4-14 Courbe rappel-précision pour une décimation de 20% du maillage initial 140
Figure 4-15 Courbe rappel-précision pour un changement réel d'échelle pour les descripteurs SHOT, IndSHOT et IndThrift 141
Figure 4-16 Courbe rappel-précision pour un bruit, de σ = 0.1xMR -en haut, et de σ = 0.3xMR -en bas 143
Figure 4-17 En haut-Exemple de lissage du modèle de l'ange -En bas- Courbe rappel précision pour un facteur de relaxation f = 0.25xMR 144
Figure 4-18 Schéma de calcul du quotient R_d utilisé pour l'évaluation du pouvoir descriptif et discriminant d'un descripteur 146
Figure 4-19 Evolution de la distance entre le descripteur d'un point référence et les descripteurs de ses k plus proches voisins (en fonction de la distance physique du $k^{ème}$ plus proche voisin) 147
Figure 5-1 Représentation –a- de la division du plan pour la construction et –b- la recherche d'une requête dans un Kd-Tree (Hemant, 2005) 154
Figure 5-2 Représentation de la table de hachage (Chen, et al., 2007) 156
Figure 5-3 Filtrage des paires de PIs selon des contraintes géométriques (Chen, et al., 2009) 156
Figure 5-4 Les 5 objets Minolta ajoutés au 20 objets utilisés pour évaluer la robustesse à la rotation 161
Figure 5-5 Scénario de reconnaissance pour la rotation: la case rouge représente la vue n°2 prise en test pour l'objet 2 et les cases vertes correspondent aux vues d'apprentissage 162
Figure 5-6 Premier scénario de reconnaissance pour la variation d'échelle 166
Figure 5-7 Deuxième scénario de reconnaissance pour la variation d'échelle 167
Figure 5-8 Troisième scénario de reconnaissance pour la variation d'échelle 167
Figure 5-9 Les deux scénarios de reconnaissance pour l'évaluation de la robustesse au bruit- en haut: scénario B1 et en bas: scénario B2 172
Figure 5-10 Courbes d'évolution du taux de reconnaissance pour les 4 détecteurs SC_HK_confi, SC_HK_C, SC_HK_FQ et SI en fonction de 4 degrés de bruit, pour le descripteur IndSHOT (à gauche) et le descripteur IndThrift (à droite) 173
Figure 5-11 Les 5 objets sélectionnés de la base Stuttgart pour l'évaluation des primitives respectivement, de haut vers le bas, dans (Hozatlı, 2009) et (Eskizara, 2009) 175

Figure 5-12 Diagrammes de Hinton pour les deux ensembles E1 et E2 de la base Stuttgart. 177
Figure 5-13 Exemples de mise en correspondance des PIs de l'objet pitbull entre la vue test n°1 (à gauche) et les deux vues d'apprentissage n°0 et n°9 (à droite) 178
Figure 5-14 Diagramme de Hinton pour les 25 modèles de la base Stuttgart (94.93%).....179
Figure 5-15 Diagramme de Hinton pour les 42 modèles de la base Stuttgart (94.90%).....180
Figure 5-16 Diagramme de Hinton pour les 20 objets de la base Carotte (80%) en utilisant SC_HK_FQ et IndThrift 182
Figure 5-17 Diagramme de Hinton pour les 37 modèles de la base RGB-D (78.10%) 184

Liste des Tableaux

Tableau 1-1 Tableaux de Classification HK selon la fonction T_p (Akagündüz, 2011) 36
Tableau 1-2 Classification des surfaces selon la paire SC (Akagündüz, 2011) 39
Tableau 2-1. Classification des approches présentées 52
Tableau 2-2. Résumé des avantages et inconvénients des approches 3D et des approches 2D/3D 53
Tableau 2-3. Résumé des avantages et inconvénients des approches globales et des approches locales 54
Tableau 2-4. Comparaison des détecteurs de points d'intérêt 59
Tableau 2-5. Différences entre SIFT 2D et SIFT 2.5D (Lo, et al., 2008) 65
Tableau 2-6 Récapitulatif des méthodes de détection de Points d'Intérêt 3D 71
Tableau 2-7 Comparaison entre les descripteurs SIFT 2D, SIFT 2.5D et SIFT 3D 75
Tableau 2-8. Matrice de confusion 77
Tableau 3-1 Valeurs de répétabilité, à deux seuils de distance, entre les angles de vues pour les 10 détecteurs 109
Tableau 3-2 Valeurs de répétabilité pour un ajout de bruit gaussien de variance 0.2 MR au maillage 115
Tableau 3-3 Temps de calcul des détecteurs en (s) 116
Tableau 4-1 Catégorisation des cinq descripteurs 129
Tableau 4-2 Pouvoir descriptif des six descripteurs avec le ratio C_{dis} et le ratio R_d 146
Tableau 4-3 Temps de calcul (en seconde) des cinq descripteurs 148
Tableau 4-4 Temps de calcul (en seconde) des phases de détection et de description pour le couple (SC_HK_C, IndSHOT) et pour l'approche du SURF 148
Tableau 5-1 Algorithme de construction et de parcours récursives d'un Kd-Tree (Fleury) 154
Tableau 5-2. Algorithme de vote des PIs (Mayo, et al., 2009) 158
Tableau 5-3 Résultats de reconnaissance en variant les types des PIs 160
Tableau 5-4 Résultats de reconnaissance pour les 9 objets Minolta- les taux les plus élevés, pour chaque détecteur, sont mis en gras et le meilleur taux du tableau est mis en rouge 162
Tableau 5-5 Résultats de reconnaissance pour les 25 objets Minolta 164
Tableau 5-6 Taux de reconnaissance par objet pour le détecteur SC_HK_confi 165
Tableau 5-7 Taux moyen de reconnaissance selon la vue, pour le détecteur SC_HK_confi sur les 9 objets Minolta 165
Tableau 5-8 Résultats du scénario 1 sur les données d'échelle artificielle des 9 objets Minolta 168

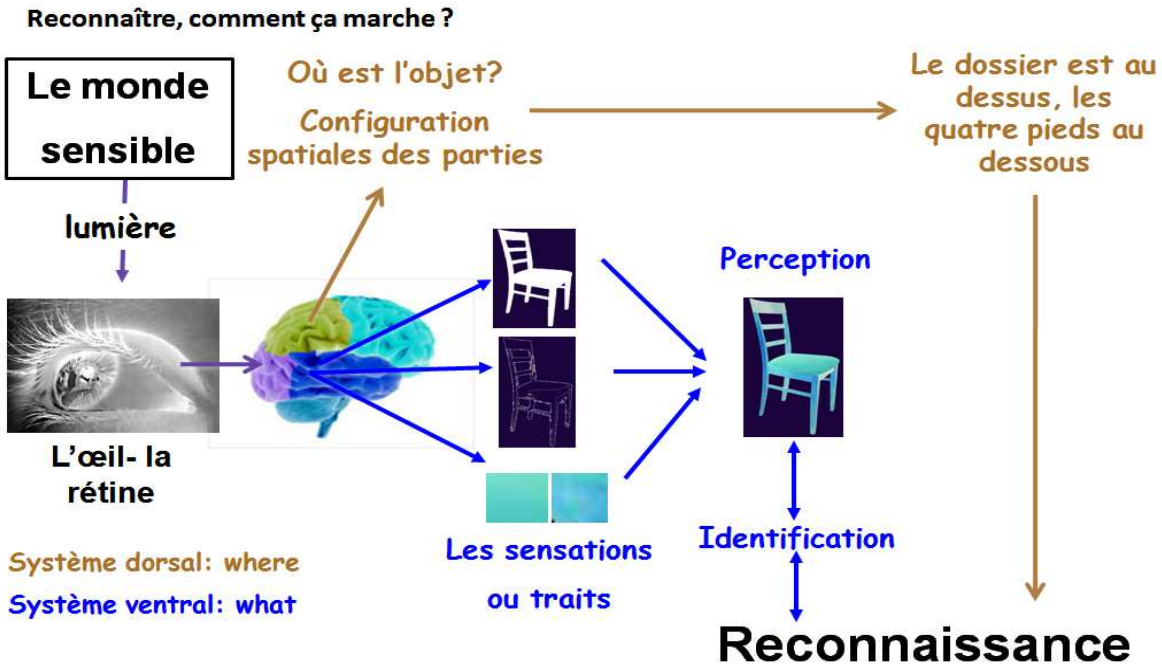
Tableau 5-9 Résultats scénario 3 sur les données d'échelle artificielle des 9 objets Minolta.....168
Tableau 5-10 Résultats du scénario 1 sur les données d'échelle réelle des 7 objets Minolta 169
Tableau 5-11 Résultats du scénario 2 sur les données d'échelle réelle des 7 objets Minolta 169
Tableau 5-12 Résultats du scénario 3 sur les données d'échelle réelle des 7 objets.....169
Tableau 5-13 Résultats du scénario B1 pour un ajout de bruit de variance 0.2 MR vs les résultats sans bruit, sur 9 objets Minolta.....172
Tableau 5-14 Résultats du scénario B1 pour un ajout de bruit de variance 0.1 MR et 0.5 MR sur les 9 objets Minolta 173
Tableau 5-15 Résultats du scénario B2 pour un ajout de bruit de variance 0.2 MR sur 9 objets Minolta.....174
Tableau 5-16 Résultats de reconnaissance sur les 5 objets Stuttgart de l'ensemble E1.....176
Tableau 5-17 Comparaison des taux de reconnaissance de notre méthode avec deux méthodes de la littérature 181
Tableau 5-18 Temps de calcul (s) de la phase de mise en correspondance 185
Tableau 5-19 Temps de calcul total des phases de détection, description et mise en correspondance pour les couples (SC_HK_C, IndSHOT) et (SC_HK_C, IndThrift) et pour l'approche du SURF 185
Tableau 5-20 Résumé sur la performance des systèmes proposés.....186

Acronymes

La littérature relative aux thèmes traités dans cette thèse est rarement traduite. Pour la plupart des acronymes employés, nous avons utilisé la dénomination la plus courante, qui est de fait en langue anglaise.

2D Bi-dimensionnel
3D Tri-dimensionnel
ACI Analyse en Composantes Indépendantes
ACP Analyse en Composantes Principales
CGAL Computational Geometry Algorithms Library
DBD Depth Buffer-based Descriptor
DFT Transformée de Fourier Discrète – Discrete Fourier Transform
DPD Distance de programmation dynamique
EGI Extended Gaussian Image
EVD Décomposition en valeurs propres – Eigen Value Decomposition
FFT Fast Fourier Transform
FOV Champs de vue – Field Of View
HK Courbure moyenne et courbure gaussienne – Mean Curvature and Gaussian Curvature
HMM Modèle(s) de Markov cachée(s) – Hidden Markov Model(s)
GMM Modèle de Mixture de Gaussiennes
KD K Dimensional
LSP Local Surface Patch
pdf fonction de densité de probabilité – Probality Density Function
PI Point d'Intérêt
SAI Spherical Attribut Images
SC Indice de forme et Intensité de courbure – Shape index and Curvedness
SHOT Signature of Histograms of OrienTations
SI Indice de forme – Shape Index
SIFT Scale Invariant Features Transform
SURF Speeded Up Robust Features
SVD Décomposition en valeurs singulières – Singular Value Decomposition
SVM Séparateur à Vaste Marge – Support Vector Machine(s)

Introduction



Comme l'illustre la figure ci-dessus, l'être humain s'est toujours référé à l'aspect visuel des objets de son environnement pour les **reconnaitre**. Cette capacité de différencier les entités est rattachée à un processus de détection et **décryptage** de **caractéristiques** propres à ces objets (des signes sur la couleur ou la forme) que nous avons stockées en **mémoire** lorsque nous avons vu ces objets précédemment.

La transposition dans nos systèmes informatiques de cette faculté de reconnaître des objets n'a pas cessé de susciter un grand intérêt, et constitue l'une des étapes essentielles pour la création d'une véritable Intelligence Artificielle. Les progrès des capteurs et des capacités de stockage et de manipulation d'information permettent aux ordinateurs de percevoir et interpréter de mieux en mieux leur environnement et les objets qui les entourent. La vision par ordinateur et en particulier la reconnaissance visuelle d'objets, si elles atteignent des niveaux satisfaisants de fiabilité et robustesse tout en fonctionnant en temps-réel (ou au moins quasi-réel), peuvent permettre de développer de nombreuses applications telles que l'accompagnement automatique de personnes à mobilité réduite, l'interprétation et l'identification de personnes pour la vidéo surveillance, l'aide et l'assistance pour des applications médicales, localisation et reconnaissance d'une position géographique, l'identification des composantes dans une chaine de fabrication industrielle, etc….

Aujourd'hui, une part grandissante des techniques de vision par ordinateur utilise l'information 3D pour attribuer une sémantique aux scènes observées. Plus particulièrement, la reconnaissance d'objets, richement étudiée en 2D, consiste à identifier et localiser des objets dans une image ou une séquence d'images. Ce concept a motivé plusieurs applications de robotiques où la tâche est de détecter et

reconnaitre des objets présents dans le champ de vue du robot. De plus en plus équipé de capteurs 3D, un besoin apparait d'analyser et interpréter efficacement les données acquises par le robot pour réaliser sa tâche. Il y a donc lieu de penser que des outils générant une représentation compacte et répondant aux critères de robustesse à des modifications de la forme de surface et d'invariance aux transformations géométriques, devront être mis en œuvre pour atteindre cet objectif.

✓ **Problématiques et motivation**

Dans une tâche de reconnaissance d'objets 3D par un système de vision par ordinateur, notre objectif est de reconnaitre un objet isolé, pour une vue donnée en requête, à partir d'une base d'apprentissage contenant quelques vues de cet objet. Certes l'information tridimensionnelle, par rapport à l'image 2D, résout la question de robustesse aux variations d'illumination, et apporte des indices sur la forme géométrique des objets pour aider à leur reconnaissance. Toutefois, la reconnaissance d'objet en 3D reste un sujet difficile et non totalement résolu.

- Tout d'abord, cette thèse considère le problème de la reconnaissance d'objets 3D dans le contexte d'une reconnaissance à partir de vues partielles d'objets. Selon l'angle de vue acquise, la partie capturée de l'objet change. Mais également, les détails de la forme sont accentués ou atténués en fonction de la résolution du capteur et de la distance de l'objet au capteur. En conséquence, au moment de la requête, la vue présentée au système de reconnaissance est plus au moins différente des vues apprises. La nécessaire robustesse à ces variations de l'aspect géométrique de la forme constitue un réel défi.
- Dans le cas ou un seul objet se présente sur l'image, la tache de le reconnaitre devrait être plus facile. Néanmoins, des régions erronées, comme l'ombre ou les occultations, peuvent apparaître sur les images de profondeur. Ces artéfacts liés aux limitations inévitables des techniques de capture, les systèmes de reconnaissances d'objets doivent y remédier.
- Par ailleurs, le système développé doit tenir compte des contraintes matérielles de la machine, du point de vue complexité de calcul et compacité en mémoire. D'un côté, le temps de calcul peut être élevé, voire prohibitif, en raison de la grande quantité de données à traiter (coordonnées x, y, z). Et d'un autre coté, la capacité mémoire nécessaire peut être un handicap pour conserver ces données.

Résoudre au mieux tous ces problèmes constitue un riche champ d'études. Face à cette problématique, les questions suivantes se posent :

- Quels procédés de traitement de ces données 3D et d'apprentissage doit-on mettre en œuvre pour extraire une description unique et robuste de la forme des objets?
- Comment peut-on par ailleurs tirer parti de l'existant en 2D et en 3D pour améliorer les représentations existantes des objets 3D, ou l'exploiter autrement?

✓ **Approche**

Dans notre recherche sur le sujet, nous nous sommes orientés vers quelques travaux, basées sur des approches dites locales. A l'issue de notre étude, nous avons suggéré une méthodologie hybride, qui combine certains aspects de ces méthodes. L'idée principale est d'extraire des primitives renseignant sur la variation locale de la forme à base de normales et de courbures. Des mesures de saillance, basées sur ces quantités différentielles, permettent de sélectionner des points d'intérêt et de décrire la surface locale autour de ces points. Cette proposition a apporté une amélioration sur la performance de la reconnaissance. Nous détaillons ici nos motivations à suivre cette voie.

- Dans la littérature, la description de la forme 3D est approchée par des méthodes divisées en deux grandes catégories: globales et locales. Les globales décrivent la forme des objets dans sa

totalité. Intuitivement, on se dit que le caractère global offre plus de facilité à identifier un objet inconnu (comme la silhouette, les contours). En revanche, en général, certains détails de la forme, qui peuvent différencier les objets d'une même catégorie par exemple, ne sont pas pris en compte par ces primitives. Dans la même optique, ces approches sont sensibles aux déformations de la topologie locale. Une représentation plus adéquate consisterait en l'extraction d'une information locale sur des zones saillantes de la surface de l'objet. Le constat qui justifie cette solution, est que plusieurs points de l'objet peuvent être redondants ou ne sont pas très informatifs de la spécificité de l'objet.

- Le potentiel d'une représentation locale de la forme a été confirmé pour la reconnaissance en 2D. En particulier, celles à base de points d'intérêt a connu un grand succès comme le SURF et le SIFT. Face aux situations d'occultations et de bruit, là encore, la localité réussit à différencier un objet d'un autre.
- Un savoir faire des approches de points d'intérêt 2D, dans notre laboratoire, a également motivé notre choix porté sur ce type de techniques.
- Il existe en outre des bases de données d'images de profondeurs publiques sur lesquelles nous pouvons tester la performance de nos algorithmes et nous comparer aux approches existantes. Les natures différentes de ces bases nous laissent imaginer divers scénarios d'usage de notre système de reconnaissance.

✓ **Plan d'étude et résumé des contributions**

Dans ce manuscrit, nous présentons tout d'abord les notions de bases utilisées tout au long de notre travail. Dans la première partie du chapitre 2, nous regroupons dans plusieurs catégories, différentes approches de la littérature proposant des solutions 2D ou 3D au problème de reconnaissance 3D. Par la suite, dans une deuxième partie, nous listons quelques approches à base de points d'intérêt 3D.
Après la présentation dans le chapitre 3 des détecteurs références de notre travail, nous détaillons ensuite les principes de notre contribution pour former les nouveaux détecteurs. Nous démontrons notre apport par les résultats exposés dans la partie évaluation comparative des détecteurs testés. Dans le chapitre 4, nous introduisons le formalisme des descripteurs qui nous ont inspiré pour la mise en place de deux nouveaux descripteurs. Cette section est suivie, dans le même chapitre, par une évaluation de la performance des différents descripteurs.
Après une énumération de quelques techniques d'apprentissage utilisées dans la reconnaissance 3D, la deuxième partie du chapitre 5 démontre le potentiel des nouvelles techniques proposées dans un système de reconnaissance d'objets.
Enfin, le chapitre conclusion propose diverses perspectives de recherche, liées aux détecteurs de points d'intérêt 3D, aux descripteurs 3D et à l'approche de mise en correspondance proposée.

Nous listons maintenant nos contributions principales :

- Constitution d'un état de l'art, non exhaustif, de différentes catégories d'approches de reconnaissance 3D
- La présentation d'un nouveau critère de détection à partir de la combinaison des deux espaces de classification de la forme locale des surfaces: SC et HK
- Proposition et implémentation de quatre nouveaux détecteurs basés sur des mesures de la courbure que nous avons appelé (SC_HK_C, SC_HK_ FQ, SC_HK_Conf et SC_HK_Con)
- Comparaison de ces détecteurs avec des détecteurs connus dans le domaine des points d'intérêt comme le Harris 3D et le SURF 3D en évaluant quantitativement la répétabilité et la robustesse

- Formulation de deux nouveaux descripteurs IndSHOT et IndTHrift en rajoutant une information sur l'indice de forme aux descripteurs SHOT et Thrift de la littérature.
- Etude et comparaison du pouvoir descriptif et de la stabilité de ces descripteurs
- Développement d'un système complet de reconnaissance d'objets 3D, comportant les étapes de détection de points d'intérêt, leur description et leur mise en correspondance sur des vues 3D de l'objet
- Evaluation de la performance de notre système, en termes de taux de reconnaissance, sur diverses bases de données d'images de profondeurs provenant de capteurs différents, et comparaison avec des systèmes existants, pour des scénarios, réels et artificiels, évaluant l'invariance aux transformations géométriques et la robustesse au bruit.
- Proposition d'une extension de la méthode proposée pour une version multi-échelle qui serait plus robuste
- Participation, dans le laboratoire, à l'enregistrement et l'exploitation de données de bases acquises avec des caméra 3D dont la base Carotte de données de Kinect.

1. CHAPITRE : TERMINOLOGIE ET NOTIONS DE BASE SUR LA 3D

L'un des principaux atouts de l'acquisition 3D du monde réel est la possibilité d'avoir une information de profondeur (distance) indépendante de la luminosité et une information sur les dimensions réelles de la scène. Cette acquisition peut se faire par différents capteurs comme les scanners 3D, les systèmes de stéréo, les scanners lidar, les caméras temps de vol, etc. Le signal capturé est souvent sous le format d'une grille ou un nuage de points 3D. Cependant, la grande quantité d'informations contenues dans ce signal est difficilement exploitable pour des applications temps-réel. Pour pallier ce problème, une représentation plus compacte, robuste au bruit et invariante à l'échelle et aux transformations géométriques, est requise.

1.1. Types de données 3D

1.1.1 Nuage de points

La sortie initiale et commune des scanners laser ou des caméras 3D est un nuage de points (Figure 1-1). Ces points sont considérés comme des échantillons de la surface de l'objet et sont arrangés sous formes de système de coordonnées en des triplets (X, Y, Z) avec Z représentant la profondeur. Ce nuage est souvent transformé, pour une meilleure visualisation, en un maillage polygonal avec une reconstruction surfacique. Selon le type du capteur, le nuage de points peut être organisé (structuré) ou désorganisé (non structuré).

Figure 1-1. Nuage de point d'un modèle 3D de voiture

1.1.2 Image de profondeurs

Une image de profondeur (range image) ou image 2.5D est une image en niveau de gris dans laquelle l'intensité est inversement proportionnelle à la distance des points de la scène vue sous un seul angle.

Les techniques de mesure de profondeurs (distances) sont multiples. Nous citons à titre d'exemple, le **scanner par temps de vol** qui utilise un faisceau laser pour sonder la profondeur. Au cœur de ce type de scanner se trouve un télémètre laser permettant de calculer la distance à la surface de l'objet étudié en mesurant le temps nécessaire au trajet d'aller-retour de l'impulsion du faisceau laser réfléchi. Puisque la vitesse de la lumière est connue, le temps de retour permet de déterminer la distance

parcourue par la lumière, qui est deux fois la distance entre le scanner et la surface de l'objet. Évidemment, l'exactitude du scanner par temps de vol dépend de la précision de la mesure du temps de retour.

La technologie capteur de profondeur 3D (3D depth sensor) à base de **caméra infrarouge** est un autre exemple et elle permet d'obtenir une image représentant les dégagements thermiques émis par l'objet observé. Le fonctionnement est similaire à celui d'une caméra RGB sauf qu'uniquement les rayons infrarouges sont émis et filtrés. La scène est bombardée par les rayons infrarouges non visibles à l'œil. Une partie de ces rayonnements sera retournée par l'ensemble des objets de la scène. Plus l'objet est loin et plus la quantité de rayonnement infrarouge réfléchie est faible. Ainsi, la caméra infrarouge mesure la distance de l'objet en fonction de l'intensité. Des niveaux de couleurs peuvent être rajoutés en fonction de la température mesurée. La Kinect™ est une caméra 3D qui comporte une source de lumière infrarouge « structurée », une caméra infrarouge et une caméra couleur (Figure 1-2).

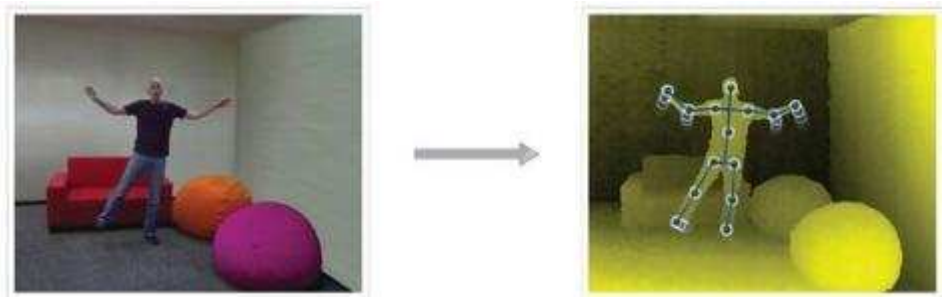


Figure 1-2. L'image RGB d'une scène réelle (à gauche) et l'image recueillie par la caméra 3D Kinect™ (à droite)

Une autre façon d'obtenir des images de profondeur est de générer à partir d'un modèle 3D complet un ensemble de **projections 2.5D** correspondant à différents angles de vue. Dans une approche d'indexation, Zaharia et Prêteux (Zaharia, et al., 2002) définissent une boîte cubique de projection sur laquelle les images de profondeur sont calculées. De même, Chaouch et Verroust-Blondet (Chaouch, et al., 2007) projettent les modèles 3D sur les directions normales associées aux vingt sommets d'un dodécaèdre régulier (Figure 1-3).

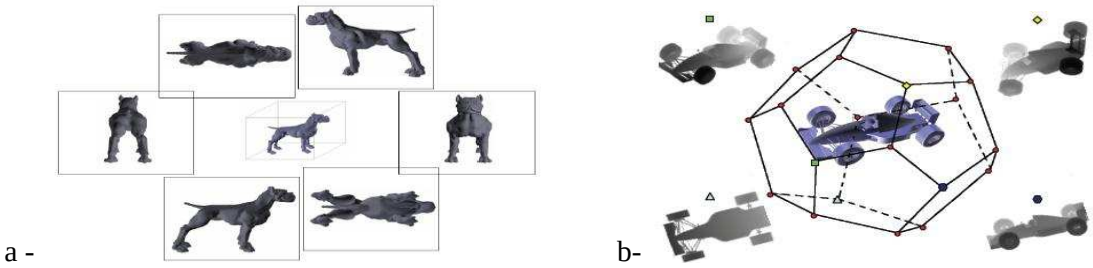


Figure 1-3. -a- Six images de profondeur associées aux six faces de la boîte de projection (Zaharia, et al., 2002) ; -b- les 20 angles de vues associés aux sommets du dodécaèdre (Chaouch, et al., 2007)

Nous mentionnons également la technique utilisant la **stéréoscopie** qui permet, à partir de deux images 2D, de produire une perception du relief. Les systèmes stéréoscopiques utilisent généralement deux caméras vidéo, situées côte à côte et visant la même scène. Un système de caméra stéréo peut

être utilisé pour déterminer, avec plus ou moins d'approximation, la profondeur des points dans la scène, par exemple à partir du point central sur la ligne située entre les deux foyers. Pour mesurer la profondeur à l'aide d'un système stéréo, il est nécessaire, d'abord, de trouver les points qui s'apparient dans les différentes images. La résolution de la problématique d'appariement est l'un des principaux problèmes de cette technique (régions d'intensité ou couleur homogène), en plus du coût des capteurs (multiplié par 2), et surtout de la critique étape de calibration de la paire stéréo.

1.1.3 Données polygonales

Pour des raisons de visualisation en temps réel de modèles tridimensionnels, les applications basées sur les graphiques et la vision par ordinateur utilisent des surfaces polygonales (polyédriques). Généralement, à partir des images de profondeurs, l'information 3D des points est convertie en des données 3D polygonales. Contrairement aux images de profondeurs qui donnent uniquement les points perçus sous l'angle de vue, les données polygonales peuvent représenter le modèle de l'objet complet. La combinaison des plusieurs images de profondeurs capturées sous différentes prises de vue permettent de construire un maillage polygonal du modèle complet. La représentation d'une surface 3D par une subdivision en un ensemble de polygones s'appelle maillage. Un maillage comporte des sommets connectés les uns aux autres par des faces ou facettes de forme polygonale. Parmi les types de maillages les plus courants, nous mentionnons la triangulation ou maillage triangulaire (trimesh) qui désigne un maillage où les facettes sont des triangles. On parle aussi des maillages volumiques lorsque les sommets sont reliés par des tétraèdres.

Pour des représentations de haute complexité, il est possible de représenter le modèle avec des niveaux de détail plus ou moins fins selon les besoins de l'utilisateur et le type d'application. L'utilisation de méthodes de simplification géométrique permet de réduire la complexité des objets pour en produire des versions simplifiées à partir de leurs descriptions originales (Figure 1-4). Par exemple, une zone avec une faible courbure (ex. zone plane) peut être représentée avec peu de sommets. Cependant, selon la technique d'acquisition, cette zone peut contenir beaucoup de points 3D redondants. Dans ce cas, une décimation est appliquée sur le maillage polygonal. Les zones à faible valeur de courbure sont sous-échantillonnées.

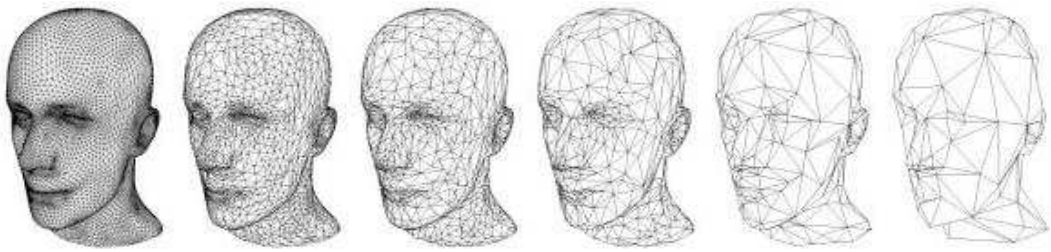


Figure 1-4. Exemple de simplification avec différentes valeurs de subdivision régulière

1.1.4 Prétraitements des données

Les données utilisées dans un système de reconnaissance 3D se présentent sous différents formats possibles, par exemple:

- nuages de points 3D
- images de profondeurs

- représentation polyédrique

Un exemple d'images de type intensité 2D, 2.5D et une représentation polygonale du modèle d'un visage est illustré dans la Figure 1-5.

Figure 1-5. Trois représentations du visage: image 2D en niveau de gris, image de profondeur, et modèle surfacique

Après acquisition de ces données, des prétraitements sont souvent appliqués, par exemple, la détection d'objets qui délimite et isole la zone d'intérêt des objets de la scène. De plus, selon la finalité de l'application et le scénario envisagé pour le système de reconnaissance, les données peuvent soit être exploitées sous le format de vues partielles, ou bien on leur applique un processus de reconstruction pour avoir le modèle 3D complet de l'objet.

Dans le traitement des données 3D réelles, le problème du bruit est souvent présent. Il peut se manifester par exemple lors de l'acquisition par un effet de bord qui apparait dans le cas où la lumière frappe le fond et le bord de l'objet au même pixel, ce qui donne des points aberrants ("outliers"). Diverses techniques se sont intéressées au traitement de ce problème dans la littérature. Nous citons à titre d'exemples :

- Pour le dé-bruitage des images de profondeurs, les techniques de filtrage des données classiques en 2D sont souvent utilisées. Par exemple, un lissage peut s'appliquer à l'information de distance (z) avec un filtre moyen ou médian adaptatif de distance. Ces filtres utilisent un masque de taille différente (ex. 3 × 3, 5 × 5, or 7 × 7) selon la valeur de distance en chaque pixel.
- Sur un nuage de points 3D, Ramaswamy et al. (Ramaswamy, 2000) utilisent les k plus proches voisins pour repérer les points aberrants : ils calculent, pour chaque point, la distance moyenne carrée aux K plus proches voisins et ordonnent les points par rapport à cette distance, et enfin extraient les n données les plus déviantes.
- Kalogerakis et al. (Kalogerakis, et al., 2009) traitent, aussi, le problème de dé-bruitage du nuage de points 3D et des points du bord automatiquement par une minimisation d'énergie. Un point du bord se présente si le nombre de ses voisins est inférieur à un seuil (Figure 1-6). Le voisinage est défini par les points les plus proches se trouvant sur le plan tangent subdivisé en 6 tranches (des angles de 60°). Un point du bord est défini par un point dont le nombre de voisins est inférieur à un seuil (égal à 3). Les bords peuvent apporter de l'information utile, il est alors judicieux de ne pas être très restrictif sur le seuil. Dans leur processus de calcul de la courbure en un point p_i, Kalogerakis et al. estiment un poids pour chaque paire des points (p_i, p_j), avec p_j est un point de son voisinage. Pour chaque point p_i, une mesure qui représente la contribution de chaque point du voisinage à l'estimation de la courbure est introduite. Un point p_j est un point aberrant si l'évaluation de cette mesure est égale à 0, ce qui signifie que la variation des normales, associée à la paire (p_i, p_j), n'est pas impliquée dans la définition de la

courbure au point p_i. Une technique de correction des points mal placés qui consiste en une minimisation d'énergie est également proposée.

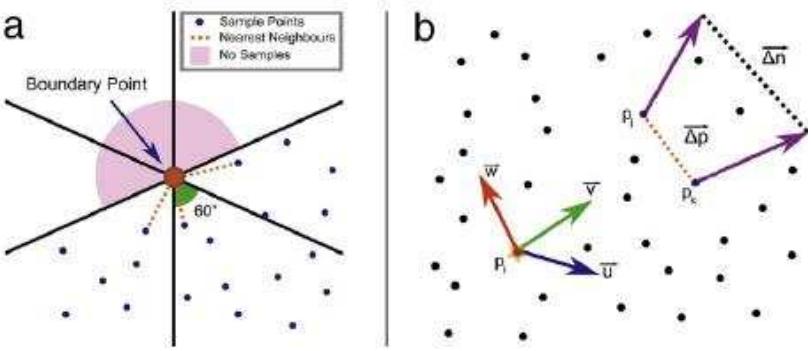


Figure 1-6. a- Point du bord (définition et conditions); b-Variation de la normale pour l'estimation de la courbure (Kalogerakis, et al., 2009)

Dans le but de trouver une représentation de haut niveau de l'apparence de l'objet, la plupart des techniques utilisées font appel à la géométrie différentielle pour extraire des primitives à base de normales et de courbures (Figure 1-7).

Figure 1-7. Mesure du caractère « plus ou moins courbé » de l'objet : courbure (à gauche), et normales (à droite).

1.2. Mesures quantitatives du degré de courbure

Dans cette partie, nous allons présenter des notions d'estimateurs de la déviation de la surface locale. Ces mesures vont être utilisées dans la suite de ce manuscrit.

1.2.1 Normale et Courbure

- **Normale d'un point 3D**

La normale désigne la quantité différentielle d'ordre 0. L'orientation dominante en un point 3D est celle de la normale de la surface locale définie par le voisinage du point. L'estimation de la normale en un point 3D constitue une problématique en soit.

Plusieurs techniques d'évaluation de la normale en un point sont possibles :

* Ajustement d'une fonction analytique continue aux données locales (le voisinage) discrètes avec les moindres carrés. Le problème d'ajustement est approché par une résolution d'équations (Al-Osaimi, et al., 2007). Plus simplement, si nous supposons que l'équation du plan contenant le point est:

$$ax + by + cz + d = 0 \text{ ; la normale est donnée par le vecteur transposé } [a\ b\ c]^T$$

* Dans le cas d'un maillage, la façon la plus simple pour calculer la normale en un sommet est de prendre la moyenne des normales des facettes auxquelles il appartient. La normale d'une facette est le produit vectoriel de deux vecteurs constitués par les sommets de cette facette. La normale en un point i est la moyenne des normales $\overrightarrow{n_{jk}}$ calculées sur son voisinage, et donnée par :

$$\overrightarrow{n_{jk}} = \frac{\overrightarrow{v_{ij}} \times \overrightarrow{v_{ik}}}{\left|\overrightarrow{v_{ij}} \times \overrightarrow{v_{ik}}\right|}$$

avec $\overrightarrow{v_{ik}}$ le vecteur formé par le point i et le point voisin k. L'opérateur x désigne le produit vectoriel. Les directions des normales sont contrôlées par le produit scalaire de deux normales. Si ce produit est négatif, une des deux directions doit être changée en son opposé (Biosca, 2008). Cependant, en présence de bruit, cette méthode échoue dans l'estimation des normales.

- **Courbure**

La courbure d'un objet géométrique est une mesure quantitative du caractère "plus ou moins ondulé ou plat" de cet objet, ou encore l'intensité du changement dans la rotation des lignes de points (arc) qui le constituent. La courbure désigne la quantité différentielle d'ordre 1. La définition formelle de la courbure est donnée par l'équation suivante, ou K est la courbure, ϕ est l'angle tangentiel et S est la longueur de l'arc :

$$K = \frac{d\Phi}{ds}$$

Équation 1.1

En prenant l'équation 2D paramétrique x = x(t) et y = y(t), cette équation peut s'écrire sous la forme:

$$K = \frac{\frac{d\Phi}{dt}}{\frac{ds}{dt}} = \frac{\frac{d\Phi}{dt}}{\sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}} = \frac{\frac{d\Phi}{dt}}{\sqrt{x'^2 + y'^2}}$$

Équation 1.2

Pour simplifier, l'Équation 1.3 de l'angle tangent peut être utilisée :

$$\tan\Phi = \frac{dy}{dx} = \frac{dy/dt}{dx/dt} = \frac{y'}{x'}$$

Équation 1.3

Le terme $\frac{d\phi}{st}$ de l'Équation 1.2 est obtenu par les dérivées par rapport à t dans l'Équation 1.3. Le calcul suivant permet d'obtenir l'Équation 1.4.

$$\frac{d}{dx}(\tan\Phi) = \sec^2\Phi\frac{d\Phi}{dt} = \frac{x'y''-x''y'}{x'^2}$$

$$\frac{d\Phi}{dt} = \frac{1}{\sec^2\Phi}\frac{x'y''-x''y'}{x'^2}$$

$$\frac{d\Phi}{dt} = \frac{1}{1+\tan^2\Phi}\frac{x'y''-x''y'}{x'^2}$$

$$\frac{d\Phi}{dt} = \frac{1}{1+\left(\frac{y'}{x'}\right)^2}\frac{x'y''-x''y'}{x'^2}$$

$$\frac{d\Phi}{dt} = \frac{x'y''-x''y'}{x'^2+y'^2}$$

Équation 1.4

Combinant les équations Équation 1.2 et Équation 1.4

, la valeur de la courbure K est obtenue par:

$$K = \frac{x'y''-x''y'}{\left(x'^2+y'^2\right)^{3/2}}$$

Équation 1.5

Cette équation peut être minimisée sous la forme:

$$K \approx \frac{d^2y}{dx^2}$$

Équation 1.6

Les courbures ont été utilisées par exemple, en reconnaissance des visages car elles permettent de décrire les détails des zones comme le nez et la bouche. Le problème essentiel dans l'estimation des courbures est la définition du voisinage au point, surtout quand le nuage de points 3D n'est pas ordonné. Des algorithmes comme les K-moyennes floues, améliorés en incorporant dans la mesure de distance les matrices de covariances floues, s'avèrent très performants pour faire émerger des structures particulières (Loménie, et al., 1999). Kalogerakis et al. (Kalogerakis, et al., 2009) ont été les premiers à présenter une méthode pour extraire les lignes de courbure directement sur le nuage de point. En effet, ils proposent une nouvelle approche pour une estimation robuste et statistique des normales et des courbures d'une surface. Notons c le point sur lequel la courbure est calculée par le point central. En utilisant le nombre minimum de points voisins les plus proches dans les six tranches de son plan tangent, le voisinage du point central c est estimé. Une échelle pour le calcul de la courbure est définie par la sphère de rayon proportionnel à la moyenne des distances du point central à ces voisins. Le tenseur de courbure est estimé premièrement par la variation de la position et des normales d'une paire de points sur ce voisinage. Cette variation est définie dans un repère de coordonnées locales orthonormées obtenue par le plan tangent. Le tenseur est raffiné en s'appuyant sur le processus IRLS (Iterative Reweighted Least Squares). En effet, l'algorithme M-estimation attribue

des poids aux variations des normales de chaque paire et les ré-estime jusqu'à convergence (Kalogerakis, et al., 2007). Comme les courbures sont calculées avec des dérivées secondes, ces techniques sont sensibles au bruit et aux petites ondulations dans la surface de l'objet et notamment dans le cas multi-résolution.

1.2.2 Courbures principales

L'extraction des courbures principales fait partie des méthodes les plus répandues pour caractériser une surface 3D. Ces courbures, en un point donné, correspondent à l'intersection des plans contenant les vecteurs des normales aux surfaces, avec les plans tangents au point en question. La valeur maximale et la valeur minimale de ces courbures forment les courbures principales et sont notées respectivement κ_1 et κ_2 (Figure 1-8). En un point d'une image X(u, v), la courbure principale et ses directions principales sont calculés par la décomposition en valeurs propres de la matrice Hessienne définie par :

$$\mathbf{H} = \begin{bmatrix} \dfrac{\partial^2 \mathbf{X}}{\partial u^2} & \dfrac{\partial^2 \mathbf{X}}{\partial uv} \\ \dfrac{\partial^2 \mathbf{X}}{\partial uv} & \dfrac{\partial^2 \mathbf{X}}{\partial v^2} \end{bmatrix}$$

Équation 1.7

Les valeurs propres de cette matrice symétrique donnent les courbures principales κ_1 et κ_2 et les vecteurs propres correspondent aux directions principales. Le signe de la courbure principale est positif si elle a la même direction que le vecteur normale et est négatif si l'inverse. Les directions principales sont toujours perpendiculaires. Ces courbures informent sur la forme locale autour d'un point et permettent de classer les points de la surface.

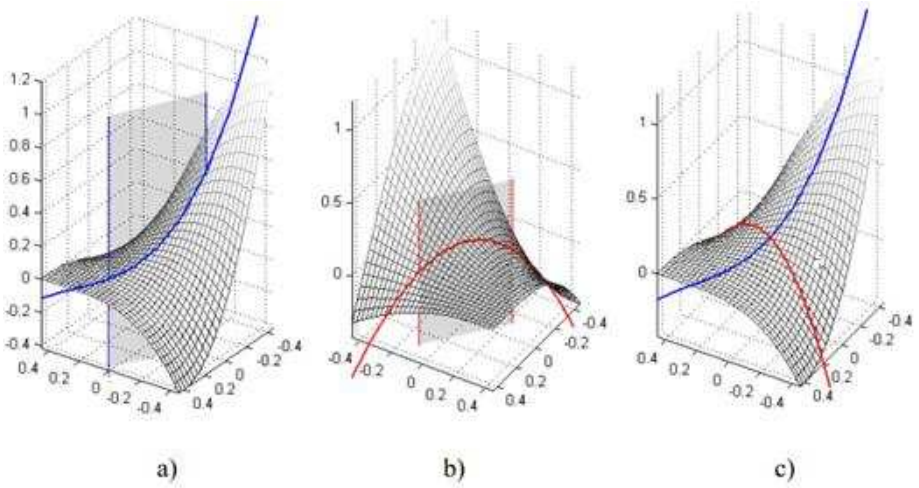


Figure 1-8. a- Le plan de la normale avec la courbure maximale. -b- Le plan de la normale avec la courbure minimale. -c- Exemple de valeurs de courbures principales

κ_1 = 1.56 et κ_2 = -2.37. La surface est de la forme d'une selle de singe (monkey saddle) d'équation: $z(x,y) = x^3 - 3 \bullet x \bullet y^2$.

Des estimateurs à base des courbures principales ont été proposés comme les courbures gaussiennes et moyennes, l'indice de forme et l'intensité de courbure.
Dans ce qui suit, nous introduisons ces estimateurs.

1.2.3 Courbure gaussienne (K) et moyenne (H)

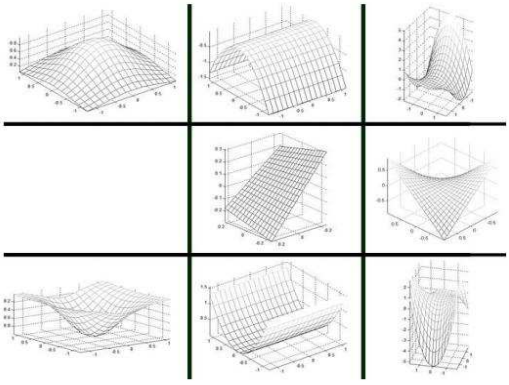
Les courbures gaussiennes K et moyennes H sont introduites par Besl (Besl, et al., 1988). K est le produit des courbures principales et renseigne sur le caractère localement elliptique ou hyperbolique de la surface. Le signe de K est positif si la surface locale a une forme de sommet ou de vallée, et est négative si la forme est une selle. La région est plane si l'une au moins des courbures principales est nulle.
La moyenne des courbures principales définit H et indique de combien la surface est courbée.
Les formes locales sont classées selon les valeurs de H et K qui sont données par :

$$H = \frac{k_1 + k_2}{2} \; ; \qquad K = k_1 . k_2$$

Équation 1.8

1.2.4 Espace HK

Besl et Jain (Besl, et al., 1988) ont été les premiers à proposer d'utiliser les valeurs de H et K pour la segmentation de la surface. Selon les valeurs de la paire HK, les régions peuvent être classées comme suit dans la Figure 1-9 :



	K>0	K=0	K<0
H<0	Convex (1) (*Elliptic or Umbilic*)	Ridge (2) (Convex *Parabolic*)	Saddle Ridge (3) (*Hyperbolic*)
H=0	(*Not possible*)(4)	Planar (5)	Minimal (6)
H>0	Concave (7) (*Elliptic or Umbilic*)	Valley (8) (Concave *Parabolic*)	Saddle Valley (9) (*Hyperbolic*)

Figure 1-9. Classification des surfaces selon les valeurs de H et K (Akagündüz, 2011)

A cause du bruit existant sur les surfaces réelles, il est impossible d'avoir les valeurs de H et K égales à zéro. Des seuils sur ces valeurs sont utilisés. Les valeurs au dessous de ces seuils sont considérées comme nulles. Une Fonction T_p (Equation 1.9) est introduite pour donner une valeur entre 1 et 9 à la surface selon la valeur de la paire (H, K) (Tableau 1-1).

$$T_p = 1 + 3\left(1 + \mathrm{sgn}_{\varepsilon_H}(H)\right) + \left(1 - \mathrm{sgn}_{\varepsilon_K}(K)\right);\ \mathrm{sgn}_{\varepsilon_X}(X)\begin{cases} +1 & if\ X > \varepsilon_X, \\ 0 & if\ |X| \leq \varepsilon_X, \\ -1 & if\ X < \varepsilon_X \end{cases}$$

Equation 1.9

Mean curvature H	Gaussian curvature K		
	$K > 0$	$K = 0$	$K < 0$
$H < 0$	Peak $T_p = 1$	Ridge $T_p = 2$	Saddle ridge $T_p = 3$
$H = 0$	None $T_p = 4$	Flat $T_p = 5$	Minimal $T_p = 6$
$H > 0$	Pit $T_p = 7$	Valley $T_p = 8$	Saddle valley $T_p = 9$

	K>0	K=0	K<0
H<0	Convex (1) (*Elliptic or Umbilic*)	Ridge (2) (Convex *Parabolic*)	Saddle Ridge (3) (*Hyperbolic*)
H=0	*(Not possible)*(4)	Planar (5)	Minimal (6)
H>0	Concave (7) (*Elliptic or Umbilic*)	Valley (8) (Concave *Parabolic*)	Saddle Valley (9) (*Hyperbolic*)

Tableau 1-1 Tableaux de Classification HK selon la fonction T_p (Akagündüz, 2011)

La Figure 1-10 illustre la classification HK dans le plan (κ_1, κ_2). Les droites séparatrices sont les seuils-zéro des valeurs de H et d'équations :

$$\kappa_1 + \kappa_2/_2 = +H_{zéro} \quad \text{et} \quad \kappa_1 + \kappa_2/_2 = -H_{zéro}$$

Équation 1.10

Les courbes séparatrices sont les seuils-zéro des valeurs de K et d'équations:

$$\kappa_1 . \kappa_2 > +H_{zéro} \quad \text{et} \quad \kappa_1 . \kappa_2 > -H_{zéro}$$

Équation 1.11

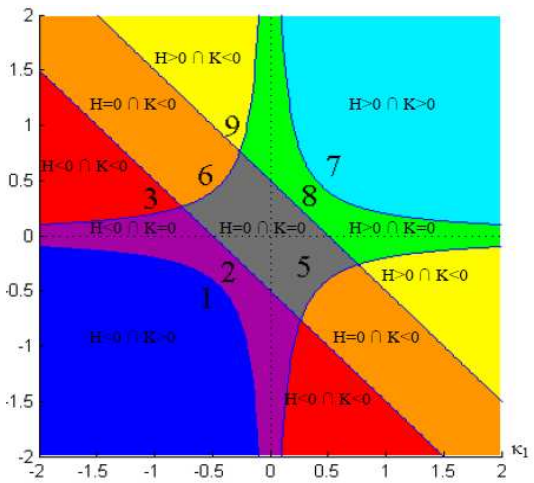


Figure 1-10. Classification HK dans le plan (k_1, k_2) (Akagündüz, 2011)

Les régions colorées correspondent au Tableau 1-1.

Comme indiqué dans l'analyse de Cantzler et Fisher (Cantzler, et al., 2001), les régions planes ne sont pas symétriques et la zone des régions paraboliques devient très fine pour les grandes valeurs de courbures. De plus, les valeurs de H et K dépendent fortement des seuils. Un décalage des valeurs des seuils entrainent des changements dans la répartition des régions dans le plan (κ_1, κ_2). Les valeurs de κ_1 et κ_2 sont invariants aux changements d'orientation mais varient pour un ré-échantillonnage ou un changement d'échelle de la surface locale. Ainsi, H et K ne sont pas invariants à l'échelle et à la résolution mais invariants à l'orientation.

1.2.5 Indice de forme (S)

Partis du constat que les courbures gaussiennes et moyennes n'informent pas suffisamment sur la forme locale de la surface, Koenderink et Doorn (Koenderink, et al., 1992) proposent une autre mesure quantitative invariante à l'échelle appelée « indice de forme » SI_p. Cet estimateur est défini en un point p par l'Équation 1.12. Modifié par Dorai et Jain (Dorai, et al., 1997), la formule de SI_p est donnée dans l'Équation 1.13.

$$SI_p = \frac{2}{\pi} \times \mathrm{arctg}\left(\frac{k_p^1 + k_p^2}{k_p^1 - k_p^2}\right);\ k_p^1 > k_p^2$$

Équation 1.12

$$SI_p = \frac{1}{2} - \frac{1}{\pi} \times \mathrm{arctg}\frac{k_p^1 + k_p^2}{k_p^1 - k_p^2}$$

Équation 1.13

Dans ces formules, k^1_p and k^2_p sont respectivement la courbure maximale et la courbure minimale. Avec la définition de l'Équation 1.13, à chaque forme va correspondre une valeur unique de SI et est classée dans l'intervalle [0,1] (sauf pour les surfaces planes ou les courbures principales sont nulles). Dans l'Équation 1.12, l'intervalle de variation des valeurs est [-1,1]. Les grandes valeurs de l'indice de

forme correspondent à des surfaces convexes et les petites valeurs représentent des surfaces concaves. La graduation continue des valeurs de l'indice de forme est subdivisée par des seuils fixes en des sous intervalles correspondants à différents types de forme. Dans la définition originale de Koenderink et Doorn, ils ne différentient pas les régions hyperboliques en trois types vallée anticlinale, hyperbolique et saddle ridge. Une autre valeur de constante (3/16) a été alors rajoutée par (Akagündüz, 2011) (Figure 1-11). Le résultat de la classification par SI est invariant à l'orientation, la rotation et la translation.

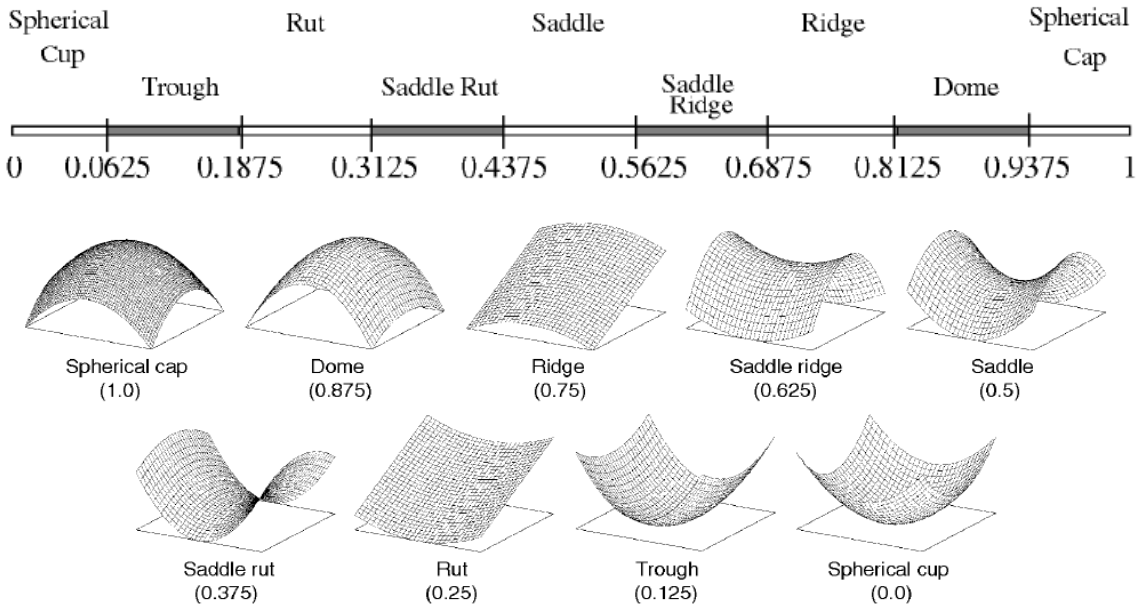


Figure 1-11. Positions des 9 types de surfaces par la classification sur l'échelle des valeurs de l'indice de forme SI (Dorai & Jain, 1997)

1.2.6 Intensité de courbure (C)

Un estimateur C de la magnitude de courbure est introduit également par Koenderink et Doorn (Koenderink & Doorn, 1992). C'est une mesure positive qui quantifie de combien est courbé le point (Équation 1.14). Un point sans courbure a une valeur nulle de C. Cette mesure est invariante à l'orientation, mais ne l'est pas pour l'échelle et la résolution.

$$C = \sqrt{\frac{(k_p^1)^2 + (k_p^2)^2}{2}}$$

Équation 1.14

1.2.7 Espace SC

Comme présenté précédemment, l'indice de forme S renseigne sur la forme et l'intensité de courbure C décide si la surface est localement plane ou pas. L'espace Indice de forme-Intensité de Courbure (SC) a été utilisé (Koenderink & Doorn, 1992) pour subdiviser les formes de surface en 9 types (sous forme de dôme, de cuvette, de selle ornière, de selle crête, etc). Par exemple, une valeur de SI entre 0.8125 et 0.9375 correspond à dôme (Tableau 1-2). Ce schéma a été utilisé aussi pour la représentation

de la forme 3D dans (Dorai & Jain, 1997) en segmentant la surface du modèle en des sous-régions selon leur typologie.

Convex (Elliptic) (1)	$S \in [+5/8,1] \cap C>C_{zero}$
Convex (Parabolic) (2)	$S \in [+3/8,+5/8] \cap C>C_{zero}$
Saddle Ridge (3)	$S \in [+3/16,+3/8] \cap C>C_{zero}$
Planar (5)	$C<C_{zero}$
Hyperbola (6)	$S \in [-3/16,+3/16] \cap C>C_{zero}$
Concave (Elliptic) (7)	$S \in [-1,-5/8] \cap C>C_{zero}$
Concave (Parabolic) (8)	$S \in [-5/8,-3/8] \cap C>C_{zero}$
Saddle Valley (6)	$S \in [-3/16,+3/16] \cap C>C_{zero}$

Tableau 1-2 Classification des surfaces selon la paire SC (Akagündüz, 2011)

1.2.8 Comparaison HK vs SC

La description de la courbure par le biais des représentations HK et SC a été comparée par Cantzler et Fisher (Cantzler, et al., 2001), en termes de pertinence de classification des surfaces et impact des seuils et des niveaux de bruit sur cette classification. La remarque est que, sur une représentation des espaces HK et SC sur le même plan (k_1, k_2), C s'annule si la surface est plane, K s'annule sur les surfaces paraboliques (cylindrical ridges et ruts) et H s'annule sur les surfaces en selle, pour des surfaces qui paraissent courbées pour l'œil humain. Leur conclusion consiste à dire que ces classifications dépendent fortement des seuils choisis et que les résultats sont comparables pour les deux représentations. Cependant, l'approche SC apparait meilleure pour des faibles seuils et dans le traitement du bruit et des scènes complexes. Dans le travail de (Li, et al., 2004), des histogrammes sur les valeurs des paires HK et SC sont calculés sur des images 2D. L'étude comparative menée montre que les histogrammes de SC sont meilleurs pour la classification. Cependant, la résolution et l'effet de l'échantillonnage des images sont ignorés dans cette comparaison. Plus récemment, dans une approche multi échelle, Bozkurt et al. (Bozkurt, et al., 2009) ont montré que l'espace HK donne des meilleurs résultats en terme de reconnaissance d'objets.

Le tableau suivant résume les points positifs et négatifs de chacun des deux espaces HK et SC.

	Points positifs	Points négatifs
HK	• Invariance à l'orientation, translation et rotation • Meilleure performance pour une approche multi-échelle	• Dépendance de deux seuils-zéro sur H et K • Pas d'invariance à l'échelle et à la résolution
SC	• Invariance à l'orientation, translation et rotation • Invariance de S à l'échelle et à la résolution • Meilleure classification pour des faibles seuils et dans le traitement du bruit et des scènes complexes	• Dépendance d'un seuil-zéro sur C

2. CHAPITRE : ETAT DE L'ART DES TECHNIQUES DE RECONNAISSANCE 3D

Dans cette section, nous classons les approches de reconnaissance d'objets 3D décrites dans la littérature. Cependant, cette liste n'a pas but d'être exhaustive. Afin de répondre à l'attente du sujet de notre thèse, nous détaillons plus particulièrement les approches basées sur les points d'intérêt.

Le schéma classique d'un processus de reconnaissance d'objets, présenté dans la Figure 2-1, distingue deux étapes: la première est l'apprentissage qui consiste à apprendre la description des objets à partir d'une base d'exemples (galerie) pour construire des classes, la seconde est la classification qui reconnait un objet inconnu (requête ou probe) à partir de sa description en lui assignant une classe.

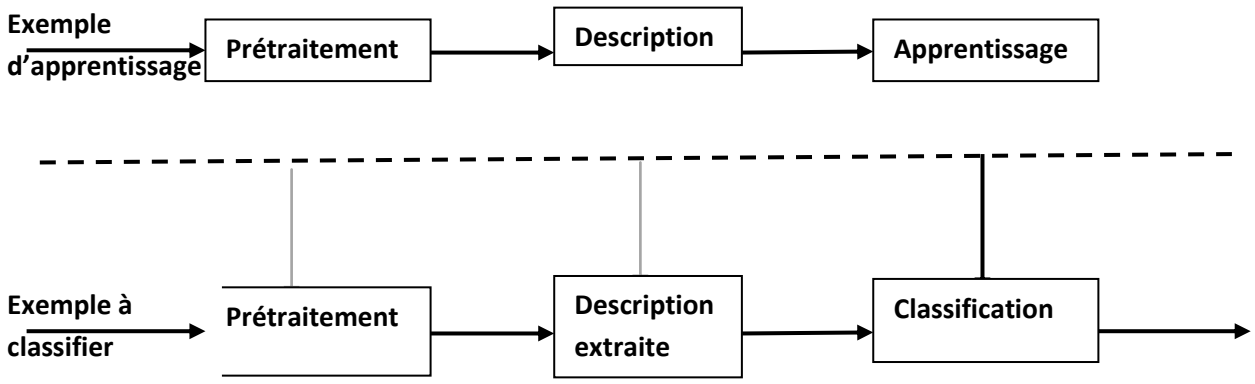


Figure 2-1. Etapes d'un système de reconnaissance

2.1. Typologie des approches de reconnaissance 3D

La détermination de la similitude entre les objets 3D consiste à définir une signature et une mesure de similarité. Outre la reconnaissance 3D, la recherche par le contenu ou « l'indexation 3D » s'intéresse à cette notion. L'analyse de la littérature des approches de ces deux domaines montre qu'il existe une variété de descripteurs qui dérivent de différents aspects des objets 3D. Nous avons choisi comme première répartition d'adopter celle évoquée par Zaharia et Prêteux (Zaharia, et al., 2002). Si certaines méthodes caractérisent les modèles 3D par leurs propriétés géométriques, d'autres se fondent sur les projections 2D des modèles. Ces projections 2D encodent une information tridimensionnelle. De ce fait, une distinction des descripteurs en deux familles est possible: famille des approches 3D et famille des approches 2D/3D. Une deuxième répartition, proposée par Campbell et al. (Campbell, et al., 2001), répertorie les méthodes de reconnaissance selon la nature de la représentation de la forme 3D et

considère : les approches locales et les approches globales (ou holistiques) (Roth, et al., 2008). Zhao et al. (Zhao, et al., 2003) rajoutent une troisième classe qui regroupe les méthodes caractérisant localement et globalement le modèle 3D.

Dans cette partie, nous présentons dans un premier temps quelques approches 2D/3D en les subdivisant en locales et globales et dans un deuxième temps les approches 3D avec la même sous-classification de localité et globalité.

2.1.1 Approches 2D/3D

Parti du constat que deux modèles 3D sont similaires s'ils ont le même aspect de la forme sous tous les angles de vues, le traitement direct de la forme 3D est remplacé par des projections 2D sous formes d'images de profondeur, des silhouettes ou encore de cartes de courbures (Diego, et al., 2010). En effet, les images de profondeur permettent de déduire les propriétés topologiques et structurales de l'objet et offrent ainsi une description riche de la forme 3D.

2.1.1.1 Approches 2D/3D globales

Les approches globales couvrent l'information de **tous** les pixels de l'image. L'idée est de projeter les données initiales dans un sous-espace qui représente les données de manière optimale selon un critère précis : par exemple, dans le cas de l'ACP (analyse en composantes principales) la variance est minimisée, dans l'ACI (Analyse en Composantes Indépendantes), l'indépendance des composantes est visée, etc. L'un des avantages de ces méthodes est qu'à partir des descripteurs calculés une reconstruction du modèle original est possible. Cependant, ces approches sont sensibles aux occultations, aux variations d'illumination et de l'échelle, et aux déformations géométriques locales (expressions du visage par exemple), ce qui nécessite une étape de normalisation lors du prétraitement. Souvent, cette normalisation automatique ou manuelle (choix de points références) détériore la performance de la reconnaissance.

• Chaouch et Verroust-Blondet (Chaouch, et al., 2007) proposent de représenter un modèle 3D par 20 images de profondeur. A chaque image est associé un descripteur de **lignes de profondeurs**. Ces lignes sont obtenues par une méthode de mise en séquences qui part de l'espace des points pour former une séquence d'éléments. Son principe est de coder une ligne de profondeur en N états d'une séquence d'observations. Chaque observation est représentée par un état parmi 5 défini en correspondance avec 5 types de région (fond intérieur, fond extérieur, profondeur croissante, profondeur décroissante, profondeur stable). Finalement, la distance de programmation dynamique (DPD) est utilisée pour le calcul de similarité entre les descripteurs des lignes de profondeurs.

• Les travaux de Vranic et Saupe (Vranic, et al., 2000) présentent le premier descripteur basé sur les images de profondeur, le « Depth Buffer-based Descriptor » (**DBD**). Pour assurer un comportement invariant géométriquement, chaque objet 3D est normalisé avec une ACP par rapport à un cube d'axes parallèles à ceux du repère intrinsèque à l'objet 3D. En projetant le modèle sur les six faces de ce cube, des images de profondeur sont calculées et transformées dans l'espace de Fourier en utilisant la 2D-FFT (Fast Fourier Transform). La signature de l'objet 3D est extraite en gardant les coefficients basse-fréquence pour chaque image de Fourier traitée.

• Les **Eigenspaces** utilisés par Murase et Naya (Murase, et al., 1995) et les **Eigenfaces** utilisés par (Tsalakanidou, et al., 2003) et par (Chang, et al., 2003) permettent un appariement holistique. Le

principe de cette technique est de projeter les images de profondeur sur un sous espace en appliquant une ACP (trouver les composantes principales). Les primitives correspondent au résultat du produit scalaire des images avec les premiers vecteurs propres obtenus (correspondant aux plus grandes valeurs propres). La signature est une somme pondérée de ces primitives. Une image test est projetée sur les vecteurs propres de la base. L'appariement utilise la classification avec une méthode de plus proche voisin (plus proche primitive selon un critère de distance).

• Dans le cadre de vérification des visages, Diego, et al (Diego, et al., 2010) calculent des **ondelettes de Gabor** sur trois types de représentations de données : des images de profondeurs, des images 2D et des images de courbures moyennes. Pour générer la carte des courbures de la surface de l'objet 3D, en chaque point, la moyenne des courbures minimales et maximales des polygones du maillage est calculée. Avec les informations des courbures, la surface du modèle 3D épouse alors la surface d'une forme élémentaire. Les filtres de Gabor peuvent être définis comme un plan d'ondes orientées dans une direction multipliée par une enveloppe gaussienne 2D. Les filtres de Gabor sont connus comme un moyen d'analyse espace-fréquence très robuste. Cette spécificité a fait des filtres de Gabor un puissant moyen d'analyse pour la classification. Ces filtres analysent l'information (profondeur, texture ou courbure) d'un objet suivant différentes résolutions et différents angles. Dans le domaine spatial, un filtre de Gabor 2D est une fonction à noyau gaussien modulé par une onde sinusoïdale plane complexe. La famille des filtres de Gabor est caractérisée par un certain nombre de résolutions, d'orientations et de fréquences, qui forment la signature de l'objet.

• **Courbes de niveau des profondeurs** : Samir, et al. (Samir, et al., 2006) proposent d'apparier deux surfaces faciales par la comparaison des courbes faciales. L'approximation grossière de la surface faciale S, par un ensemble fini de courbes de niveau fermées, génère les courbes de niveau d'une fonction de profondeur où les indexes correspondent aux valeurs de cette fonction (Figure 2-2). Les courbes de niveau des profondeurs de deux surfaces de visages sont comparées par un outil de l'analyse Riemannienne des formes qui est la distance géodésique. Après extraction des courbes faciales à partir des images de profondeurs, une fonction d'angle est calculée pour chaque courbe et permet d'estimer les longueurs géodésiques. Une métrique sur les formes faciales est déduite en cumulant les distances entre les courbes faciales.

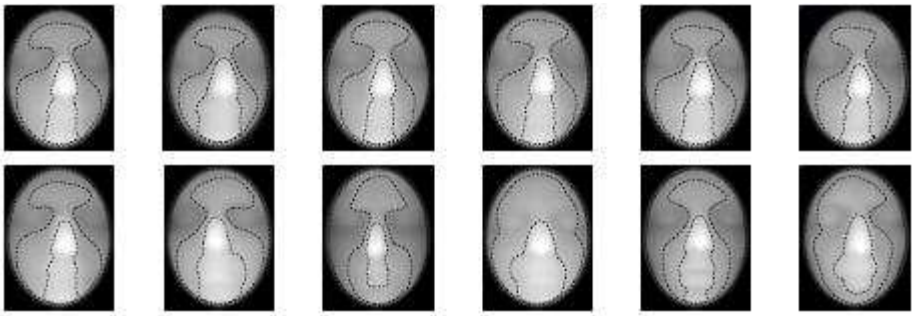


Figure 2-2. Représentation des courbes faciales. En haut : six expressions faciales de la même personne. En bas: la même expression faciale de six personnes différentes (Samir, et al., 2006)

2.1.1.2 Approches 2D/3D locales

L'une des approches les plus populaires pour représenter la surface est celle basée sur les descripteurs locaux ou une signature qui décrit la région locale. Les positions de la surface où sont estimés les

descripteurs locaux peuvent être simplement choisis d'une façon exhaustive ou aléatoire. Dans le cas d'une sélection exhaustive, la redondance dans des régions, ayant une légère variation de la forme, rend cette technique inefficace. La sélection aléatoire peut négliger certaines zones ayant une structure géométrique distincte ce qui réduit la performance de l'algorithme. Ainsi, extraire un ensemble représentatif de points primitives permet de remédier à ces problèmes. Plusieurs méthodes de reconnaissance d'objet, dont celles de Johnson et Hebert (Johnson, et al., 1999), et Frome, et al (Frome, et al., 2004)), ont suggéré d'utiliser un jeu de points choisis aléatoirement pour calculer les descripteurs de surface. D'autres ont tenté d'extraire efficacement des points. Dans le cas des approches locales qui détectent des régions saillantes, les invariants de l'objet peuvent être par exemple les coins, l'entropie, les niveaux de gris, les contours, la texture, la symétrie, etc. En reconnaissance de visages, des modèles peuvent être générés à partir des relations entre les formes géométriques ou la luminance des différentes parties de l'objet (le nez, les yeux, la bouche, le visage,…). A chaque région est attribué un descripteur de primitives. Ces primitives locales doivent répondre aux critères suivants:

– assez discriminantes pour différencier un objet d'un autre,
–reproductibles pour être détectables au même endroit de l'objet dans différentes scènes,
–riches en information caractérisant la topologie, la forme ou la texture locale l'objet.
–invariantes aux changements d'illumination, du bruit, d'échelle, de vues…
–robustes au: bruit, occultations partielles et variations intra-classes.

En général, une seule primitive ne permet pas de satisfaire tous ces critères, une combinaison de primitives est souvent utilisée. Les algorithmes d'appariement pour les primitives locales mettent en correspondance, soit directement les primitives, soit leurs distributions statistiques. Dans le cas particulier où la région saillante est réduite à un point, la notion de Point d'Intérêt (PI) est introduite. Dans ces approches, deux étapes sont nécessaires: d'abord, déterminer les régions de l'image qui doivent être comparées aux régions dans la seconde image (détection). Ensuite, générer des signatures invariantes pour comparer les régions sélectionnées (descripteurs). En 2D, les investigations pour caractériser des détecteurs et descripteurs locaux des PIs ne cessent de s'enrichir. Avec le succès de cette approche locale, plusieurs travaux cherchent à adapter à la tridimensionnalité les méthodes du 2D.
Nous listons ci-dessous quelques une des approches 2D/3D locales.

• Des primitives locales boostées sont introduites par Viola et Jones (Jones, et al., 2003) pour associer des régions rectangulaires des images de visages à différentes localisations, échelles et orientations. Une mesure de similarité consiste en une combinaison linéaire de primitives f_i qui sont des filtres appliqués à une paire d'images :

$$f_i(I_1, I_2) = \begin{cases} \alpha & si\ |\phi_i(I_1) - \phi_i(I_2)| > t_i \\ \beta & sinon \end{cases}$$

Le terme t_i est un seuil sur les primitives et ϕ_i une fonction scalaire qui représente un filtre rectangulaire. Dans la Figure 2-3, les filtres rectangulaires sont calculés en additionnant les intensités de tous les pixels dans les régions sombres et en soustrayant la somme des intensités de tous les pixels dans les régions claires. Les entrées de l'étape d'apprentissage sont constituées d'une grande base de **filtres rectangulaires** et un jeu d'exemples positifs et négatifs. L'algorithme d'Adaboost est utilisé pour trouver les meilleurs filtres (ϕ_i), les seuils (t_i) et les poids (α et β) pour créer un classificateur qui sépare les exemples positifs (des mêmes paires de visage) des exemples négatifs (des paires de visages différentes).

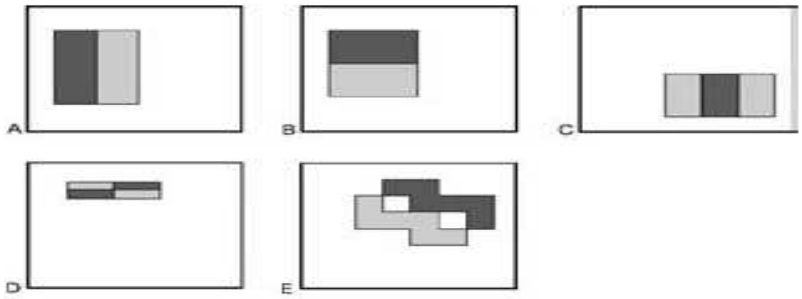
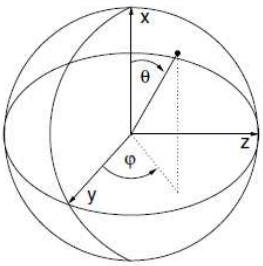


Figure 2-3. Exemples de filtres rectangulaires (Jones, et al., 2003)

•**L'histogramme de primitives locales** (local feature histogram) a été exploré dans le travail de Hetzel et al. (Hetzel, et al., 2001) pour reconnaitre des objets de forme libre sur des images de profondeurs. La signature a regroupé la **profondeur** des pixels, les **normales** de la surface er les **courbures**. La combinaison de ces primitives dans un histogramme multidimensionnel offre un cadre probabiliste permettant de gérer entre autres le problème de l'occultation. L'avantage de l'histogramme des distances est son invariance à la translation, aux rotations et à l'échelle si la distance est normalisée. Par contre, les distances normalisées sont sensibles aux intervalles de la profondeur perçue et au bruit du fond. Une paire d'angles des normales définies en coordonnées sphériques est considérée pour une représentation 2D des normales (Figure 2-4). L'indice de formes (défini dans le chapitre 1) est utilisé pour représenter les courbures de la surface. Les courbures peuvent être calculées directement par les dérivées premières et secondes, ou indirectement comme le taux de changement des orientations des normales dans une région locale. La robustesse aux angles de vues et la discrimination de l'information sont prouvés par ces primitives. La reconnaissance est établie par un appariement direct des histogrammes ou par une reconnaissance probabiliste.



$$\phi = \arctan\left(\frac{n_z}{n_y}\right), \theta = \arctan\frac{\sqrt{(n_y^2 + n_z^2)}}{n_x}$$

Figure 2-4. Représentation des normales en coordonnées sphériques (Hetzel, et al., 2001)

•**Une approche hybride :** (Mian, et al., 2006) proposent un algorithme de reconnaissance hybride: locale et globale. La localité consiste à segmenter le visage en trois régions qui sont appariées avec un algorithme d'ICP (Iterative Closest Point) modifié. L'ICP est un algorithme utilisé pour l'alignement en minimisant une distance quadratique carrée entre les points les plus proches dans un processus itératif. Pour l'appariement holistique un algorithme basé sur l'ACP est utilisé.

2.1.2 Approches 3D

Les données considérées pour les approches 3D se présentent sous formes de nuage de points ou sous forme de représentation polygonale (maillage). Ces approches tentent de caractériser la forme 3D avec une description indépendante de la topologie 3D. En effet, la topologie d'une forme peut varier sous l'effet : des déformations internes (expressions du visage) ou externes (bruit et occultations), ou de la façon de modéliser la forme comme la résolution du nuage de points (nombre de points), ou le nombre des facettes (tessellation du maillage). Ces techniques peuvent être classées selon leur portée en locale ou globale.

2.1.2.1 Approches 3D globales

Nous pouvons subdiviser les méthodes holistiques 3D en deux catégories: approches surfaciques et approches volumiques.

• **Distribution de forme :** Cette signature proposée par (Osada, et al., 2001) pour un modèle 3D polygonal se base sur une fonction de forme qui mesure les propriétés géométriques globales d'un objet. L'apport principal de cette méthode est, en premier lieu, le calcul de la signature avec la spécificité aléatoire de l'échantillonnage permettant de construire une distribution continue de probabilité. En deuxième lieu, une simple comparaison des distributions de probabilité est requise contrairement aux méthodes qui font appel à une reconstruction d'objets solides, une mise en correspondance de primitives ou un ajustement des modèles. Des primitives simples à mesurer sont considérées: la distance entre le centre de gravité et un point de la surface, la distance entre deux points, l'angle entre trois points, l'aire au carré du triangle formé par trois points et le volume au carré du tétraèdre formé par quatre points. Les points concernés par le calcul des primitives sont choisis arbitrairement sur les facettes, conférant à cette méthode une invariance aux modifications du nombre de facettes. De plus, l'échantillonnage pondéré et la normalisation permettent d'assurer une robustesse au bruit et une invariance par rapport aux changements d'échelle. Les modèles sont comparés en évaluant des mesures de similarité avec la norme L_N, des fonctions de densité de probabilité (pdfs) et des fonctions de distribution cumulative (cdfs).

• Les **EGI** (Image Gaussienne Etendue, Extended Gaussian Images) (Horn, 1984) caractérisent une fonction synthétisant l'information d'orientation en tout point de la surface de l'objet 3D (Figure 2-5). Un histogramme d'orientation, en tant qu'approximation discrète des EGI, est calculé sur la sphère unité de Gauss. En effet, l'information des normales de la surface peut être mappée sur la sphère unité de Gauss subdivisée en cellules. Le mapping utilisé consiste à associer l'inverse de la courbure gaussienne de chaque point de la surface de l'objet avec le point correspondant ayant la même normale dans la sphère gaussienne. Ce mapping est réversible si l'objet a partout une courbure gaussienne positive. Si u et v sont les coordonnées du point dans la surface originale, ξ et μ les paramètres du point dans la sphère de Gauss (peuvent correspondre par exemple à la longitude et latitude), l'EGI est définie alors par : $G(\xi,\mu) = \frac{1}{k(u,v)}$ avec k la courbure gaussienne qui est égale au produit de la courbure maximale et la courbure minimale:

$$k = \kappa_1\kappa_2;\ \kappa_N(\theta) = \kappa_1 cos^2\theta + \kappa_2 sin^2\theta.$$

Pour une sphère de rayon R, l'EGI est $G(\xi,\mu) = R^2$.

Dans les versions plus élaborées, cette fonction est pondérée par l'aire des facettes ou leur distance à l'origine d'un repère préalablement spécifié. L'EGI est invariant par translation et par rotation. La faiblesse majeure de cette approche est que plusieurs objets convexes peuvent avoir le même EGI. L'appariement entre deux objets revient à mettre en correspondance les deux histogrammes des EGI par le calcul de la distance Euclidienne.

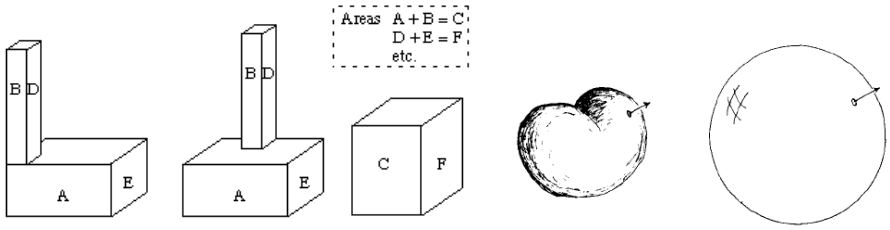


Figure 2-5. A gauche-Exemples d'objets avec le même EGI-A droite- un objet avec la sphère de Gauss (Horn, 1984)

•Une approche qui s'inspire de la théorie de la reconnaissance par composantes (Recognition By Components, RBC) proposée par (Biederman, 1987) consiste à identifier à l'aide d'un appariement de superquadriques des primitives de forme appelées géons. Ces **géons** décrivent qualitativement des volumes élémentaires (sphères, cylindres, cônes, etc.) d'un corps. A ce titre, ils permettent une identification sommaire mais rapide des objets. (Medioni, et al., 2000) proposent une modélisation par sous-parties volumiques des objets 3D à base de géons. Après une décomposition de l'objet 3D en géons, ces éléments sont hiérarchiquement organisés dans un arbre de description avec des relations d'adjacence. Cette structure hiérarchique permet en outre de calculer plus efficacement une mesure de similarité fondée sur un coût de transition issu d'une méthode d'appariement par graphes. Les images à attribut sphérique (Spherical Attribut Images : **SAI**) (Hebert, et al., 1995) permettent également de reconnaitre des objets de topologie proche de la sphère.

• Le **SF3D** (descripteur par spectre de forme 3D) présentés par Zaharia et Préteux dans (Zaharia, et al., 2002). Le SF3D est un histogramme de l'indice de forme qui fournit une représentation intrinsèque des caractéristiques géométriques locales de la surface 3D. Cet indice est défini par la valeur de la coordonnée angulaire de la représentation polaire du vecteur de courbures principales. Invariant aux transformations euclidiennes et aux homothéties, ce descripteur n'est cependant pas robuste aux représentations topologiques.

• La transformée de Fourier proposée par (Vranic, et al., 2001), permet de caractériser l'information spatiale des objets 3D dans l'espace fréquentiel. L'extraction des primitives se fait après une étape de normalisation en utilisant une ACP continue et une voxelisation du modèle 3D. Les composantes du vecteur primitives sont les valeurs absolues des coefficients de la transformée de fourrier discrète 3D (**3D DFT**) appliquée aux voxels. Ces descripteurs sont invariants à la translation, la rotation, le changement d'échelle et la réflexion.

• Le **DH3DO**, introduit par (Zaharia, et al., 2002) est un descripteur dérivé de la transformée de Hough 3D. Son principe consiste à accumuler des points sur des plans de R^3. En effet, un plan dans l'espace euclidien peut être transformé, dans l'espace de Hough, en un seul point défini par les coordonnées sphériques (r, θ, φ). Par conséquent, une collection de plans dans l'espace euclidien correspond à une collection de points dans l'espace de Hough. Le DH3DO a été proposé pour optimiser le DH3D suite aux faiblesses que présentaient le SF3D. Il permet, en effet, d'assurer une invariance aux transformations géométriques et une robustesse vis-à-vis des représentations topologiques multiples.

2.1.2.2 Approches 3D locales

Kazhdan et Funkhouser (Kazhdan, et al., 2002) proposent une description locale de la forme 3D par un clustering de la masse sur des sphères concentriques. Des composantes de fonctions définies par l'intersection de la surface du modèle 3D avec un ensemble de sphères concentriques sont appelées **harmoniques sphériques**. Cette représentation, invariante aux rotations, se présente sous forme d'une grille 2D où la valeur d'une case (i, j) est la norme de la composante j de la fonction sphérique dans la sphère de rayon i. La comparaison de deux représentations harmoniques se fait par le calcul de la distance euclidienne. La Figure 2-6 illustre les différentes étapes du processus du calcul des harmoniques sphériques.

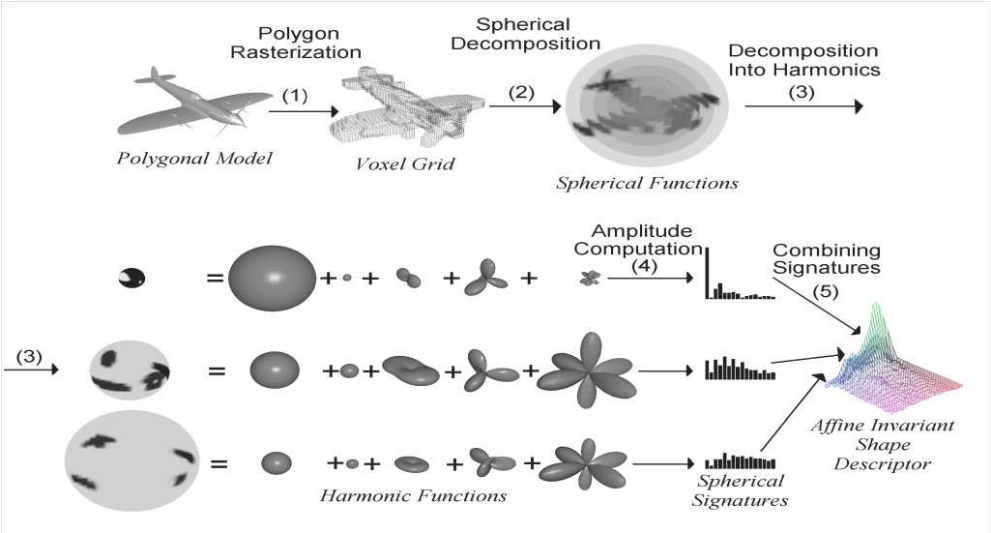


Figure 2-6. Etapes de calcul des harmoniques sphériques (Kazhdan, et al., 2002)

•Le **Spin image**, introduit par (Johnson, et al., 1999) sur un maillage de données, a une efficacité confirmée dans la reconnaissance d'objets pour des scènes complexes. L'appellation « Image » vient de la représentation 2D des zones et « Spin » s'inspire du processus de génération de l'image qui est similaire à une filature « sheet spinning » sur les normales. Ce descripteur est défini sur une zone locale (patch) cylindrique et se base sur la notion de point orienté (déterminé par sa position 3D et sa normale). La normale d'un sommet d'une facette est calculée par l'ajustement d'un plan aux points qui lui sont connectés dans la facette. Chaque point (sommet) est caractérisé par ses deux coordonnées cylindriques (coordonnée radiale et coordonnée d'élévation). Un histogramme 2D cumule les valeurs des paires de coordonnées. Une réduction de la dimension de l'ensemble des spins est réalisée à l'aide d'une ACP pour former un EigenSpin Images stocké dans la galerie. La distance l_2 est utilisée pour comparer les spins lors de l'appariement.

•**3D Shape Context** : Le principe de ce descripteur, proposé par Frome et al. (Frome, et al., 2004), est similaire au spin image à la différence du support qui est sphérique. La sphère est segmentée en des sous volumes en divisant, uniformément sur les dimensions d'élévation et d'azimut, et en logarithmique sur la dimension radiale. Au niveau de l'histogramme, cette méthode considère que la densité varie indépendamment entre les bins, et normalise chaque bin par l'inverse de la densité des points et le volume du bin. Un degré de liberté sur la direction azimut doit être supprimé lors de la mise en correspondance des descripteurs.

• Un histogramme géométrique de paires (**Pairwise Geometric Histogram** (**PGH**)): (Ashbrook, et al., 1998) approximent les surfaces des modèles par un maillage triangulaire. Chaque facette t_i est représentée par un histogramme 2D, h_i, qui décrit la relation entre cette dernière et les facettes voisines. Les entrées de l'histogramme sont constituées par l'angle entre les normales des facettes et un intervalle de distances perpendiculaires entre le plan de la facette t_i et celui de tous les points de la facette t_j (Figure 2-7). L'histogramme cumule ces mesures pondérées par le produit des aires des deux facettes. La mise en correspondance entre les facettes est effectuée entre leurs histogrammes respectifs par la mesure de la métrique Bahattacharayya donnée par :

$$D_{ij} = \sum_{\alpha,d} \sqrt{h_i(\alpha,d)}\sqrt{h_j(\alpha,d)}$$

Équation 2.1

Ces histogrammes sont robustes aux transformations rigides et stables vis-à-vis du bruit et de l'omission de données de la surface.

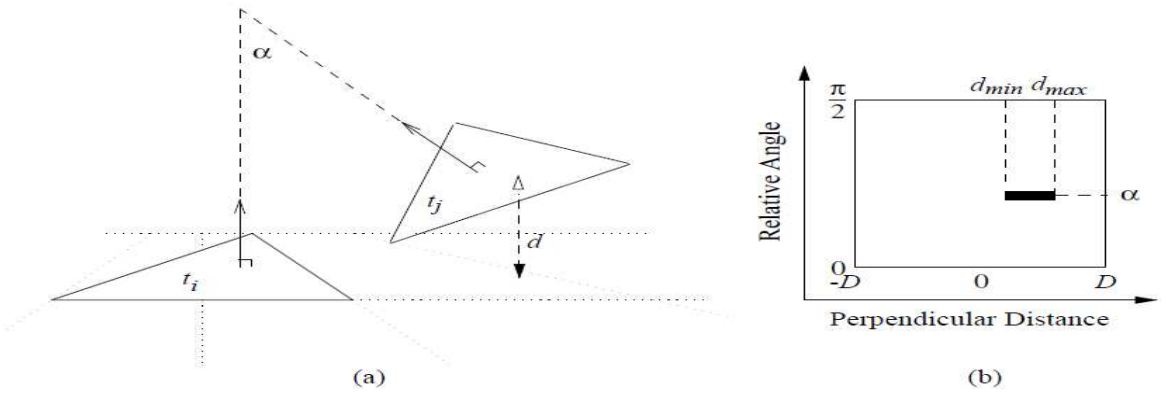


Figure 2-7. (a)- Les mesures géométriques caractérisant la relation entre deux facettes (b)- L'histogramme de paire géométrique représentant cette relation

• Swadzba et Wachsmuth (Swadzba, et al., 2008) proposent une classification basée sur les caractéristiques 3D pour des pièces scannées avec une caméra temps de vol. Ils extraient des structures plates (des patches) en combinant la méthode de croissance de région et l'approche RANSAC (Random Sample Consensus) (Fischler, 1981). Les vecteurs de primitives sont définis par des **histogrammes** cumulant les tailles relatives des zones plates, les **angles** entre les normales des zones voisines, et les rapports entre les tailles des paires des zones. Trois classificateurs sont utilisés: un réseau de neurones avec une couche cachée, un séparateur à vaste marge (SVM) avec un noyau polynomial et un modèle de mixture de gaussiennes (GMM).

• Hilaga, et al. (Hilaga, et al., 2001) suggèrent une approche topologique par comparaison des graphes de Reeb multirésolution (**MRGs**). Un MRG représente la structure de squelette de la forme 3D à différent niveaux de résolutions. Une fonction continue qui calcule la distance géodésique sur la forme 3D est utilisée pour construire les nœuds des graphes. L'objet est partitionné en régions en utilisant cette fonction. Chaque nœud du graphe est une composante connexe dans une région. Cette représentation est invariante par translation, rotations et tesselation. La similarité entre deux formes est estimée en effectuant une mise en correspondance des nœuds des diagrammes de Reeb. En partant des

diagrammes les plus grossiers, la mise en correspondance est propagée jusqu'aux diagrammes les plus fins. La Figure 2-8 illustre un exemple de graphe construit par la fonction de hauteur.

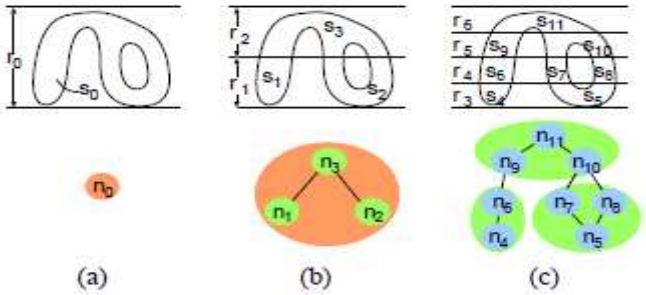


Figure 2-8. Un graphe Reeb Multi-résolution utilisant la fonction de hauteur (Hilaga, et al., 2001)

• **Détection des lignes crêtes et ravines**

Dans l'article (Song, et al., 2005), Song et al. ont proposé une approche rapide, flexible, robuste au bruit et invariante aux transformations géométriques. Les lignes crêtes et ravines sont extraites en utilisant une fonction paramétrique lissée. L'approche adoptée est fondée sur l'ajustement local d'une surface polynomiale bivariée en chaque sommet du maillage original M. Pour chaque sommet p, le voisinage du premier ring (Figure 2-9) est utilisé pour formé un nouveau point *p'* (barycentre des voisins). Un nouveau maillage *M'*, plus lisse que le premier maillage M, est formé. Pour chaque sommet *p'*, est estimée une normale avec la méthode NelsonMax. Par la suite, un polynôme cubique local h (Équation 2.2) est ajusté à *p'* et à son voisinage obtenu à partir de k ring. Les dérivées de ce polynôme donnent les courbures qui sont attribuées au point p du maillage M. Les minimas et maxima des courbures sont détectés par une approximation de différence finie et forment les lignes crêtes et ravines (Figure 2-10).

$$h(x,y) = \frac{1}{2}\,(b_0x^2 + 2b_1xy + b_2\mathrm{y}^2) + \frac{1}{6}\,(c_0x^3 + 3\,c_1\mathrm{x}^2y + 3\,c_2xy^2 + c_3\mathrm{y}^3)$$

Équation 2.2

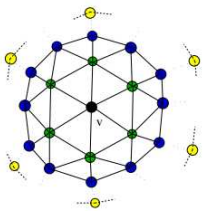


Figure 2-9. Point v (en noir) et ses anneaux de voisinage. Le premier anneau est composé par les sommets verts. L'anneau 2 est composé par les sommets verts et bleus

Figure 2-10. Extraction des crêtes (en bleu) et des ravines (rouge) sur différents modèles (Song, et al., 2005)

• Li et Guskov (Li, et al., 2005) détectent un jeu de **points saillants** en construisant une représentation multi-échelle de la surface 3D. L'ensemble des primitives est l'ensemble des maxima et minima de la **différence des normales** entres deux niveaux d'échelle voisins. Ces primitives sont robustes aux variations surfaciques. L'appariement s'effectue entre vues 2,5D, ou entre une vue 2,5D et un modèle 3D reconstruit. Ils organisent la signature du modèle 3D dans un arbre KD-Tree pour faciliter la recherche du plus proche voisin.

• Une technique de détection de points d'intérêt en se basant sur la variation de la forme des visages 3D avec une **représentation en tenseur** est présentée par Mian et al. (Mian, et al., 2008). Elle se base sur la quantification d'une surface locale ajustée au voisinage d'un point en une grille 3D définie dans une base de coordonnées locale. La définition de cette base utilise le point, sa normale et les directions principales de la surface de son patch. Le descripteur est l'ensemble des points de la grille. Pour réduire la dimension du descripteur, une ACP est appliquée. En appariement, deux graphes sont formés par projection des primitives du modèle probe et du modèle galerie avec une ACP. La similarité entre deux visages revient à calculer la similarité entre les deux graphes obtenus. Dans (Mian, et al., 2006), Mian et al. ont comparé deux méthodes pour le calcul de la base de coordonnées locale : la première utilise la normale du point et la localisation du nez et la deuxième utilise la normale du point et l'orientation du descripteur SIFT calculée sur l'image 2D. La première technique a été retenue étant donné sa stabilité. Chaque point p et son voisinage l sont transformés dans les coordonnées de la nouvelle base. Le patch est translaté et mis en rotation pour qu'il soit aligné avec la base locale et que le point p soit l'origine. Le patch est ré-échantillonné en une nouvelle grille 3D uniforme. Cet échantillonnage forme le descripteur local au point p. Ce descripteur donne des performances meilleures aux spins images dans des scènes complexes.

2.1.3 Conclusions et remarques

Nous présentons un tableau récapitulatif des approches décrites et deux tableaux comparatifs des apports et lacunes des approches 3D versus les 2D/3D et de les approches globales versus les locales:

Approches 3D			**Approches 2D/3D**	
Globales		**Locales**	**Globales**	**Locales**
Surfacique	**volumique**			
EGI (Horn, 1984) Distribution de forme (distance, angle, taille) (Osada, et al., 2001) SF3D (Zaharia, et al., 2002) 3D DFT (Vranic, et al., 2001) DH3DO (Zaharia, et al., 2002)	Géons (Medioni, et al., 2000) SAI (Hebert, et al., 1995)	Harmoniques sphériques (Kazhdan, et al., 2002) RANSAC + Histogramme des normales et des tailles des patches (Swadzba, et al., 2008) Spin Images (Johnson, et al., 1999) 3D Shape Context (Frome, et al., 2004) Pairwise Geometric Histogram (PGH) (Ashbrook, et al., 1998) MRG (Hilaga, et al., 2001) Différence des normales sur points saillants (Li, et al., 2005) Point saillant et représentation en tenseur (Mian, et al., 2008) Détection des lignes crêtes et ravines (Song, et al., 2005)	Depth Buffer-based Descriptor (Vranic, et al., 2000) Lignes de profondeur (Chaouch, et al., 2007) Courbes de niveau des profondeurs (Samir, et al., 2006) Eigenfaces (Tsalakanidou, et al., 2003) (Chang, et al., 2003) Ondelettes de Gabor (Diego, et al., 2010)	Filtre rectangulaire de Viola et Jones (Jones, et al., 2003) Représentation en tenseur (Mian, et al., 2006) Histogramme de profondeurs, normales, et courbures (Hetzel, et al., 2001)

Tableau 2-1. Classification des approches présentées

	Avantages	Inconvénients
Approches 3D	- Souligne la structure tridimensionnelle -Invariance aux transformations géométriques (transformations euclidiennes et homothéties) et aux conditions d'acquisition (éclairage)	-Complexité et temps de calcul (exemple : génération du maillage)
Approches 2D/3D	-Représentation compacte et réduction de la complexité en temps de calcul -Profiter des approches existantes du domaine 2D -Représentation explicite de la forme 3D à travers la profondeur	-Sensibilité à la variation de la pose (rotation) -Possibilité de perte d'information (par cause de la structure 2D) (Mayo, et al., 2009)

Tableau 2-2. Résumé des avantages et inconvénients des approches 3D et des approches 2D/3D

- Pour conclure, le 3D offre de décrire la variation de la forme 3D tout en surpassant les limitations dues aux angles de vues et aux variations de l'éclairage. Avec le succès qu'a connu l'utilisation du 3D dans le domaine de la reconnaissance des visages plusieurs personnes se sont penchées sur la multi-modalité 3D-2D. Dans (Bowyer, et al., 2004), une revue des méthodes 3D et multimodales 3D+2D est présentée. Tsalakanidou, et al. (Tsalakanidou, et al., 2003) utilisent la multi-modalité 3D avec les images de couleur qui a permis d'attendre un score de bonne reconnaissance de 99% sur une base de 40 personnes. Chang, et al. (Chang, et al., 2003) utilisent les images 3D et 2D avec une approche basée sur l'ACP et montre que la multi-modalité donne un meilleur score que la mono-modalité (98,8% pour le rank1). Dans le travail de (Mian, et al., 2006), les descripteurs SIFT 2D et les primitives locales 3D inspiré de la représentation en tenseur sont fusionnés en utilisant la règle de la somme pondérée des coefficients de confiance et un score d'identification de 95,5% dans la base FRGC v2 est atteint.
- Cette fusion des données peut être approchée sur deux niveaux : fusion des primitives (pour former un seul vecteur) ou fusion des scores des classificateurs. Diego et al. propose, dans (Diego, et al., 2010), une approche qui fusionne 3 représentations: images de texture 2D, 2.5 images et images de courbures moyennes. Les matrices de noyaux linéaire ou gaussien calculées sur les trois types de données sont fusionnées pour former un seul noyau avec la méthode de combinaison PWS (Probability Weighting Scheme). Le noyau résultant est entrainé dans un SVM (Support Vector Machine).

	Avantages	Inconvénients
Approche globale	-Permettre la reconstruction de la forme 3D originale (complétude) -Description de toute la structure de l'objet (sans omission de parties) -Adéquate pour des structures rigides et proprement segmentées et pour des objets ayant une haute convexité	-Sensibilité aux occultations, au bruit, à l'échelle et aux déformations locales du mouvement (expressions du visage) -Nécessité d'une étape de normalisation
Approche locale	-Robustesse aux occultations -Invariance à l'échelle et aux angles de vue (dans certaines approches) -Réduction du nombre de données traitées (lors de l'appariement) -Moins sensible au bruit qu'une approche globale	-Limitation à des parties de la surface pouvant pénaliser des zones significatives de la forme -Choix d'un critère saillant et d'un voisinage (un patch)

Tableau 2-3. Résumé des avantages et inconvénients des approches globales et des approches locales

- Les surfaces peuvent être définies dans le système des coordonnées du capteur ou dans le système de coordonnée de l'objet. Dans le premier système, la description de la surface varie quand l'angle de vue change et l'appariement entre des surfaces doit s'accompagner par leur alignement. En outre, pour représenter une surface à partir de vues multiples, une représentation séparée doit être stockée pour chaque angle de vue différent. Un système de coordonnée relatif à d'objet décrit la surface de l'objet indépendamment de l'angle de vue, une seule représentation de la surface décrit toutes les vues. Les scènes réelles sont en général complexes avec des objets multiples. Le problème qui se pose dans ce cas est le faux appariement des surfaces qui ne font pas partie de l'objet. Générer un système de coordonnées propre à l'objet devient alors difficile. Les primitives locales de la scène permettant de réduire cet effet.
- La forme d'un objet 3D est intrinsèquement indépendante de sa taille et de sa position spatiale. Un bon descripteur, mis à part sa propriété discriminante, doit être invariant aux transformations géométriques (transformations euclidiennes et homothéties) et robuste aux représentations topologiques multiples (bruit, artéfacts).
- Les modèles 3D se situent dans des positions, orientations et échelles arbitraires dans l'espace 3D. Une étape de normalisation est souvent requise pour satisfaire les propriétés d'invariances géométriques et topologiques de l'objet 3D. Elle consiste à trouver le repère de coordonnées canoniques (une position, orientation et échelle canoniques). L'analyse en composantes principales (ACP) et le remaillage adaptatif sont les processus les plus généralement utilisés pour remédier à ces problèmes. Pour l'appariement, selon l'utilisation ou pas de la normalisation, deux catégories

de méthodes apparaissent (Kazhdan, et al., 2002). Celles ayant recours à des méthodes de normalisation pour aligner les modèles et définir par la suite une représentation de la forme selon cette orientation comme le cas de l'EGI, la transformée de Hough 3D, l'ACP (Chang, et al., 2003). La deuxième classe regroupe les méthodes invariantes à la rotation et donc n'ayant pas d'étape de normalisation par rapport à la position. Pour citer des méthodes parmi d'autres dans cette catégorie, les harmoniques sphériques (Kazhdan, et al., 2002) et la distribution de forme (Osada, et al., 2001) en font partie.

- Il est clair que le choix de la représentation de la forme d'un objet est problématique et dépende de l'attente finale du système de reconnaissance. Entre une représentation de zones planaires qui fournit beaucoup de primitives invariantes aux occultations et à l'échelle et qui requière peu de paramètres, et à l'inverse une représentation avec des formes génériques donnant peu de primitives et demandant beaucoup de paramètres (Johnson, et al., 1999), la nature globale ou locale de la représentation constitue un axe fort de comparaison de ces méthodes. Par conséquent, des systèmes de reconnaissance adoptent une représentation géométrique combinant les deux représentations (locale et globale) comme les spins images. Un autre compromis apparait entre la capacité du stockage exigé, la complexité informatique et la performance de la reconnaissance. En effet, les représentations de l'objet doivent permettre un appariement rapide pour que la reconnaissance se fasse en temps réel. En outre, la représentation doit être compacte pour avoir le plus possible de modèles dans la base.

  - ✓ La technique locale utilisant les points d'intérêt 2D a pu faire profiter le monde de la reconnaissance 3D. Nous enchaînons dans la partie suivante avec la description de ces approches et leur correspondance en 3D.

2.2. Détection et description des points d'intérêt en 2D

Comme nous l'avons mentionné précédemment, l'approche 3D locale avec points d'intérêt s'inspire des techniques 2D pour définir un détecteur et un descripteur 3D. Dans cette partie, nous exposons les approches utilisées pour le cas 2D, puis dans la partie suivante les techniques utilisées pour le cas 3D.

Rappelons qu'une région d'intérêt est une partie de l'image représentant une propriété intéressante. Le point est un cas particulier des régions d'intérêt. Héritant du caractère local, les points d'intérêts permettent essentiellement de remédier aux occultations de l'image. Parti du fait que dans l'image plusieurs points ont des caractéristiques plus significatives que d'autres, H. Moravec (Moravec, 1977) a introduit la notion de points d'intérêts depuis 1977. Son idée consistait à utiliser la fonction d'auto-corrélation afin de déterminer la meilleure position du point saillant, de façon à ce que toute position voisine contienne moins d'informations.

La problématique essentielle en détection est de trouver des points d'intérêts (PI, appelés aussi keypoints) d'une façon automatisée et suffisamment robustes pour l'appariement. Pour ce faire, un critère primordial à respecter est la répétitivité des points sous plusieurs conditions de prise de vue. Après extraction, un voisinage est défini autour de chaque point et est représenté par un descripteur (vecteur de primitives). Ce descripteur doit être discriminant et robuste au bruit, aux erreurs de détection et aux déformations géométriques et photométriques. La dimension du descripteur est importante d'autant son influence sur le temps de calcul. En dernière étape, les représentations locales des modèles tests sont comparées avec celles de la base d'apprentissage en se basant sur une mesure

de similarité (ex distance Euclidienne entre descripteurs).Nous abordons ces étapes en détails dans ce qui suit.

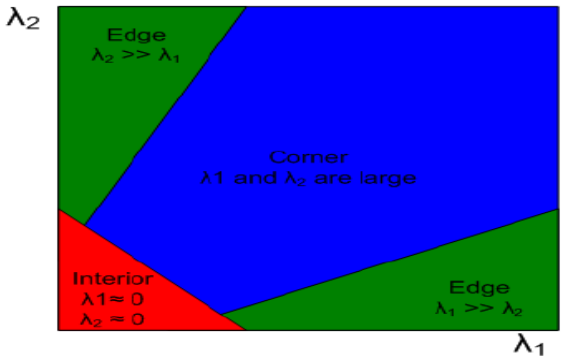
2.2.1 Détection

Les travaux sur la détection ont été initiés par le détecteur de coins et de contours de Harris. Dans la catégorie de détecteur de blobs, le SIFT introduit par Lowe et Herbert (1999) est invariant au changement d'échelle. Les SURFs proposés par Tuytelaars et Van Gool (2006) sont un bon compromis entre robustesse au changement d'échelle et temps de calcul (Roth, et al., 2008). Nous présentons les deux types de détecteur connu : détecteurs de coins et détecteurs de blobs (ou région).

❖ **Détecteurs de coins**

Ces détecteurs localisent les points et régions d'intérêt qui présentent une importante structure dans l'image et donc ne sont pas convenables pour la détection des régions uniformes. Les primitives des coins ont été utilisées dans plusieurs travaux pour obtenir une robustesse et une stabilité pour la reconnaissance d'objets.

Détecteur basés sur la matrice de Harris : est défini par Harris et Stephens (Harris, et al., 1988). En se basant sur le calcul de la fonction d'auto-corrélation, Harris et Stephens se ramenaient à l'étude de valeurs propres de la matrice de Harris (matrice de moments de second ordre). Sachant que les valeurs propres de cette matrice de Harris représentent les courbures principales de la fonction d'auto-corrélation, nous avons trois cas qui se présentent :



(a) région homogène, si les deux valeurs propres sont faibles,

(b) un contour (transition) si l'une des valeurs propres est très grande par rapport à l'autre

(c) un coin (point d'intérêt), si les deux valeurs propres sont élevées.

Figure 2-11. Distributions des types de région en fonction des deux plus grandes valeurs propres de la matrice de Harris (Harris, et al., 1988).

Cet opérateur permet de ne garder que les structures en coins, c'est à dire de courbure suffisamment grande. Malgré leur taux élevé de répétabilité, ces détecteurs restent sensibles au bruit et au changement d'échelle. Dufournaud et al (Dufournaud, et al., 2000) ont proposé une version multi-échelle de l'opérateur de Harris dans laquelle les points d'intérêt correspondent aux maximas locaux du détecteur de Harris appliqué à plusieurs échelles.

Détecteur basés sur la matrice de Harris-Laplacien : Le détecteur de Harris-Laplace (Mikolajczyk, et al., 2004) détecte les points en utilisant la fonction de Harris adaptée à l'échelle. Par la suite, les points maxima du Laplacien de gaussien sont choisis à une échelle caractéristique. Le Laplacien correspond à la trace de la matrice de Harris. Ces détecteurs sont invariants à l'échelle et aux transformations affines.

❖ **Détecteur de Blobs**

Ces détecteurs permettent de remédier aux lacunes du détecteur des coins, concernant les zones uniformes et les régions sans transitions apparentes (lissées). En effet, ce détecteur extrait les blobs de l'image qui sont des zones homogènes en termes d'intensité.

Détecteur basés sur la matrice Hessienne : se base sur un principe similaire à celui des détecteurs de Harris. Le critère de sélection dépend du déterminant de la matrice Hessienne constituée des dérivées secondes de l'image. Ces dérivées ont permis d'avoir une forte réponse aux niveaux des blobs et des contours et une invariance à la rotation.

Détecteur de Régions Saillantes Basé sur l'Entropie **(**EBSR**):** Kadir et Brady introduisent le détecteur de régions saillantes qui s'appuie sur la fonction de densité de probabilité (pdf) des valeurs d'intensité. La pdf de l'entropie est estimée par les valeurs de l'histogramme des niveaux de gris dans un patch circulaire à une échelle et une position précise. Ce détecteur est invariant au changement d'échelle et à la rotation. Un autre travail de Kadir (Kadir, et al., 2004) propose une extension de cet algorithme pour assurer l'invariance affine. Il considère une recherche exhaustive sur les déformations elliptiques du patch qui reste très coûteuse en temps de calcul.

Le SIFT (Scale Invariant Features Transform) : Une approximation du Laplacien par la fonction de différence de gaussiennes (DoG) qui calcule la différence des images convoluées avec des noyaux gaussiens à plusieurs échelles est utilisée par David Lowe (Lowe, 2004) pour introduire le SIFT. Nous présentons dans ce qui suit les étapes suivies par le détecteur SIFT pour générer l'ensemble des PIs :

1- Détection des extrema de l'échelle de l'espace: utilise la fonction DOG pour identifier les points d'intérêt qui sont invariants par rapport à l'échelle et à l'orientation.
2- Localisation des PIs: En chaque position candidate, adapter un modèle détaillé pour déterminer la position et l'échelle. Les PIs sont sélectionnés en se basant sur la mesure de leur stabilité.
3- Attribution d'orientations : Afin d'obtenir une représentation invariante par rotation, il est d'usage d'attribuer à chaque point d'intérêt des orientations principales, extraites de l'histogramme circulaire d'orientation des gradients au voisinage du point. Une ou plusieurs orientations sont attribué à chaque position du PI en calculant la direction du gradient de l'image locale. Les données de l'image sont transformées relativement à l'orientation, l'échelle et la position de chaque primitives pour assurer une invariance à ces transformations.

   Dans le cas de la reconnaissance d'objets, les estimations locales de l'échelle permettent de manipuler différents dimensions d'objet. Ce détecteur est reconnu par son invariance à la translation, rotation et échelle de l'image et son invariance partielle aux changements d'illumination et aux déformations locales de l'image.

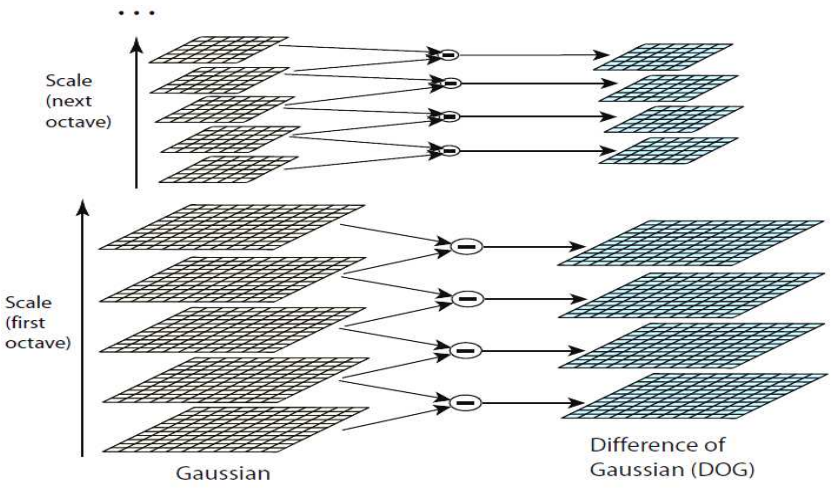


Figure 2-12. Processus de production des images de différence de gaussienne (Lowe, 2004)

Le SURF (Speeded Up Robust Features): Son principe est de calculer le déterminant de la matrice Hessienne en utilisant l'image intégrale pour extraire la localisation et l'échelle des points d'intérêt (Bay, et al., 2006). La matrice de Hessien est approximée par un ensemble de filtres de type « box ». Cette approximation peut être évaluée très rapidement en utilisant l'image intégrale, indépendamment de leur taille. Malgré les grandes approximations, la performance de détection de points est comparable aux résultats obtenus avec les gaussiennes discrétisées. Le déterminant estimé de la Hessienne approximée représente la réponse de blob dans l'image. Les extrema sont recherchés dans de petits voisinages en position et en échelle (typiquement 3 x 3 x 3). Une étape d'interpolation a pour but d'améliorer la localisation des points d'intérêt en espace et en échelle.

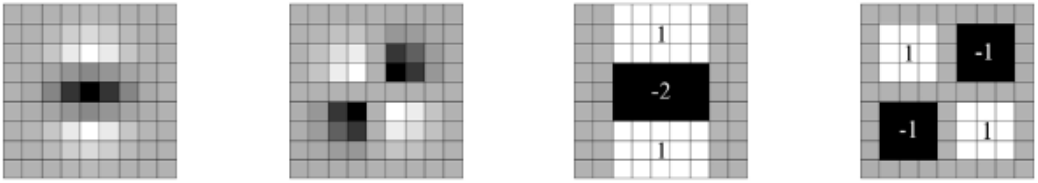


Figure 2-13. De gauche à droite : La gaussienne dérivée partielle discrétisée de second ordre dans la direction y et la direction xy, respectivement. L'approximation de la gaussienne dérivée partielle de second ordre dans la direction y et la direction xy, respectivement.

Conclusion: Souvent les détecteurs de blobs et les détecteurs de coins sont combinés adhérant ainsi à une meilleure couverture des régions et points saillants l'image. Maja et al. dans (Rudinac, et al., 2009) ont combiné simultanément 3 détecteurs (Hessian Affine, Harris Affine et Hessian Laplace) puis ont réduit le nombre de PIs en première phase par une approche de réduction spatiale (les points proches dans un certain voisinage) et en deuxième phase par une estimation de l'entropie des régions locales (garder les points les plus riches et plus représentatives).

Le tableau suivant résume les caractéristiques des différents détecteurs abordés dans cette partie:

Détecteur	Catégorie	Invariance	Temps de calcul	répétitivité	Nombre de détection
Harris	Coins	Rotation	Moyen	élevé	Important
Hessian	Région	Rotation	Rapide	élevé	Important
Harris-Laplacien	Coin	Echelle	Moyen	élevé	Moyen
DoG	Région	Echelle et rotation	Rapide	élevé	Moyen
EBSR	Région	échelle et rotation	Long	moyenne	Faible

Tableau 2-4. Comparaison des détecteurs de points d'intérêt

Le potentiel de ces détecteurs est complété par le calcul des descripteurs sur les points d'intérêts.

2.2.2 Descripteurs locaux

Le champ des descripteurs calculés sur les PIs est vaste (les dérivées de gaussiennes, les moments invariants…). Nous en avons retenu deux qui sont très fréquemment utilisés dans le monde de la reconnaissance étant donné leur efficacité : le SIFT et le SURF.

❖ **Le descripteur SIFT**

Le descripteur SIFT (Lowe, 2004) se base sur le calcul du gradient dans chaque point à l'échelle de la région entourant le PI. Cette région est divisée en 4 ou 16 sous régions. Pour chaque région, un histogramme de degrés de quantification égale à 8 cumule l'orientation du gradient spatial de l'intensité de la zone. Dans la Figure 2-14, le descripteur SIFT est un vecteur de dimension 2 x 2 x 8 = 32. Le descripteur SIFT usuelle est de taille 4 x 4 x 8 = 128.

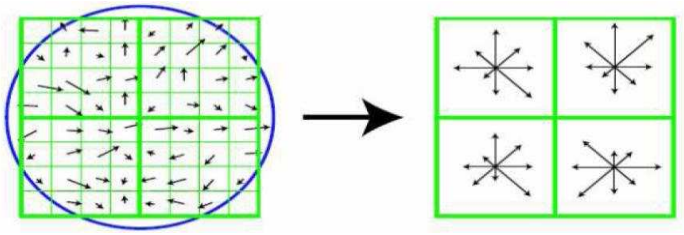


Figure 2-14. Histogramme SIFT (Lowe, 2004)

Ke and Sukthankar (Ke, et al., 2004) proposent d'appliquer une ACP sur le gradient de l'image. L'ACP a permis de réduire la dimension du descripteur à 36, lui conférant ainsi une mise en correspondance plus rapide. Toutefois, cette méthode s'est avérée être moins discriminante que le SIFT dans une deuxième étude comparative de Mikolajczyk et Schmid (Mikolajczyk, et al., 2005) et son lent temps calcul des primitives pénalise l'étape d'appariement. Dans le même article, les auteurs ont proposé une variante du SIFT, appelé GLOH, qui s'est avéré être encore plus discriminante pour une même dimension du descripteur. Cependant, le GLOH est plus couteux en temps de calcul.

❖ **Le descripteur SURF**

La signature locale en un point d'intérêt est l'ensemble des orientations principales de la fenêtre rectangulaire entourant ce point (Bay, et al., 2006). Ces régions d'intérêt sont divisées en 4 x 4 blocs. Chaque sous régions est caractérisée par un vecteur v défini par:

$v = (\sum d_x, \sum d_y, \sum|d_x|, \sum|d_y|)$; avec dx et dy sont les réponses aux ondelettes de Haar dans une direction horizontale et verticale. La dimension de ce descripteur est 4x4x4 = 64. Ce descripteur est avantageux par l'invariance aux rotations et changement d'échelle et par le faible temps de calcul.

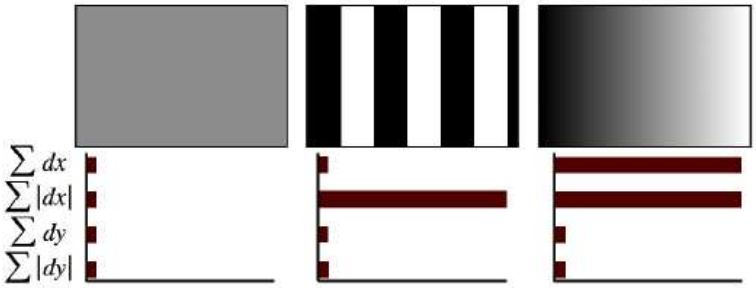


Figure 2-15. Descripteurs SURF calculés: sur une région homogène à gauche, en présence de fréquence sur l'axe x au milieu et sur une région d'intensité dégradée à droite (Bay, et al., 2006)

2.3. Détection et description des Points d'Intérêt (PI) en 3D

2.3.1 Détecteurs

L'efficacité des Points d'Intérêt a été prouvée pour la reconnaissance d'objets dans des images 2D. Dans le domaine 3D, le détecteur des PIs doit être invariant aux transformations rigides 3D. Dans cette section, nous revisitons d'une façon non exhaustive les méthodes de l'état de l'art pour la détection de points d'intérêts 3D. Deux grandes classes sont à distinguer: celle des détecteurs à échelle fixe et celle des détecteurs invariants à l'échelle. Par ailleurs, selon le type des données utilisées, nous subdivisons ces méthodes en deux sous classes: celles qui utilisent des images de profondeurs ou qui mappe le maillage 3D sur un plan 2D pour pouvoir exploiter la structure en treillis de l'image 2D et appliquer les algorithmes connus dans la 2D (nous appellerons cette classe : carte 2D) ; et celles qui utilisent des nuages de points ou des maillages 3D (que nous appellerons représentations 3D).

2.3.1.1 Détecteurs à échelle fixe

Les détecteurs à échelle fixe cherchent des PIs à une échelle constante spécifiée en paramètre dans l'algorithme. Ces approches calculent une mesure associée à chaque point (propriété de chaque sommet) ou à une région. Les PIs sont sélectionnés en maximisant la qualité de la mesure en question dans un espace de voisinage définie par l'échelle.

❖ **Sur une carte 2D**

• **Détecteur SI (LSP)** : Un exemple des approches utilisant une mesure de qualité des points est la méthode des patches de surfaces locales (Local Surface Patch, LSP) proposée par (Chen, et al., 2007).

La qualité d'un sommet est définie par l'indice de forme SI présenté dans le chapitre 1. L'idée est de détecter des points d'intérêt sur des zones de forte variation de la forme et correspondant à des extremums locaux des valeurs de SI de leur voisinage. Plus de détails de ce détecteur seront donnés dans le chapitre suivant.

• Classification selon SI : Cette approche se base sur la construction d'un espace de classification de forme avec un estimateur de l'indice de forme. Ce détecteur est proposé dans (Hozatlı, 2009), où chaque point est classé selon la valeur de son indice de forme en 7 types de régions (Figure 2-16), à l'exception des surfaces planes pour lesquelles les courbures principales s'annulent. Les points appartenant à la même classe sont groupés en utilisant le concept d'étiquetage des composantes connexes. Les composantes les plus informatives sont sélectionnées pour représenter chaque objet. Les centres des deux composantes les plus larges de chaque type de régions sont retenus comme représentants du modèle (Figure 2-17). Ainsi, 14 PIs sont détectés au maximum par objet.

Class	Region-type	Shape-index
Dome	Elliptic	$[5/8, 1)$
Ridge	Parabolic	$[3/8, 5/8)$
Saddle ridge	Hyperbolic	$[1/8, 3/8)$
Plane	Hyperbolic	Undefined
Saddle-point	Hyperbolic	$[-1/8, 1/8)$
Saddle-rut	Hyperbolic	$[-3/8, -1/8)$
Rut	Parabolic	$[-5/8, -3/8)$
Cup	Elliptic	$[-5/8, -1)$

Figure 2-16 Typographie des surfaces selon la valeur de l'indice de forme

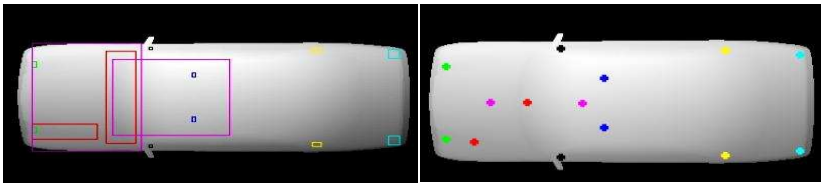


Figure 2-17. A gauche-Etendu des 14 composantes connexes détectées-A droite- les centres de ces composantes. Les couleurs représentent les différents types de surfaces, à savoir Dome, Ridge, Saddle Ridge, Saddle Point, Saddle Rut, Rut et Cup.

❖ Sur la représentation 3D

• **L'Intrinsic Shape Signatures (ISS)** : Comme méthode basée sur une mesure de qualité de régions, l'Intrinsic Shape Signatures (ISS) (Zhong, 2009) calcule une décomposition en valeurs propres (Eigen Value Decomposition (EVD)) de la matrice de covariance des points appartenant au voisinage. L'ISS utilise la magnitude de la plus petite valeur propre pour prendre seulement les points ayant une large variation dans chaque direction principale. Le ratio entre deux valeurs propres successives sert à exclure les points ayant une répartition similaire sur les directions principales.

• **Carte 2D des longitudes latitudes** : Atmosukarto et Shapiro (Atmosukarto, et al., 2008) définissent une signature basée sur la représentation d'une carte 2D des longitudes latitudes aux points saillants (Figure 2-18). Ils utilisent un classificateur pour trouver les points et les annoter selon les scores de prédiction. En effet, en premier lieu, ils extraient sur les points du maillage une propriété locale par le biais d'un opérateur bas-niveau. Cinq opérateurs ont été testés: la valeur absolue de la courbure gaussienne, la courbure moyenne, l'indice de forme, la magnitude de la normale et l'intensité de courbure. Puis, ces valeurs sont convoluées avec un filtre gaussien pour réduire le bruit et former les primitives. Ensuite, des histogrammes qui cumulent ces primitives sur un voisinage de chaque point sont utilisés pour entrainer dans un premier classificateur SVM. Et finalement, ce classificateur permet de classer les points saillants et non saillants avec un score de confiance attribué à la décision. La carte 2D des longitudes latitudes consiste à associer à chaque point 3D p_i (p_{ix} , p_{iy} , p_{iz}) de la surface de l'objet les deux mesures : Longitude : $\theta_i = \arctan\left(\frac{p_{iz}}{p_{ix}}\right)$ et la latitude : $\phi_i = \arctan\left(\frac{p_{iy}}{\sqrt{{p_{ix}}^2+{p_{iz}}^2}}\right)$. Ces valeurs de longitudes latitudes sont subdivisées en un nombre fixé de bins étiqueté par le point saillant y correspondant.

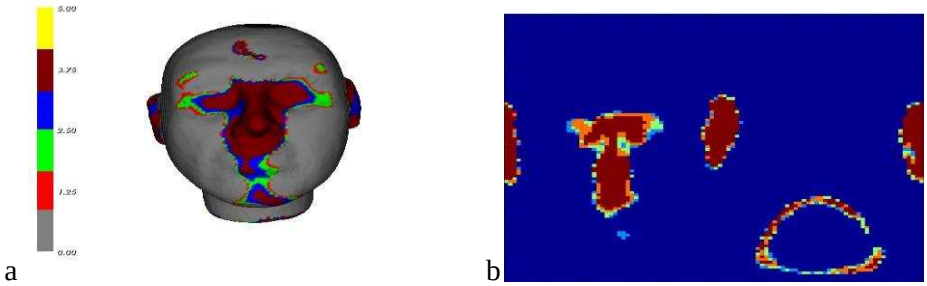


Figure 2-18. a-Modèles de points saillants sur un visage 3D- b- La carte 2D longitude-latitude Points saillants avec un indice de confiance 0.95−0.96: rouge, 0.96−0.97:vert, 0.97−0.98: bleu, 0.98−0.99: rouge foncé et 0.99−1.0: jaune. Les points gris ne sont pas des points saillants

• Pour la **KeyPoint Quality (KPQ),** Mian et al. (Mian, et al., 2009) abordent la détection de PIs dans un modèle 3D et dans des vues 2,5. Ils procèdent pour le modèle 3D comme suit: En chaque point échantillonné (sommet), la surface locale est découpée à l'aide d'une sphère de rayon r (déterminé automatiquement) autour du point. La vue locale obtenue est mise en rotation jusqu'à l'alignement de sa normale avec l'axe des Z. Après suppression des points avec occultation, une ACP est appliquée sur le reste des points de la vue qui sont par la suite projetés sur les vecteurs propres. Un ratio r calculé sur les deux premiers axes principaux de la surface locale permet de renseigner sur la symétrie de la surface en question. Une base de coordonnées 3D est extraite de la surface seulement si ce ratio est supérieur à 1. Un deuxième seuil t >1 est introduit pour contrôler le bruit. Au final, un point est détecté s'il vérifie r > t. Cette première passe permet d'éliminer les points non significatifs en seuillant le ratio entre les longueurs maximales sur les deux premiers axes principaux. L'invariance à la translation dans cette approche est due au fait que le PI est choisi comme origine de la base de coordonnées (la matrice de covariance tient compte seulement de 3 rotations). Dans le cas des vues partielles, essentiellement, une modification est apportée: une condition sur les bords est ajoutée pour éliminer les PIs du bord. Une surface découpée dans le bord va avoir une valeur de r grande. Dans le cas d'un nuage de points ordonné selon une grille de X Y Z, la condition considère les points qui sont à une distance r du bord. Avec les surfaces symétriques (cône, sphère, parabolique, pics…), ce critère de sélection ne permet pas de prendre les centres, mais plutôt les points qui l'entourent. Un critère de

sélection supplémentaire des PIs est déterminé par une mesure Q de la qualité du point d'intérêt, qui est une combinaison empirique des courbures calculées sur la surface lissée, ré-échantillonnée et approximant les données alignées (sera présentée dans la partie « Facteur de qualité de PIs »). Un seuil mesurant la distance minimale entre les PIs permet la sélection finale (Figure 2-19).

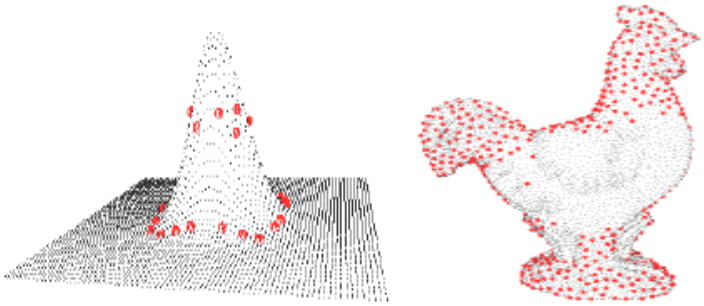


Figure 2-19. Points détectés sur une surface symétrique (gaussienne 2D) et sur le modèle d'un coq (Mian, et al., 2009)

• **<u>Points saillants pondérés</u> :** Tangelder et Veltkamp (Tangelder, et al., 2003) proposent une approche locale fondée sur des représentations par points saillants pondérés permettant la prise en compte de la position spatiale relative de l'objet polyèdre. Ils commencent par appliquer une ACP pour normaliser la position de l'objet et le décompose ensuite en une grille 3D. Une signature se compose de points saillants pondérés associés à des cellules non vides de la grille. Pour l'opération de détection et de pondération des points d'intérêt, ils comparent trois méthodes qui utilisent les sommets et les facettes qui lui sont adjacentes:

a- Méthode de courbures gaussiennes : qui calcule la courbure gaussienne associée à un sommet par la formule : $c(v) = \frac{d(v)}{a(v)}$ avec d(v) est le défaut angulaire qui est égal à 2π moins la somme des angles intérieurs des facettes issues de v, a(v) est l'aire associé au sommet v qui est égale à la somme des aires des facettes divisé par leur nombre de sommets. Pour chaque cellule non vide, le sommet ayant la plus grande valeur absolue de la courbure gaussienne est choisi comme point saillant. Le poids attribué à ce point est M (|c|)= 1 – 1/ (1+ |c|).

b- Méthode de variation des normales: le point saillant p(c) est la moyenne de la zone pondérée des sommets dans une cellule: $p(c) = \left(\sum_{j=1}^{M} w_j v_j\right) \Big/ \sum_{j=1}^{M} w_j$, avec M le nombre de sommets dans la cellule, v_j le j-ième sommet de la cellule, w_j est le a(v) déjà défini précédemment. Pour chaque cellule, la normale moyenne n est calculée en fonction des normales des facettes dans cette cellule pondérées par leur aire : $n = \sum_{i=1}^{N} a_i n_i$. La paire (point, poids) : (p(c), $1 - (\|n\| / \sum_{i=1}^{N} a_i)$) est ajoutée à la signature.

c- Méthode du mi-point: Dans le cas où les polygones sont non orientés, cette méthode attribue à chaque cellule de la grille un point saillant qui est le centre de masse de tous les sommets en lui assignant un poids unitaire.

• **Harris 3D** (Sipiran, et al., 2010) L'algorithme propose une extension de la méthode de Harris (Harris, et al., 1988) de détection de coins pour des maillages 3D. Le principe de cette méthode se base sur les étapes suivantes :

- Extraire le voisinage de chaque sommet. Ce voisinage peut être spatial, adaptatif ou dans un anneau.
- Construire un système canonique local en appliquant une ACP sur le voisinage.
- Ajuster une surface quadratique au voisinage normalisé.
- Calculer les dérivées sur la surface ajustée. Le lissage gaussien appliqué sur les dérivées rend cette méthode robuste aux changements géométriques.
- La fonction d'auto-corrélation construite à partir des dérivatives permet d'évaluer l'opérateur de Harris. Ainsi, une réponse est calculée pour chaque sommet.
- Deux façons pour sélectionner les PIs sont proposées: soit en prenant un nombre fixe de sommets ayant les plus grandes réponses, soit par une approche de regroupement (clustering) pour avoir des PIs bien distribués. L'algorithme de ce dernier sera donné dans le chapitre Détection.

La Figure 2-20 représente les points de Harris détectés sur des modèles 3D synthétiques.

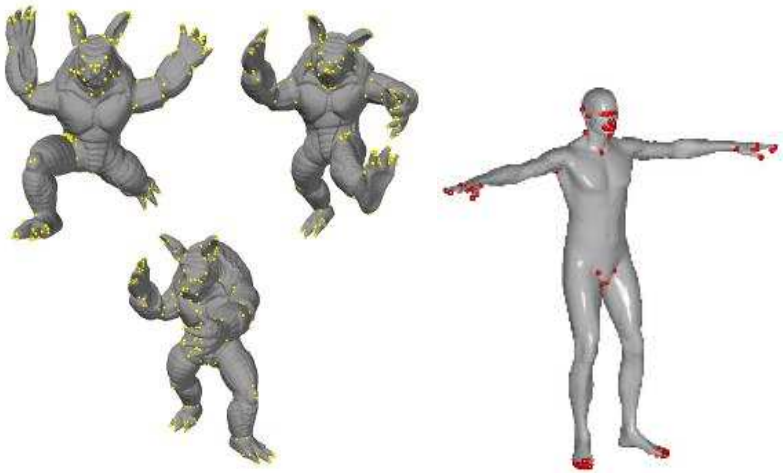


Figure 2-20. Points Harris 3D sur des modèles synthétiques (Sipiran, et al., 2010)

2.3.1.2 Détecteurs invariants à l'échelle

Les détecteurs invariants à l'échelle, comme est le cas pour les images 2D, recherchent des points d'intérêt dans un espace d'échelles du maillage. Les PIs sont détectés à différentes échelles et une échelle est leur associée. Pareillement qu'aux méthodes à échelle fixe, ces approches calculent une mesure de qualité qui est, par contre, associée à chaque position spatiale et à chaque échelle. Les PIs sont sélectionnés en maximisant la mesure de qualité sur les niveaux spatiales et échelles.

❖ **Sur une carte 2D**

• **SIFT 2D** : Un algorithme pour la reconnaissance de visage 3D basé sur l'extraction des PIs par le SIFT 2D est proposé par Mayo et Zhang (Mayo, et al., 2009). L'idée de base est de faire pivoter chaque nuage de point 3D autour des axes x, y ou z et de projeter les points 3D sur des images 2.5D à chaque étape de la rotation. Le détecteur et descripteur SIFT 2D est appliqué sans modification sur les images de profondeur. En phase de test, un visage inconnu subit d'abord la même technique

d'extraction des PIs dans plusieurs vues et puis il est reconnu par un algorithme d'appariement pondéré par un système de vote des PIs.

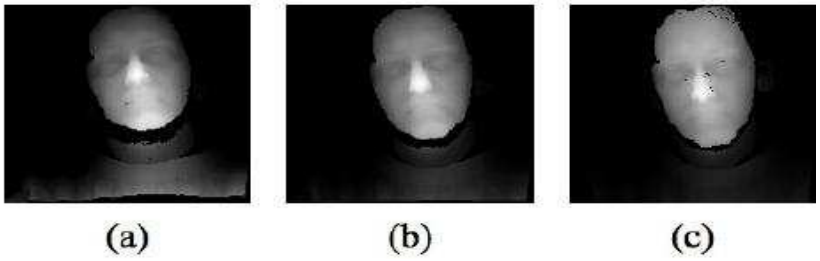


Figure 2-21 . Exemples d'images 2,5D frontales générées par la rotation du même nuage de point 3D autour de l'axe x avec un angle de (a) -10°, (b) 0° et (c) +10° (Mayo, et al., 2009)

• **SIFT sur les images de profondeurs**: Lo et Siebert (Lo, et al., 2008) ont implémenté une nouvelle version du SIFT adaptée aux images de profondeurs. Un masque de segmentation est appliqué sur l'image pour isoler la zone d'intérêt. Pareillement au SIFT 2D, l'extraction des courbures est utilisée pour détecter les PIs. Les courbures principales sont calculées en fonction des dérivées premières et secondes de gaussiennes. Un indice de forme constitué par la distribution de ces courbures permet de décrire les différents types de surfaces autour du PI. Par la suite, un masque gaussien circulaire permet d'échantillonner l'image. Les orientations et les magnitudes du gradient de l'image locale sont calculées avec les dérivées premières gaussiennes. Un histogramme des orientations cumule les fréquences de ces orientations dans 360 bins. L'histogramme des orientations est stabilisé par la convolution trois fois avec un noyau gaussien. De plus, la zone locale autour du PI subit une rotation selon l'orientation canonique du point pour assurer une invariance aux points de vue avant de calculer les descripteurs. Ainsi, le descripteur final est formé par un indice de forme et par l'histogramme des orientations. Le tableau suivant décrit les principales différences entre le SIFT 2D et le SIFT 2.5.

Etapes	Traitement	SIFT 2D	SIFT 2.5D
Localisation des Points d'Intérêt	Seuil du contraste	0.3	0.003
	Seuil du ratio	10	5
	Méthode d'extraction de la courbure	Matrice Hessienne	1ère et 2ème dérivée gaussiennes
Assignement de l'orientation canonique	Nombre de bins dans l'histogramme d'orientation	36	360
	Lissage de l'histogramme	Gaussienne pondérée+ fenêtre circulaire avec un sigma=1.5*échelle du point	Convolution par Noyau gaussienne de taille 17 et sigma=17 ; appliqué 3 fois

Tableau 2-5. Différences entre SIFT 2D et SIFT 2.5D (Lo, et al., 2008)

• **SIFT sur les images des indices de forme** : Dans (Bayramoglu, et al., 2010), l'algorithme SIFT est appliqué à l'image construite à partir des valeurs de l'indice de forme sur le modèle original (Figure 2-22). Une comparaison avec les méthodes du détecteur SI (LSP) (Chen, et al., 2007) et du SIFT 2.5D (Lo, et al., 2009) est faite. L'extraction des PIs sur les images SI parait plus efficace que la sélection des extrema sur les histogrammes des SI autour du PI ou l'application du SIFT directement sur des images de profondeurs. L'explication donnée à ce résultat est la suivante : d'une part, chaque valeur SI en un point contient une information sur le voisinage de ce point alors que le pixel dans une image de profondeur détient seulement la profondeur du point. D'autre part, les régions riches en informations sont visuellement accentuées par le comportement non linéaire de la fonction SI et par le caractère discriminant du SIFT.

Dans la Figure 2-23 sont illustrées des mises en correspondances entre les paires des PIs obtenus par la méthode SIFT sur les images SI.

Figure 2-22. Schéma de l'application du SIFT sur une image SI

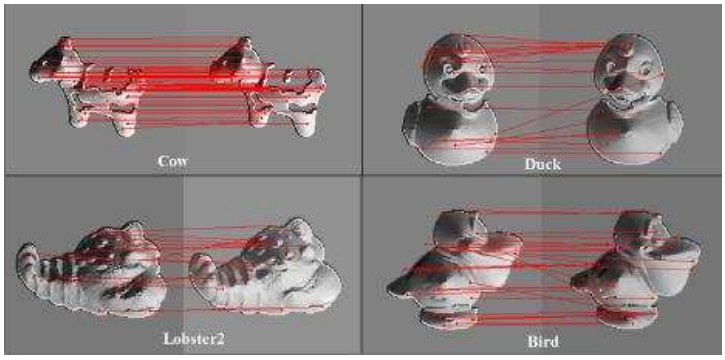


Figure 2-23. Illustration sur des images SI, des mises en correspondances entre les paires de PIs obtenus par la méthode SIFT

• **SC multi- échelle et HK multi- échelle**

Comme introduit dans le chapitre 1, la paire formée par l'indice de forme et l'intensité de courbure permet classer les formes locales. Rappelons que l'intensité de courbure renseigne sur la quantité de courbure et que l'indice de forme décrit le type de la forme. Pareillement à la classification SC, l'espace HK donne une typologie des surfaces locales des points. Pour ces deux méthodes, le calcul se fait sur un voisinage local autour des points. Si la taille de ce voisinage est constante, seulement les PIs ayant une taille inférieure à cette taille seront extraits sur la surface. Akagündüz et Ulusoy (Akagündüz, et al., 2009) proposent une approche en espace d'échelle sur la surface et la courbure. Une comparaison de ces deux espaces de classification en termes de reconnaissance d'objet est menée. L'espace d'échelle est obtenu par une pyramide de gaussiennes, correspondant à un filtrage gaussien et un sous- échantillonnage des données de l'image de profondeurs. Au niveau S de la pyramide est appliqué le facteur 2^S. En utilisant U et V, les valeurs de la dimension de la surface, et S la dimension

d'échelle, l'espace d'échelles UVS est construit. L'algorithme calcule la mesure de qualité (HK) ou (SC) en chaque point et les primitives obtenues pour les composantes des UVS sont connectées si elles ont le même type de surface. Pour chaque composante connexe est estimé un descripteur composé par les valeurs: type, position, échelle, rayon, volume et vecteur normal. L'espace d'échelle basé sur HK donne de meilleurs résultats en termes de reconnaissance d'objets que celui basé sur SC. Dans la Figure 2-24 est illustré le résultat de l'extraction des dix primitives les plus larges sur trois objets.

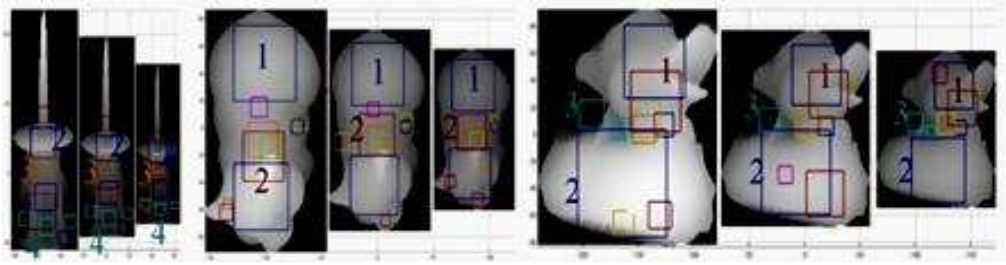


Figure 2-24. Les dix primitives les plus larges extraites pour trois objets;
pour chaque objet est représenté le modèle original au milieu, la version en augmentant l'échelle de 1.2 fois (à gauche) et la version en réduisant l'échelle de 0.8 (à droite).Le type et la taille sont représentés par les carrés colorés (bleu : peak, rouge : saddle ridge, pourpre : convex cylinder, cyan : pit, jaune : saddle valley, vert : concave cylinder, orange : hyperbolic, gris : plane)

❖ Sur la représentation 3D

• **MeshDoG**: Proposé par Zaharescu et al (Zaharescu, et al., 2009), le MeshDoG construit un espace d'échelle à partir du maillage 3D en utilisant des convolutions itératives avec un noyau gaussien. Lors de la création de cet espace, l'algorithme lisse la valeur d'une fonction scalaire f définie en chaque sommet sans modifier la géométrie du maillage. Cette fonction peut correspondre à la courbure moyenne, courbure gaussienne ou à l'information photométrique du sommet. La détection se fait en deux étapes : les extrema de la fonction du Laplacien (approximée par la différence entre les échelles adjacentes- différence de gaussiennes DoG) sont détectés à travers les échelles, suivi par une suppression des non –maximum en utilisant un voisinage de 1-anneau, sur l'espace et à travers les échelles adjacentes. Deuxièmement, un filtrage des PIs est appliqué en fixant le nombre maximum de PIs détectés à une valeur égale à un pourcentage du nombre des sommets de la maille et en éliminant les réponses ne correspondant pas à des coins. Dans (Unnikrishnan, et al., 2008), Unnikrishnan et Hebert construisent un espace d'échelles à la manière du MeshDoG. La fonction scalaire utilisée est un invariant calculé avec l'opérateur de Laplace-Beltrami et correspond au déplacement d'un point tout au long de sa normale d'une quantité proportionnelle à la courbure moyenne (H).

• **Mesh SIFT** : Dans l'approche de Maes et al. (Maes, et al., 2010), l'espace d'échelle contient les versions lissées du maillage et convoluées avec un filtre binomial. La courbure moyenne est calculée pour chaque sommet et à chaque échelle. Soit H_i la courbure moyenne d'un sommet à une échelle i. La différence entre les échelles adjacentes est calculée comme suit :

$$DH_i = H_{i+1} + H_i$$

La valeur DH de chaque sommet est comparée à celle des ses voisins se trouvant : sur la même échelle, sur l'échelle d'avant et sur l'échelle d'après. Un PI est détecté si la valeur de son DH est la plus grande ou la plus petite par rapport à celle de tous ses voisins.

• **Heat Kernel Signature** Sun et al (Sun, et al., 2009) utilisent comme mesure de qualité le noyau de chaleur calculé sur la maille: la résolution de l'équation de chaleur sur l'espace et le temps permet de construire un équivalent des espaces d'échelle. Les maxima du noyau sont alors choisis comme PIs.

• **Le 3D SURF** Knopp et al (Knopp, et al., 2010) introduisent une extension du SURF à la 3D, à la manière de l'approche de Willems et al. (Willems, et al., 2008) qui définissent une Hessienne 3D pour des primitives spatio-temporelles. La saillance maximale est extraite sur plusieurs octaves obtenues par l'application de gaussiennes aux données voxélisées. Plus de détails sur ce détecteur seront donnés dans le chapitre détection.

• **Paires de points saillants** : Li et al. (Li, et al., 2007) utilisent la détection des points saillants avec une projection linéaire sur une surface lissée et une seule version d'échelle des caractéristiques multi-échelles (Li, et al., 2005). En effet, l'approche de sélection des PIs se base sur une représentation en espace d'échelle d'une surface: projection des points 3D et de leur normale sur une version lissée de la surface. Pour ce faire, ils procèdent comme suit : une surface S est lissée à l'aide d'un opérateur résiduel défini pour les points de S, une échelle et une direction des normales lissée avec une gaussienne normalisée. Deux représentations de la surface S sont construites pour deux échelles. Puis, la différence des normales entre ces deux nivaux permet de trouver un voisinage. Les maximas du voisinage défini représentent les points saillants. Ces points sont utilisés par paire. Deux PIs forment une paire si la distance entre eux est comprise entre deux seuils fixés. Des nouvelles entités sont calculées pour la paire et donnent une position **p** qui est la moyenne des positions des deux points, un facteur **d** qui est fonction de la norme des deux positions et une direction **n** (Figure 2-25).

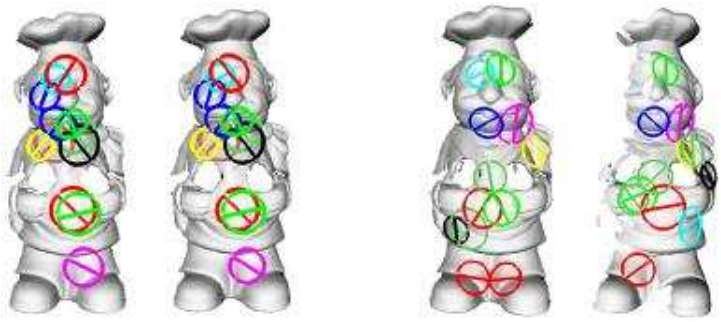


Figure 2-25. Mise en correspondance des paires de points détectées pour : deux niveaux d'échelles différents (à gauche) et deux vues différentes (à droite) du même objet.

• **L'approche multi-échelle basée sur l'intensité de courbure C**

L'approche de Ho et Gibbins (Ho, et al., 2009) est schématisée dans l'algorithme de la Figure 2-26. L'idée principale dans cette approche est que l'augmentation de la taille du voisinage local des points est similaire à l'application d'un filtre lisseur. Ce qui au final évite d'apporter une modification directe sur la surface 3D. L'intensité de courbure C des points est calculée en tout point et à différentes échelles (différentes tailles du voisinage local). Les points maxima et minima en intensité de courbures

à travers les échelles sont sélectionnés comme d'éventuels PIs. Une mesure de confiance est attribuée et est basée sur la déviation de leur intensité de courbure par rapport à celle du voisinage.
A part la détection de primitives à plusieurs échelles, un autre avantage de cette approche est la réduction de l'effet du bruit sur les modèles et son adaptation à des données en format maillée ou en format de points désordonnés. De plus, l'échelle à laquelle la sélection du PI est faite peut être utilisée comme support (la région des voisins entourant le PI) pour l'étape de description.

Algorithm 1 Multi-scale Curvedness-based Feature Extraction Algorithm

Data:

$\mathbf{P} = \{\mathbf{p}_i \in \mathbf{R}^3\}$: set of 3D points sampled from the surface.

$\mathbf{R} = \{r_k\}$: a set of scales.

Algorithm:

1: **for** $r \in \{r_k\}$ **do**
2: **for** $\mathbf{p} \in \{\mathbf{p}_i\}$ **do**
3: Find the neighbourhood N_r at scale r
4: Fit a jet to N_r
5: Compute principal curvatures k_1 and k_2
6: Compute the curvedness $c_{\mathbf{p}}$

$$c_{\mathbf{p}} = \sqrt{(k_1^2 + k_2^2)/2}$$

7: **end for**
8: Keypoints are positions $\mathbf{p}$ having extremum values $c_{\mathbf{p}}$ both in the neighbourhood of radius r_k as well as over the above and below scales (r_{k-1}, r_{k+1}).
9: **end for**

Figure 2-26. Algorithme multi-échelle basée sur l'intensité de courbure C (Ho, et al., 2009)

• **Détecteur de Sparse points** : Ce détecteur utilise une représentation d'espace d'échelles avec une pyramide de Gaussiennes, et a été proposé par (Castellani, et al., 2008). La mesure de qualité est le déplacement de chaque sommet par rapport à sa position d'origine après application du filtre Difference-of-Gaussians (DoG) (Lowe, 2004). L'algorithme passe par les deux phases connues dans le SIFT : phases inter et intra octaves illustrées dans le schéma de la Figure 2-27. Dans la phase intra-octave, différents opérateurs gaussiens sont appliqués au maillage source. Ensuite, la carte d'espace est calculée, puis normalisée. Finalement, une détection en intra-octave permet d'extraire les PIs.

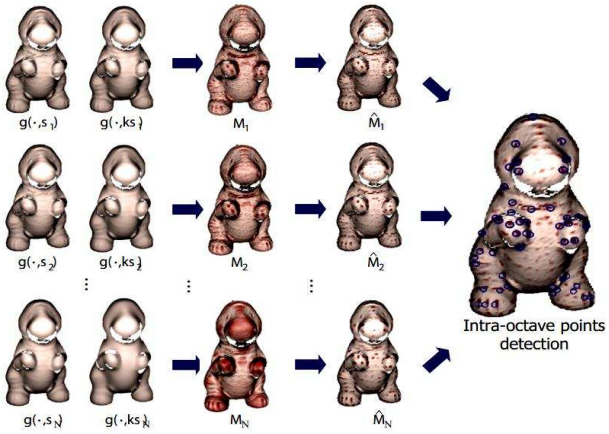


Figure 2-27. Schéma de la phase inter-octave et intra-octave pour extraire les PIs

• **THRIFT:** Flint et al. (Flint, et al., 2007) ont proposé une version 3D de la Hessienne pour mesurer la saillance des points candidats. Une fonction de densité est approximée par un échantillonnage régulier des données dans l'espace représenté par les cases B_{ijk} définis par :

$$B_{ijk} = \{(x,y,z) \in \mathbb{R}^3 : \alpha i \leq x < \alpha(i+1), \\ \beta j \leq y < \beta(j+1), \\ \gamma k \leq z < \gamma(k+1)\}$$

Soit n(B) le nombre de points contenus dans la case B. Une carte de densité normalisée est donnée par :

$$D(i,j,k) = \frac{n(B_{ijk})}{\underset{(i,j,k)\in I}{\operatorname{argmax}}\left\{n(B_{ijk})\right\}}$$

L'espace d'échelle est construit sur cette fonction de densité. Une recherche des maxima locaux du déterminant de la Hessienne est appliquée pour trouver les PIs.
L'inconvénient majeur de cette approche reste le cout du ré-échantillonnage régulier dans l'espace du maillage et l'instabilité du déterminant de la Hessienne utilisée pour les maxima locaux dans le cas de des données bruitées.

• **KPQ-SI** : Finalement, dans (Mian, et al., 2009), les PIs invariants à l'échelle sont obtenus à partir d'un ensemble de PIs à échelle fixe, extraits à différentes échelles par le détecteur KPQ introduit dans le paragraphe précédant. L'échelle caractérisant le point est donnée par celle correspondant au maximum global sur l'axe échelle du ratio entre les longueurs maximales sur les directions principales (qui n'est pas seuillé, contrairement au cas de l'échelle fixe).

Un résumé des différentes méthodes abordées dans cette partie est donné dans le tableau suivant :

Détecteurs à échelle fixe		Détecteurs invariants à l'échelle	
Sur carte 2D	**Sur représentation 3D**	**Sur carte 2D**	**Sur représentation 3D**
Local Surface Patches (LSP) (Chen, et al., 2007) Classification selon SI (Hozatlı, 2009)	Intrinsic Shape Signatures (ISS) (Zhong, 2009) KeyPoint Quality (KPQ) (Mian, et al., 2009) Points saillants pondérés (Tangelder, et al., 2003) Harris 3D (Sipiran, et al., 2010)	SIFT 2D (Mayo, et al., 2009) SIFT 2.5D (Lo, et al., 2008) SIFT sur image SI (Bayramoglu, et al., 2010) SC et HK multi- échelle (Akagündüz, et al., 2009)	MeshDoG (Zaharescu, et al., 2009) MeshSIFT (Maes, et al., 2010) Heat Kernel Signature (HKS) (Sun, et al., 2009) 3D-SURF (Knopp, et al., 2010) Paires de points saillants (Li, et al., 2007) Courbure multi-échelle (Ho, et al., 2009) Sparse Points (Castellani, et al., 2008) THRIFT (Flint, et al., 2007) KPQ-SI (Mian, et al., 2009)

Tableau 2-6 Récapitulatif des méthodes de détection de Points d'Intérêt 3D

2.3.1.3 Mesures de la performance de détection des Points d'Intérêt 3D

En reconnaissance, un choix judicieux des PIs permet de minimiser le temps de calcul et éviter l'ambiguïté dans le calcul des descripteurs et de la mise en correspondance. Pour ce faire, des mesures de performance de la détection sont introduites.

2.3.1.3.1 Facteur de qualité des PIs

Mian et al. (Mian, et al., 2009) définissent le PI comme le point satisfaisant 3 contraintes :

- une grande répétitivité dans les vues 2,5D et le modèle 3D du même objet.
- une unique base de coordonnées 3D peut être définie dans la surface voisine pour extraire des primitives locales invariantes.
- le voisinage du PI doit être discriminant et caractérise le point d'une façon unique.

Pour ne pas avoir à choisir les points d'une façon aléatoire dans le cas où plusieurs points répondent aux contraintes, ils proposent une mesure de qualité Q du PI basée sur le calcul de la courbure principale en chaque point d'une grille subdivisée en n x n cellules d'une surface locale lissée. Cette mesure est définie par:

$$Q_k = \frac{1000}{n^2} \sum |K| + \max(100K) + |\min(100K)| + \max(10\,\kappa_1) + |\min(10\kappa_2)|\,;$$

$K = \kappa_1\kappa_2$ est la courbure gaussienne et $\kappa_1 et\ \kappa_2$ sont les courbures principales. Les keypoints sont triés selon la valeur de leur mesure de Q_k. Les points avec un rayon égal à la distance minimale permise d_1 qui est égale à deux fois le pas d'échantillonnage de la surface sont rejetés.

Des résultats encourageants sont obtenus comme le montrent les figures suivantes :

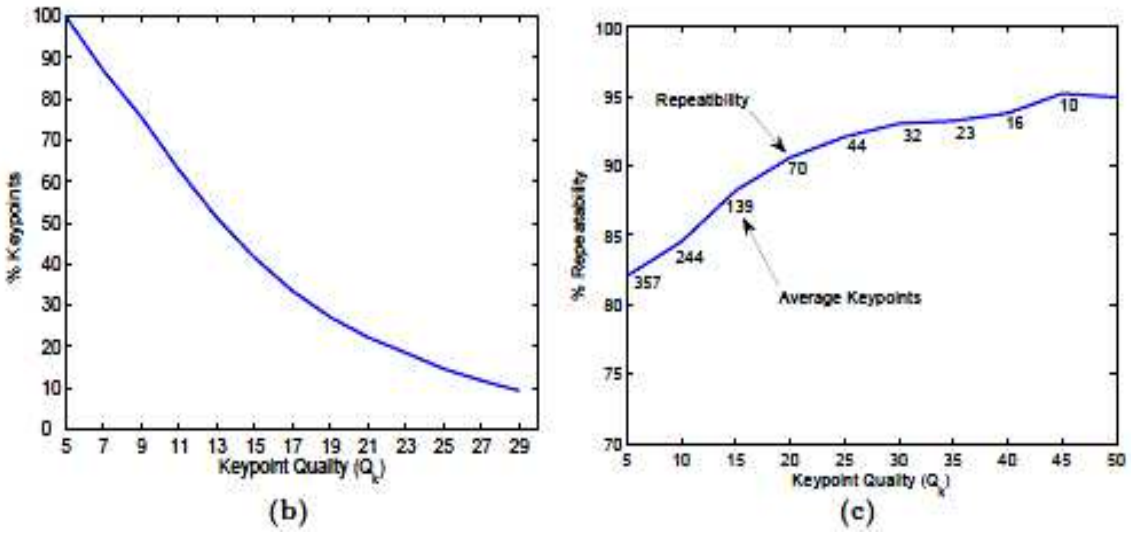


Figure 2-28. A gauche- Pourcentage des PIs en fonction de la qualité Q_k -A droite- Répétabilité de l'identification des PIs en fonction de la qualité Q_k (Mian, et al., 2009)

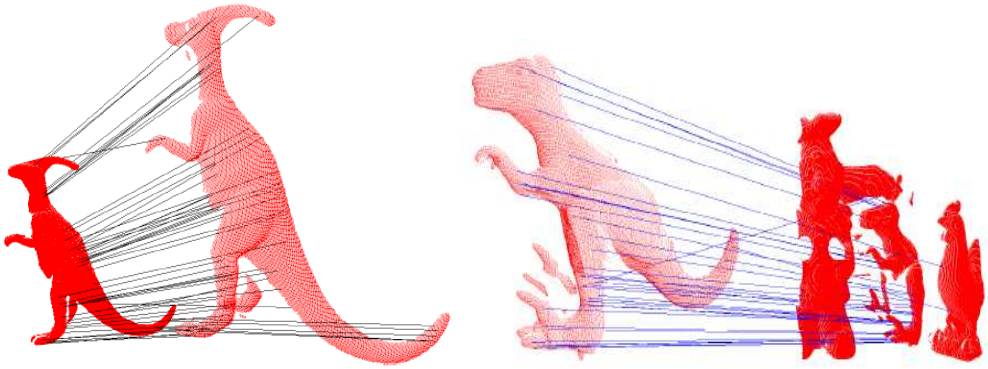


Figure 2-29. a- Invariance au changement d'échelle-b- Invariance au changement d'échelle dans une scène avec des occultations (Mian, et al., 2009)

2.3.1.3.2 Répétabilité

Les points détectés doivent être invariants aux déformations, changements d'échelles et d'angles et robustes au bruit. Ce qui veut dire que la détection doit sortir le même point après une déformation ou ajout de bruit au modèle original.

Dans (Schmid, et al., 2000), l'évaluation de la performance des détecteurs de PIs est faite par la mesure de répétabilité absolue, qui est la proportion des PIs dont la position est la même en passant d'une vue à une autre pour le même objet. Un PI est dit répétable si :

$$\underbrace{\left\| R_{ms} K_m^i + t_{ms} \right.}_{\text{Modèle du keypoint après rotation et translation dans la scène}} - \underbrace{\left. K_s^j \right\|}_{\text{keypoint dans la scène}} < \underbrace{\varepsilon}_{\text{Seuil de répétabilité}}$$

Ainsi pour mesurer la répétabilité des PIs entre différentes vues/échelles, nous considérons deux vues: vue 1 et vue 2 du même objet. La transformation réelle T (rotation ou homothétie) étant connue entre les deux vues, nous calculons la distance entre les positions de chaque PI détecté dans la vue 1 et ayant subi la transformation T et le PI le plus proche détecté dans la vue 2.

2.3.2 Descripteurs de points d'intérêt 3D

Descripteur LSP: indice de forme vs cosinus des normales : Le descripteur LSP (Local Surface Patches) est introduit par Chen et Bhanu (Chen, et al., 2004).Un histogramme 2D est formé en cumulant les points d'un voisinage au point référence, dans des cases selon deux axes relatifs à la valeur de l'indice de forme et au cosinus de l'angle entre la normale et celle du point référence. La représentation LSP inclut le type T_p, le centroïde de la zone décrite, et l'histogramme 2D. Plus détails de ce descripteur seront données dans le chapitre Descripteurs.

• **Mesh-HoG :** Zaharescu et al (Zaharescu, et al., 2009) proposent, pour un PI, ce descripteur qui est calculé sur un support géodésique de région, proportionnel à la dimension de la surface. Le gradient

3D de chaque sommet du voisinage est calculé sur l'échelle détectée. Le descripteur est invariant à la rotation par le choix d'un système de coordonnées locales. Un histogramme de gradient est calculé à la fois spatialement et en utilisant les orientations. Les histogramme 3D sont concaténés et normalisés. Au final, un descripteur de taille 96 est obtenu.

• **Descripteur SIFT 3D :** Scovanner et al. (Scovanner, et al., 2007) introduisent un descripteur SIFT 3D pour des vidéos avec le temps comme troisième dimension. Similairement au 2D, le gradient spatio-temporel (L_x, L_y, L_t) est calculé sur les point 3D (x, y, t). Ils définissent la magnitude et l'orientation du gradient comme suit :

$$m_{3D}(x,y,t) = \sqrt{L_x^2 + L_y^2 + L_t^2}\,; \qquad \theta(x,y,t) = tan^{-1}\left(\frac{L_y}{L_x}\right); \qquad \varphi(x,y,t) = tan^{-1}\left(\frac{L_t}{\sqrt{L_x^2+L_y^2}}\right)$$

A chaque point est assigné un couple d'angles (θ, φ) qui renseigne sur l'orientation du gradient en 3D. Un histogramme 2D qui cumule la distribution des bins de ces angles est fourni comme signature. Les points d'intérêt sont choisis aléatoirement pour permettre une utilisation en temps réel.

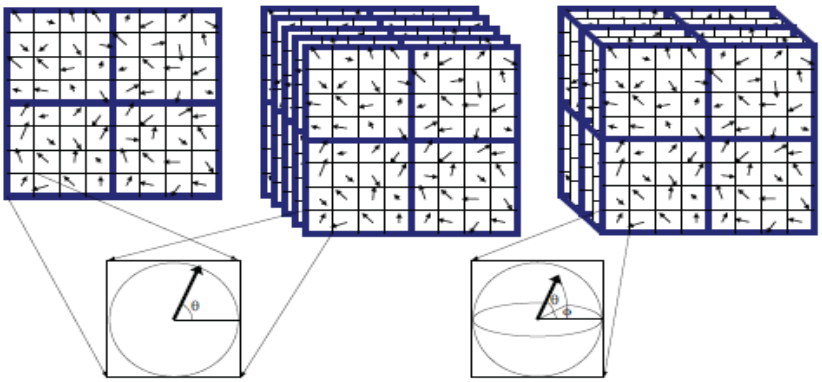


Figure 2-30. L'image gauche : le descripteur SIFT 2D normal. Milieu : le descripteur SIFT 2D multiple utilisé sur une vidéo sans modification de la méthode originale. Droite : le descripteur SIFT 3D avec ses sous-volumes 3D, chaque sous-volume est accumulé dans son propre sous-histogramme. Ces histogrammes forment le descripteur final. (Scovanner, et al., 2007)

Etapes	Traitement	SIFT 2D	SIFT 2.5D	SIFT 3D
Extraction du descripteur	Information retenue	Orientation (θ)	Indice de forme et orientation	Orientation (θ, ϕ)
	Echantillonnage	4x4	9 sous régions par des enveloppes gaussiennes	4x4x4
	Dimension du descripteur	128	153	2048= ((4x4x4) x (8x4))

Tableau 2-7 Comparaison entre les descripteurs SIFT 2D, SIFT 2.5D et SIFT 3D

• **Descripteur Thrift :** Le descripteur Thrift, proposé par Flint et al. (Flint, et al., 2007) s'inspire du formalisme du SIFT 2D. L'information de la forme locale est décrite à partir des entités différentielles de premier ordre à savoir les normales des points sur le support. Chaque PI est décrit par un histogramme qui regroupe, pour tous ses voisins, le cosinus de l'angle entre deux normales calculées sur deux supports du même point. En comparaison avec un histogramme qui cumule seulement les coordonnées 3D des points, cette représentation est plus robuste et plus descriptive de la structure locale de la surface. Le formalisme de ce descripteur sera détaillé dans le chapitre Descripteurs.

• **Descripteur SURF 3D :** Dans le travail de Knopp et al. (Knopp, et al., 2010), un descripteur SURF 3D invariant à la rotation et à l'échelle, est calculé pour chaque point d'intérêt. La formation d'un système local au PI se fait par un échantillonnage uniforme des réponses des ondelettes de Haar sur 3 axes, avec une distance du PI, proportionnelle à l'échelle. Par la suite, pour garantir plus de robustesse à un changement dans la position, chaque réponse est pondérée par une gaussienne centrée sur le PI. Les vecteurs des réponses dans tous les cônes ayant une certaine ouverture d'angle sont additionnés et l'orientation dominante est définie par la direction du vecteur résultant le plus long. Toutes les réponses sont projetées sur cette direction et une seconde orientation est obtenue avec une fenêtre glissante. Ces deux directions définissent le système de coordonnées locales. Ce descripteur est calculé sur une grille NxNxN. Sur chaque cellule de la grille, un vecteur de 6 dimensions de la réponse des ondelettes de Haar est stocké.

• **Description par modèles de Markov cachées :** Castellani et al. (Castellani, et al., 2008) suggèrent une description locale basée sur les modèles de Markov cachés (HMM). L'originalité de cette approche d'apprentissage est l'expression des variations de la géométrie locale du modèle dans le domaine 3D. Le HMM est entraîné d'une façon non supervisée en utilisant l'information contextuelle du voisinage 3D.

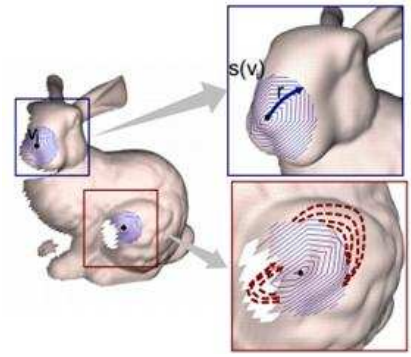


Figure 2-31. Description des points d'intérêt : à gauche, les positions de deux PIs en noir avec leur chemin spiral. En haut à droite, un zoom sur le point v_i dont le chemin spiral est limité par le rayon géodésique r. En bas à droite, la construction d'une spirale en présence d'un trou dans la maille. Les pointillés rouges correspondent à la correction pour arranger la spirale (Castellani, et al., 2008)

• **Spin image** Ce descripteur de surface (dont le principe a été brièvement présenté dans 2.1.2.2) est utilisé dans un contexte de points d'intérêt. Li et Guskov (Li, et al., 2007) appliquent ce descripteur au niveau des points saillants extraits des images de profondeur. Dans leur méthode, un cylindre construit autour de chaque point saillant est subdivisé en K x L cellules. Un histogramme des distances entre le point sélectionné et ses voisins contenus dans les K x L cellules est utilisé comme signature de la zone. Une Analyse en Composantes Principales permet la réduction de la dimension des spins images. Le temps de calcul du spin dépend de la résolution de la scène. Plus de détails sur la formation de ce descripteur seront présentés dans le chapitre Descripteurs.

• **Descripteurs SHOT et CSHOT :** Le SHOT (Signature of Histograms of OrienTations), proposé par Tombari et al. (Tombari, et al., 2010), fait partie de la catégorie des descripteurs basés sur les histogrammes de normales. La performance du SHOT a été prouvée dans une comparaison avec le spin image, l'Exponential Mapping (EM) et le Point Signatures (PS). Ce descripteur, en plus de son pouvoir descriptif de la forme de la surface 3D d'une façon répétable et robuste au bruit, occultation, translations et rotations, présente un énorme gain en temps de calcul. La définition d'un système de référence local 3D unique et non ambiguë (RF). L'introduction de l'information géométrique sur les positions des points contenus dans un support est faite par un ensemble d'histogrammes 3D locaux définis sur une grille 3D superposée sur le support. Pour chaque secteur de la grille, les valeurs des cosinus des angles entre la normale référence (celle du PI) et celles des ces voisins sont cumulées pour former l'histogramme des normales. L'estimation des normales se fait par le calcul d'une nouvelle matrice de covariance comme combinaison linéaire des distances des points appartenant à un support sphérique du PI. Les vecteurs propres de cette matrice forment des directions orthogonales, répétables et robustes au bruit. Plus récemment, la version CSHOT (Tombari, et al., 2011) du SHOT combine une information sur la forme et sur la texture. Ce descripteur est une combinaison d'un histogramme de normales et d'un histogramme de couleurs. L'histogramme de couleurs est formé par les valeurs de la somme absolue (norme L_1) entre les triplets RGB du point référence et le triplet des points voisins. Le descripteur final est la juxtaposition de ces histogrammes. Ces deux descripteurs feront parties du chapitre Descripteurs.

2.4. Mesures de la performance

Afin de clôturer ce chapitre, nous proposons de présenter quelques mesures utilisées dans la littérature pour quantifier la performance d'un système de reconnaissance.

Evaluer une méthode de reconnaissance revient à attribuer une « note » à la corrélation (le degré de similitude entre l'échantillon candidat et un échantillon prélevé précédemment). La note obtenue par l'échantillon candidat doit dépasser une valeur de seuil déterminée pour que l'échantillon soit déclaré «correspondant ». Cette corrélation est une fonction statistique dépendant du type de technologie, du matériel, de l'algorithme et des réglages opérationnels du système. Une même corrélation peut être ou ne pas être bonne suivant le contexte. Nous distinguons principalement quatre mesures de performance utilisées pour l'évaluation du système d'identification :

* Taux de bonne reconnaissance : mesure le pourcentage pour lequel le système parvient à identifier les objets.

* Taux de faux rejets (False Reject Rate) : indique dans quelle mesure un système donné ne réussit pas à apparier des échantillons provenant du même objet.

* Taux de fausses acceptations (False Acceptance Rate) : indique dans quelle mesure un système donné apparie des échantillons ne provenant pas du même objet.

* Point d'équivalence des erreurs : ou taux d'exactitude croisée, est déterminé par le point d'intersection entre la courbe du taux de fausses acceptations et la courbe du taux de faux rejets. Le taux d'erreur est plus faible au point d'intersection.

* Précision et rappel : D'autres mesures peuvent être utilisées pour évaluer la performance de la reconnaissance: Précision (precision) = TP/ (TP+FP); Rappel (recall) ou sensibilité = TP/ (TP+FN) ; Accuracy = (TP+TN)/(TP+FN+FP+TN); Spécificité = TN/ (FP+TN) ; avec la matrice de confusion suivante :

Réel	Pos	Neg
Pos	TP : True Positive	FP : False Positive
Neg	FN : False Negative	TN : True Negative

Tableau 2-8. Matrice de confusion

Le rappel mesure la capacité du descripteur à sélectionner tous les modèles pertinents, alors que la précision mesure la capacité du système à rejeter tous les modèles non pertinents. La méthode Précision-Rappel est utilisée dans la littérature pour évaluer la performance des classificateurs et des systèmes d'indexation et de recherche par le contenu.

* Courbe ROC : Pour comparer des algorithmes d'apprentissage indépendamment de la distribution des modalités de la variable à prédire et des coûts d'erreur, une courbe ROC (Receiver Operating Characteristics) est définie. Cette courbe indique la capacité du classificateur à placer les positifs

devant les négatifs. Elle met en relation dans un graphique le taux de faux positifs chez les NEG en abscisse (1 - spécificité: erreur de prédiction) et les taux de vrais positifs (sensibilité: en ordonnée). L'intérêt de cette courbe est de visualiser rapidement la puissance d'un modèle.

* Mean Average precision (MAP) (Jones, et al., 1976): La moyenne des précisions pour chaque requête est calculée et la valeur moyenne des moyennes des précisions sur toutes les requêtes donne la MAP.

* Temps de calcul est une mesure de performance très importante surtout dans le cas des applications temps réel.

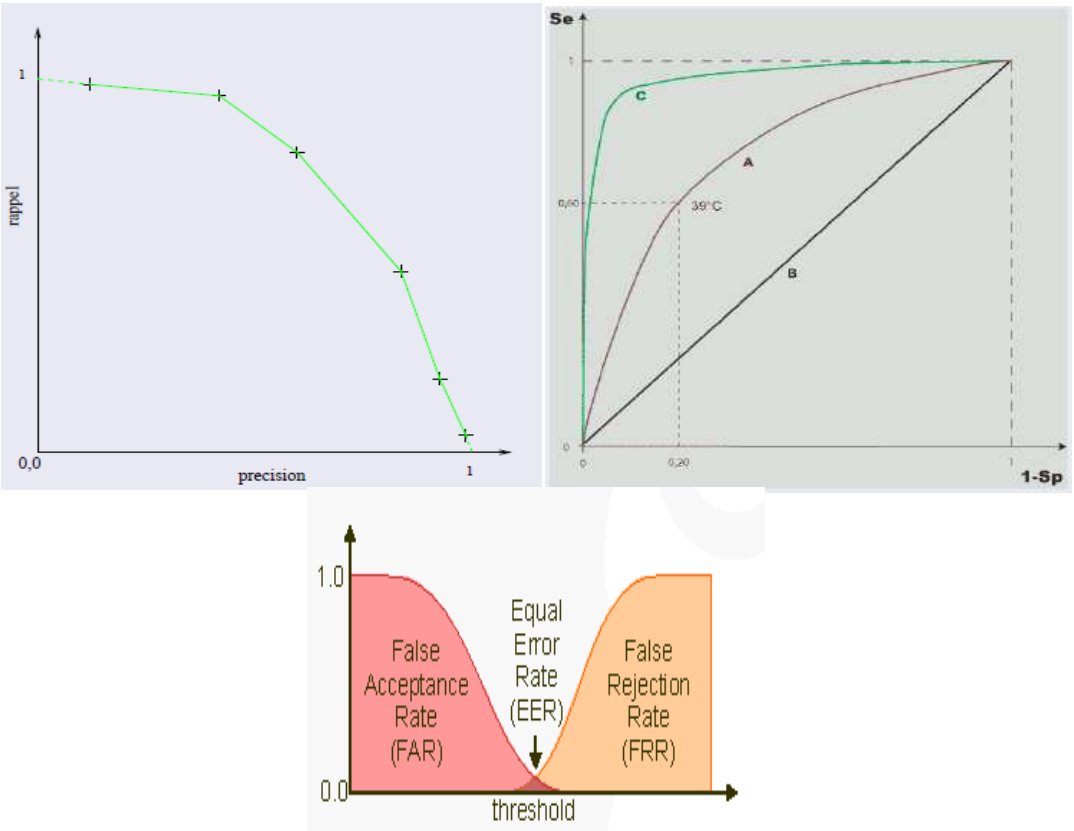


Figure 2-32. Courbes présentant: 1-précision-rappel- 2- ROC -3- Taux de reconnaissance

2.5. Conclusions

Dans ce chapitre, nous avons vu comment des travaux de recherche ont approché le domaine de reconnaissance d'objets 3D. Nous avons classé ces approches en des sous-catégories: 3D ou 2D/3D et locale ou globale. Ensuite, nous avons mis l'accent sur les techniques locales qui ont présenté des solutions pour des problématiques comme la partialité de l'information, présence d'occultation et de bruit, variation intra- et inter-classes des objets, variation de la topologie, compacité de la représentation, etc. La sous-catégorie des approches par points d'intérêt est la plus utilisée. Dans ce volet, nous avons exposé le principe de quelques détecteurs que nous avons classés en détecteurs à échelle fixe et détecteurs invariants à l'échelle, et selon sur quoi ces détecteurs opèrent, nous avons introduit les typologies : détecteurs sur carte 2D et détecteurs sur représentation 3D. Parmi ces

détecteurs, ceux qui ont inspiré le plus notre travail sont : Local Surface Patches (LSP) (Chen, et al., 2007), Classification selon SI (Hozatlı, 2009), KeyPoint Quality (KPQ) (Mian, et al., 2009), Harris 3D (Sipiran, et al., 2010), SC et HK multi- échelle (Akagunduz, et al., 2009), SURF 3D (Knopp, et al., 2010), Courbure multi- échelle (Ho, et al., 2009), THRIFT (Flint, et al., 2007).

Dans un deuxième volet, celui de l'étape de l'attribution d'une signature aux points d'intérêt, nous nous sommes intéressés plus particulièrement aux descripteurs suivants : LSP, THRIFT, SURF 3D, Spin (Johnson, et al., 1999), SHOT (Tombari, et al., 2010) et CSHOT (Tombari, et al., 2011).

Les méthodes mentionnées dans ce paragraphe seront détaillées dans les chapitres suivants, et forment une base de référence et de comparaison pour nos contributions aux techniques de détection, de description et de reconnaissance pour les images 3D.

3. CHAPITRE : DÉTECTION

Le but de l'extraction de primitives est de passer de l'espace initial des données à un espace de représentation compacte et riche en information. Des moyens comme la morphologie mathématique, la détection des contours et de coins, ou les courbures sont utilisés pour caractériser les propriétés locales de l'espace des données initiales. Selon l'application finale, la fiabilité d'une représentation de primitives est fortement rattachée à sa stabilité pour des variations liées aux transformations de l'espace, à la résolution, au bruit et aux occultations. En particulier, les courbures de la surface renseignent sur la forme de l'objet et sont robustes à des variations d'angles de vues. Dans cette thèse, pour caractériser la forme des objets dans des vues partielles, nous adoptons une approche de détection de points d'intérêt 3D basée sur les courbures de la surface. Dans ce chapitre, nous exposons, dans une première partie, une sélection de méthodes de détecteurs inspirés de la littérature et nous introduisons notre principale contribution. Dans une deuxième partie, nous mettons l'accent sur notre apport à travers les comparaisons avec les détecteurs présentés. Les notions de bases utilisées dans ce chapitre ont été présentées dans les deux sections précédentes.

3.1. Principes des détecteurs testés

3.1.1. Seuillage des valeurs de SI

Le détecteur, que nous avons nommé « SI », est un détecteur à échelle fixe proposé par (Chen, et al., 2007) basé sur un seuillage des valeurs de l'indice de forme (Équation 1.13) des points. Les points d'intérêt sont sélectionnés aux points ayant une valeur SI localement extremum par rapport aux SI de leur voisinage. Un point p est considéré comme point d'intérêt si son SI_p est un extremum local des SI de son voisinage et est significativement plus grand ou plus petit que la moyenne des SI sur le voisinage. Mathématiquement, SI_p doit satisfaire l'Équation 3.1:

$$\begin{cases} SI_p = min(SI_k)\,;k \in voisinage\ \ et\ \ SI_p \leq\ (1-\beta)\times\mu \\ SI_p = max(SI_k)\,;k \in voisinage\ \ et\ \ SI_p \geq\ (1+\alpha)\times\mu \end{cases} \quad 0 \leq \alpha, \beta \leq 1$$

Équation 3.1

Dans cette équation, μ est la moyenne des indices de formes sur les valeurs voisines. Notons que dans Équation 3.1, les paramètres α et β contrôlent le nombre de points d'intérêt sélectionnés. La stabilité au bruit et aux transformations de la surface des résultats d'une classification basée sur SI, est notable par rapport à une classification basée uniquement sur les courbures gaussiennes et moyennes. De plus, contrairement à ces deux derniers estimateurs dépendant de seuils pour détecter les valeurs de courbure proches de zéro, la formulation de SI sépare les types de surfaces par des seuils fixes. Avec

ce détecteur, la sélection de PIs se fait sur les zones convexes (les maximums en SI) et concaves de la forme (les minimums en SI). Dans Figure 3-1, les pixels clairs représentent les formes convexes et les pixels sombres indiquent les formes concaves.

Dans le travail proposé par (Chen, et al., 2007), le calcul des valeurs de SI se fait sur l'image de profondeur. Nous proposons de faire ce calcul directement sur le nuage de point en passant par un maillage qui s'oriente par l'organisation 2D de l'image de profondeur.

Dans la sélection de PIs, les valeurs de SI du voisinage sont pris en compte ce qui permet de faire une meilleure sélection basée sur la géométrie locale de la forme. La taille du voisinage reste un paramètre décisif pour cette approche créant un compromis entre l'efficacité de la détection et le temps de calcul.

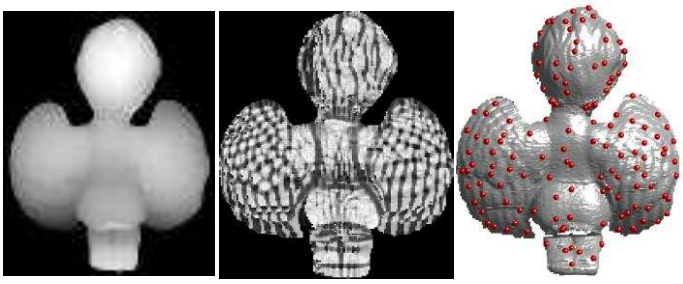


Figure 3-1-Gauche-Image de profondeur du modèle d'un ange - Milieu- Image de l'indice de forme en niveau de gris–Droite- Les PIs détectés.

3.1.2 Sélection des Points Harris 3D

Après avoir associé une valeur de l'opérateur de Harris à chaque sommet, les auteurs (Sipiran, et al., 2010) proposent deux façons de sélectionner les points d'intérêt. En premier, les maximums locaux des réponses de Harris h sont extraits, ce qui correspond aux sommets v vérifiant la condition suivante :

$$h(v) > h(w),\ \forall w \in ring_1(v)$$

Le $ring_1$ correspond aux sommets directement liés au sommet v dans le maillage. Ensuite, deux approches sont proposées pour sélectionner l'ensemble des PIs finaux :

· **Sélection des points ayant la plus grande réponse de Harris**: Où il est question d'extraction d'une portion constante de points d'intérêt. L'extraction, seulement, des points de grande saillance peut pénaliser certaines parties de l'objet.

· **Regroupement des PIs** (Clusters). Cette approche permet d'obtenir une distribution des PIs sur toute la surface de l'objet. L'ensemble des points présélectionnés selon la valeur de l'opérateur de Harris sont ordonnés dans un ordre décroissant puis, l'algorithme de regroupement (clustering) suivant est leur appliqué :

Algorithm 1 Interest Points Clustering

Require: Set P of pre-selected interest points in decreasing order of Harris operator value
Ensure: Final set of interest points

1: Let Q be a set of points
2: $Q \leftarrow \emptyset$
3: **for** $i \leftarrow 1$ to $|P|$ **do**
4: **if** $min_{j \in [1,|Q|]} \|P_i - Q_j\|_2 > \rho$ **then**
5: $Q \leftarrow Q \cup \{P_i\}$
6: **end if**
7: **end for**
8: Return Q

Figure 3-2 Algorithme de regroupement des points d'intérêt (clustering)

La valeur de ρ peut être une fraction de la diagonal de la boite englobant de l'objet et sa valeur détermine le nombre de PI final. Nous avons utilisé le code de Harris 3D donné par l'auteur pour notre comparaison des performances des détecteurs. Les paramètres réglables par l'utilisateur de cet algorithme concernent la taille du voisinage (nombre de rings), le paramètre K pour le calcul de la réponde de Harris des points, le nombre de rings pour la détection des maximas et le type de l'algorithme de sélection final (Fraction ou Clustering) avec la valeur du paramètre de sélection. Pour la suite, nous allons appeler le Harris qui utilise la sélection par Fraction « Harris_fract » et celui qui utilise le regroupement « Harris_clust ».
La Figure 3-3 représente un exemple du résultat de la sélection des points d'intérêts avec les deux détecteurs Harris_fract et Harris_clust.

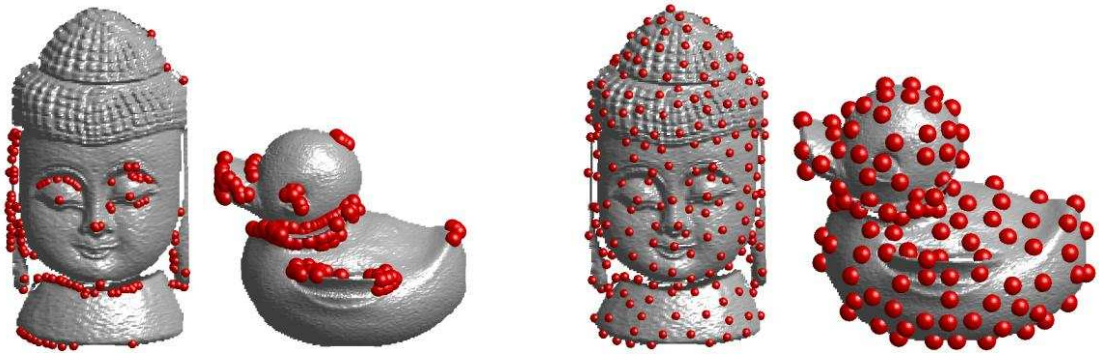


Figure 3-3- A gauche -Sélection des points avec la plus grande réponse d'Harris et - A droite - Sélection des points Harris avec regroupement.

3.1.3 SURF 3D

Un autre détecteur de l'état de l'art faisant une extension d'une méthode connue en 2D est le SURF 3D introduit par (Knopp, et al., 2010). Différemment aux autres approches muti-échelles, le 3D SURF construit des espaces d'échelle à partir de la version **voxélisée** du maillage original. Une mesure de saillance S est définie pour chaque bin de la grille cubique sur plusieurs échelles (trois octaves). S

correspond à la valeur absolue du déterminant de la matrice Hessienne H des dérivées gaussiennes secondes L, déduites des moyennes des box de filtrage.

$$S(\boldsymbol{x},\sigma) = \left|H(\boldsymbol{x},\sigma)\right| = \left|\begin{pmatrix} L_{xx}(\boldsymbol{x},\sigma) & L_{xy}(\boldsymbol{x},\sigma) & L_{xz}(\boldsymbol{x},\sigma) \\ L_{yx}(\boldsymbol{x},\sigma) & L_{yy}(\boldsymbol{x},\sigma) & L_{yz}(\boldsymbol{x},\sigma) \\ L_{zx}(\boldsymbol{x},\sigma) & L_{zy}(\boldsymbol{x},\sigma) & L_{zz}(\boldsymbol{x},\sigma) \end{pmatrix}\right|$$

Équation 3.2

Les valeurs propres de la matrice H ne sont tous de même signe, ce qui correspond à des points détectés sur des zones sous forme de blobs et de selles. La sélection finale des PIs est faite avec un algorithme de suppression non-maximale. Chaque point extrait lui est associée une échelle.

Dans notre étude comparative, nous allons utilisée la version (exécutable) fournie par les auteurs et qui calcule le détecteur et le descripteur à la fois. Les paramètres fixés par l'utilisateur sont : la dimension du cube englobant l'objet, la distance entre les bords du cube et le modèle, et le seuil utilisé par la suppression non-maximale. Nous remarquons dans l'exemple donné dans la Figure 3-4, que les positions des PIs SURF ne sont pas forcement sur la surface de l'objet, étant donné qu'elles correspondant à des centres de voxels.

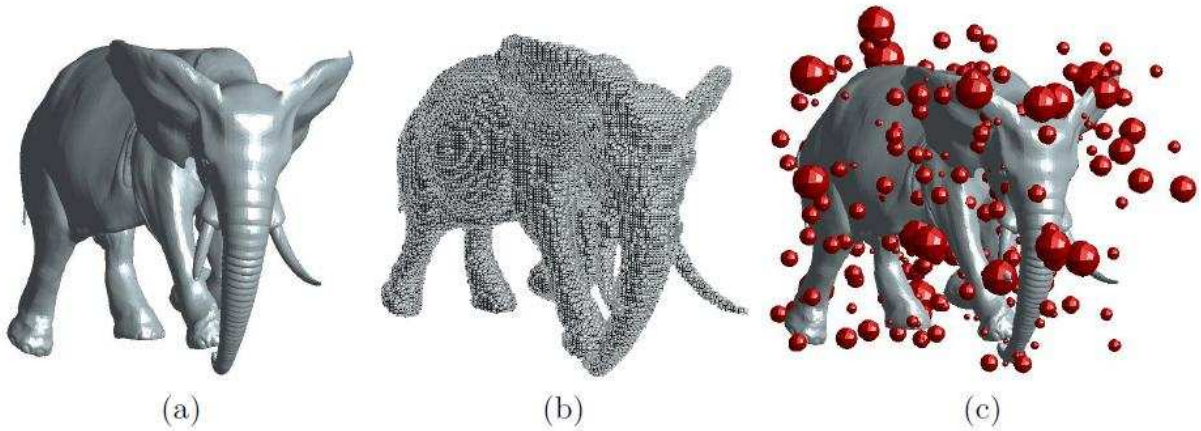


Figure 3-4 Illustration des PIs détectés par le SURF 3D. La forme (a) est voxélisée en une grille cubique (b). Les PIs détectés (c), et représentés par des sphères dont le rayon illustre l'échelle. (Knopp, et al., 2010)

3.1.4 Facteur de qualité

Le troisième détecteur implémenté est basé sur une mesure de qualité du PI introduite par (Mian, et al., 2009) qui l'ont utilisée pour classer les PIs détectés. En chaque point, nous associons la mesure de qualité Q_k que nous rappelons sa formulation:

$$Q_k = \frac{1000}{r^2} \sum |K| + \max(100K) + |\min(100K)| + \max(10\,k_p{}^1) + |\min(10k_p{}^2)|\,;\;\; K = k_p{}^1 k_p{}^2$$

Équation 3.3

Dans cette équation, $K = \kappa_1\kappa_2$ est la courbure gaussienne et $k_p{}^1 = \kappa_1\ et\ k_p{}^2 = \kappa_2$ sont les courbures principales.

Les fonctions de sommation, maximum et minimum sont calculées sur le voisinage du point. Le pouvoir descriptif des valeurs de courbure étant indépendant de leur signe, ce sont les valeurs absolues qui sont utilisées dans la formule, afin d'éviter que des contributions de signes opposés ne s'annulent. Les valeurs des constantes multiplicatives sont choisies empiriquement pour donner le poids approprié à chaque terme. Pour des surfaces lissées, les valeurs des courbures principales sont en général plus petites que 1 et comme la courbure gaussienne est le produit des courbures principales, elle est alors d'ordre 10 fois plus petite que les courbures principales. Cette relation a été confirmée expérimentalement par (Mian, et al., 2009). Ainsi, le terme multiplicatif pour K est choisi 10 fois plus grand que les courbures principales. Les données d'apprentissage ont montré que le premier terme de l'équation est 100 fois plus petit que les courbures principales. En effet, la valeur de la courbure gaussienne, pour la plupart des points sur la surface lissée, est faible et qui une fois moyenner sur la surface donne une valeur qui est 10 fois plus petite que les valeurs maximum et minimum absolu de la courbure gaussienne. Pour finir, tous les termes de l'équation sont multipliés par 10 pour avoir une mesure de qualité Q toujours supérieure à 1.

Après le calcul des valeurs du facteur qualité de tous les points du nuage, nous les ordonnons selon leur valeur Q et les groupons à l'aide de l'algorithme 1 de clustering (Figure 3-2). L'image de droite dans la Figure 3-5 illustre la carte des valeurs du facteur qualité des points du modèle de l'ange. Les pixels les plus clairs correspondent aux valeurs de Q les plus importantes et sont, en effet, situés aux zones les plus descriptives (à fortes variations locale de forme par rapport au voisinage).

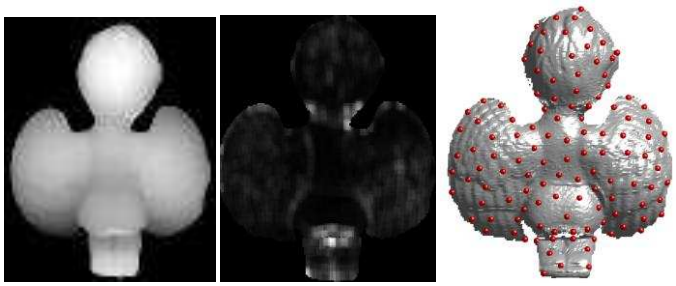


Figure 3-5-A gauche-Image de profondeur d'un modèle d'ange –Au milieu- L'image correspondant aux valeurs de la mesure facteur de qualité, en niveau de gris- A droite- Les PIs détectés.

3.1.5 La classification HK et SC

Rappelons que l'idée ici est de construire un espace de classification de formes en utilisant la paire des courbures Moyenne-Gaussienne (HK) ou le couple Indice de forme-intensité de Courbure (SC). Le type d'une surface locale est reconnu par l'intermédiaire de la valeur de la paire considérée. Ces couples sont géométriquement invariants et renseignent sur l'ordre second de l'information géométrique de la forme.

Typiquement, pour la classification HK, la fonction type T_p (Equation 1.9) associe à chaque valeur du couple H et K un type unique correspondant à une forme particulière. Neuf valeurs de type sont considérées. De plus, deux seuils sur les valeurs de H et K permettent de classer les surfaces planes.
Dans l'espace Indice de forme-Intensité de Courbure (SC), l'indice de forme SI renseigne sur la forme et l'intensité de Courbure C définie le dégrée de courbure. Comme l'illustre bien la Figure 3-6, les détails de la forme sont visuellement accentués.

Pareillement, que la représentation HK, SI permet d'avoir une graduation continue et l'espace SC subdivise les formes de surface en 9 catégories. Dans la Figure 3-7 illustrant cette catégorisation, les droites séparatrices sont des valeurs constantes de S: S = ±5/8, ±3/8 et ±3/16. Le cercle autour de l'origine est le seuil pour les valeurs de C à zéro et d'équation: $\sqrt{(\kappa_1^2 + \kappa_2^2)} = C_{zéro}$.

Nous utilisons ces deux classifieurs pour extraire les régions saillantes qui sont sélectionnées au niveau des surfaces en forme de dôme, de cuvette, de sphère, de selle ornière et de selle crête, et qui correspondent aux primitives primaires que l'humain peut observer. Un grand nombre de PIs résulte suite à cette extraction. Nous proposons dans notre implémentation de ces détecteurs d'appliquer un processus de filtrage des points selon leur distance relative par l'algorithme 1 de clustering. Nous désignons ces deux détecteurs par « HK » et « SC ».

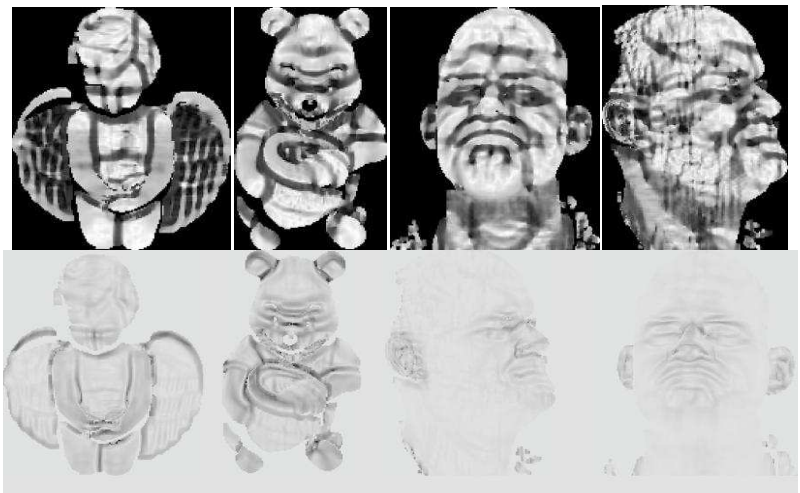


Figure 3-6 Images correspondant à -en haut- mesures de SI et -en bas- mesures de C

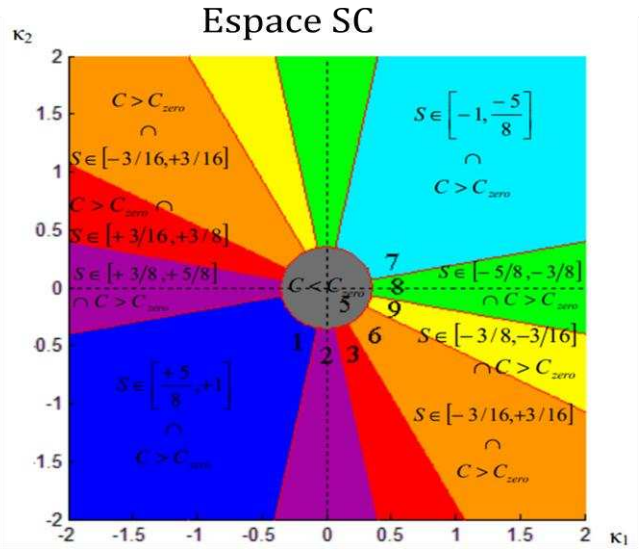


Figure 3-7 Classification SC dans le plan (κ_1, κ_2) (Akagündüz, 2011)

3.1.6 Combinaison des critères

a) Version mono-échelle

Théoriquement, les deux classifications SC et HK convergent vers le même résultat. Cependant, comme nous pouvons le remarquer sur les graphes de la Figure 3-7 de l'espace SC et ceux de l'espace HK, des différences dans la classification de quelques régions subsistent. Par exemple, la région plane, définie dans l'espace HK (Figure 1-10), est caractérisée autrement dans l'espace SC (régions 1, 2, 5, 9, 10, 11, 14). Pareillement, les régions elliptiques et paraboliques se diversifient dans les deux espaces quand les valeurs de courbures principales augmentent. Une confusion entre les deux types de surface est remarquée pour les régions 4, 6, 13, 15. Et finalement, la définition des régions 3, 7, 8, 12, 16, 17 restent ambigus.

De ce fait, nous suggérons, dans un premier temps, de combiner les deux espaces en un espace couplé HK&SC pour augmenter l'efficacité du résultat de classification (

Figure 3-8 et Figure 3-9). Seulement les régions communes sont classées. En d'autres termes, nous ne sélectionnons un PI que s'il est saillant pour les deux classifications SC et HK. L'intersection garde les points dont les deux labels du type correspondent à une région saillante (dôme, sphérique ou selle). Au final, notre résultat est validé par deux mesures de détection de PIs et nous réduisons le nombre de point représentatifs de la forme initiale (Figure 3-10). Quant au type de régions choisies comme saillantes, nous les définirons selon la nature des objets qui constituent la base de données. Par exemple, si des objets dont les surfaces sont planes existent, nous allons sélectionnés, outre les points déjà mentionnés, les points de courbure nulle.

Convex (Elliptic) (1)	$H<-\lvert H_{zero}\rvert \cap K>+\lvert K_{zero}\rvert \cap S\in[+5/8,1] \cap C>C_{zero}$
Convex (Parabolic) (2)	$H<-\lvert H_{zero}\rvert \cap K<\lvert K_{zero}\rvert \cap S\in[+3/8,+5/8] \cap C>C_{zero}$
Saddle Ridge (3)	$H<-\lvert H_{zero}\rvert \cap K<-\lvert K_{zero}\rvert \cap S\in[+3/16,+3/8] \cap C>C_{zero}$
Planar (5)	$H<\lvert H_{zero}\rvert \cap K<\lvert K_{zero}\rvert \cap C<C_{zero}$
Hyperbola (6)	$H<\lvert H_{zero}\rvert \cap K<-\lvert K_{zero}\rvert \cap S\in[-3/16,+3/16] \cap C>C_{zero}$
Concave (Elliptic) (7)	$H>+\lvert H_{zero}\rvert \cap K>+\lvert K_{zero}\rvert \cap S\in[-1,-5/8] \cap C>C_{zero}$
Concave (Parabolic) (8)	$H>+\lvert H_{zero}\rvert \cap K<\lvert K_{zero}\rvert \cap S\in[-5/8,-3/8] \cap C>C_{zero}$
Saddle Valley (6)	$H>+\lvert H_{zero}\rvert \cap K<-\lvert K_{zero}\rvert \cap S\in[-3/16,+3/16] \cap C>C_{zero}$

Figure 3-8 Formulation mathématique de la classification du couple HK&SC (Akagündüz, 2011)

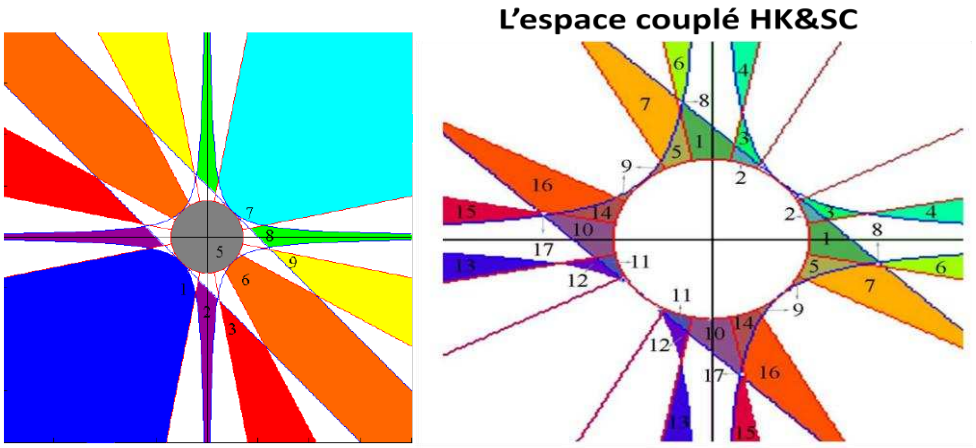


Figure 3-9 L'espace couplé HK&SC de classification dans le plan (k_1, k_2) (Akagündüz, 2011)

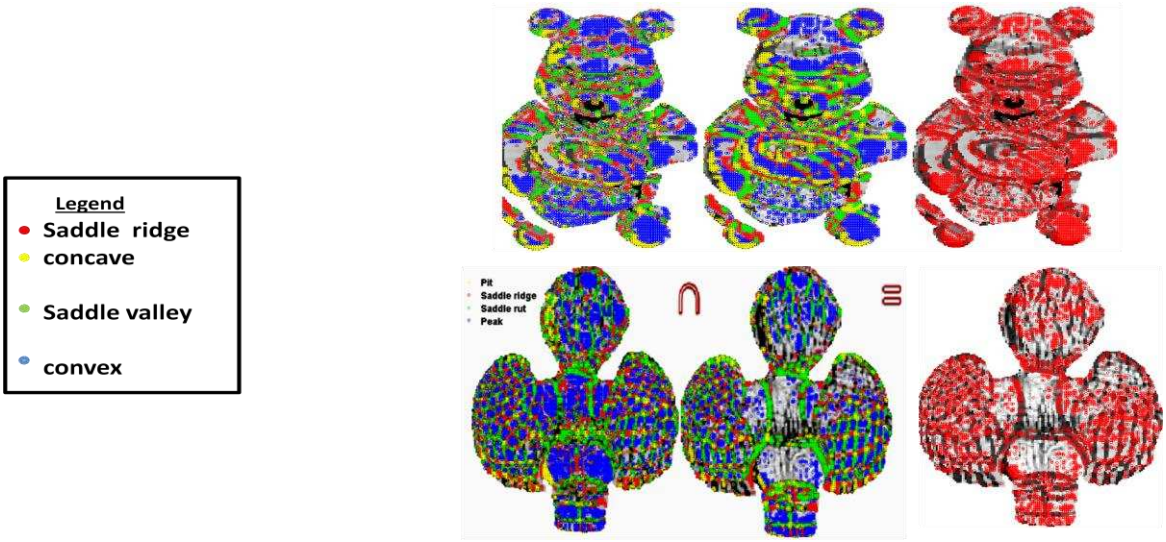


Figure 3-10 Résultat du processus de croisement de deux espaces SC et HK sur deux modèles d'objets.

Nous proposons, dans un deuxième temps, de rajouter aux critères de sélection SC et HK, un critère de tri. Nous parcourons le nuage point par point et sélectionnons les points répondant au critère de saillance de la paire (SC, HK). La sélection finale des PIs pour le détecteur SC_HK se fait de deux façons. :

- La première façon de faire est de trier l'ensemble des PIs trouvé selon une valeur de mesure de tri puis d'appliquer le principe de filtrage des PIs basé sur l'écart spatial des coordonnées 3D (algorithme 1de clustering Harris) sur l'ensemble des points ordonnés. Nous avons utilisé un premier critère de tri basé sur l'intensité de courbure C pour sélectionner les valeurs maximales. Nous appelons la combinaison SC_HK avec le critère de tri sur C, détecteur « SC_HK _C». Nous introduisons un deuxième détecteur qui classe les PIs selon la valeur du facteur qualité FQ des points. Nous notons ce dernier détecteur « SC_HK_FQ ». Un troisième détecteur basé sur un tri avec une valeur de confiance γ utilisée par Ho et Gibbins (Ho, et al., 2009) et qui mesure la déviation de leur intensité de courbure

par rapport à celle du voisinage. La formule de cet indice est donnée dans la Figure 3-11. Nous notons ce dernier détecteur « SC_HK_Conf» et appelons la mesure γ par indice « Conf ».

$$\gamma(\mathbf{p}, r_k) = \frac{|c_{\mathbf{p}} - \mu_{N_{\mathbf{p}}}|}{\sigma_{N_{\mathbf{p}}}} \qquad \begin{aligned} \mu_{N_{\mathbf{p}}} &= \frac{\sum_{\mathbf{p}_j \in N_{\mathbf{p}}} c_{\mathbf{p}_j}}{n} \\ \sigma_{N_{\mathbf{p}}} &= \sqrt{\frac{\sum_{\mathbf{p}_j \in N_{\mathbf{p}}} (c_{\mathbf{p}_j} - \mu_{N_{\mathbf{p}}})^2}{n-1}} \end{aligned}$$

Figure 3-11 Formule de l'indice de confiance γ basée sur l'intensité de courbure C (Ho, et al., 2009)

- Pour la deuxième sélection, nous procédons à un étiquetage de chaque point du nuage de points selon le critère considéré, puis nous extrayons toutes les composantes connexes. La sélection finale se fait en prenant le centre des composantes qui répondent au critère de tri final. Nous allons considérer uniquement le critère basé sur la valeur maximale en C et qui donne le détecteur « SC_HK_C_Con ». Nous introduisons la notion de labellisation des composantes connexes comme suit :

- **Labellisation des composantes connexes**

L'étiquetage des composantes connexes est utilisé pour exprimer la connexité ou grouper les pixels connectés dans une image. La connexité est la relation entre deux pixels ou plus. Deux pixels sont dits connexes s'ils sont adjacents et de valeur de pixel égale. 4 ou 8- connexités sont les plus utilisées. Pour la connexité d'ordre 4, les 4 voisins (dessus, dessous, gauche et droite) sont vérifiés. Pour la connexité 8, s'ajoute une vérification des voisins dans les 4 diagonales. Le groupement des pixels connexes forme une composante connexe. Chaque composante connexe est étiquetée par un nombre différent. Deux pixels p_1 et p_2 sont connectés s'il existe un chemin les reliant. Durant le processus de parcours des pixels, pour un pixel p_1 et ses quatre voisins, si un des voisins a été déjà étiqueté dans les étapes précédentes, p_1 est étiqueté comme son voisin. Et si aucun pixel voisin n'a été étiqueté, p_1 est marqué par une nouvelle étiquette. Ce processus est répété pour toute l'image. Au final, les composantes connexes sont séparées.

- **Algorithme général de combinaison**

Nous présentons les étapes générales de l'algorithme de combinaison des critères :

- Pour chaque point du nuage, générer le maillage (détection de la boite englobante + calcul de la distance moyenne entre les points, utilisée pour construire le maillage)
- Détection des points du bord
- Calculer les mesures de saillances sur chaque patch (SC et HK) avec un voisinage N_p proportionnel à la diagonale de la boite englobante et extraction des points saillants
- Calculer une valeur de mesure (C ou FQ ou Conf) des PIs
- Regrouper les points détectés :
  - sous forme de composantes connexes (label=types selon SC- Type selon HK) : Garder les composantes connexes les plus larges et prendre le PI ayant la plus grande intensité de courbure C comme représentant
  - ordonner et regrouper avec l'algorithme1 l'ensemble des PIs selon l'un des critères : C ou FQ ou Conf

Contrairement au tri selon le critère C calculé sur points PIs seulement, les mesures FQ et Conf des PIs font intervenir les mesures de courbure du voisinage N_p de ces points. Un paramètre, ParSelect_PIs

qui est un ratio permettant de prendre une proportionnel à la taille de la diagonale de l'objet, est utilisé pour filtrer les PIs selon la distance spatiale entre eux.

Dans la Figure 3-12, le processus d'extraction des PIs en utilisant le détecteur SC_HK_C_Con est exposé. Nous notons que l'étape des composantes connexes permet de réduire considérablement le nombre final de PIs. Les représentants de chaque composante connexe sont situés sur des régions saillantes du visage (nez, front, bouche, menton, joue, oreille) puisqu'ils sont choisis sur les valeurs de courbure maximale.

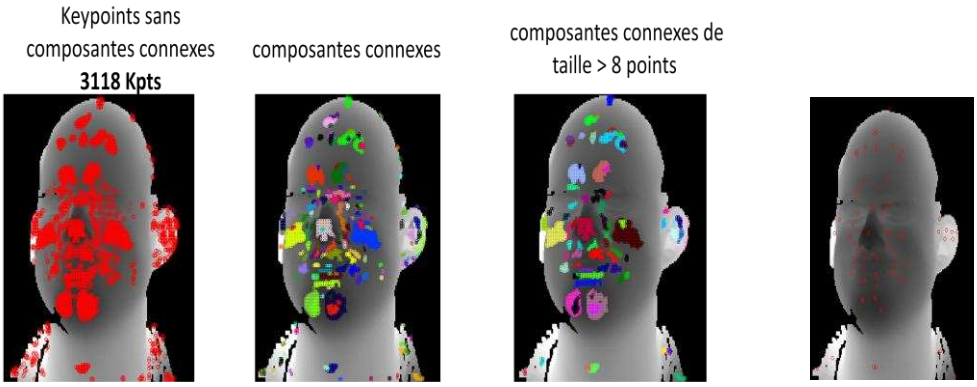


Figure 3-12 Illustration du processus d'extraction des PIs en utilisant le détecteur SC_HK_C_Con

b) Version Multi-échelle

Similairement à l'idée introduite par l'étude de (Ho, et al., 2009), nous proposons, pour une version multi-échelles de notre travail, de s'appuier sur l'augmentation de la taille du voisinage au lieu d'appliquer un lissage par une pyramide de gaussiennes. Le principe est de calculer la mesure de saillance sur différents niveaux de voisinages (échelles) et de sélectionner le niveau pour lequel la valeur de saillance est extremum par rapport aux valeurs de ses voisins à l'échelle en question et aux valeurs des échelles voisines. L'algorithme adopté est alors le suivant :

- **Algorithme de la version Multi-échelle**
  - **Données :**

  $P = \{p_i \in R^3\}$: ensemble des points 3D du modèle.

  $R = \{r_k\}$: un ensemble d'échelles.

  - **Algorithme :**

  1: **for $r \in \{r_k\}$ do**

  2: **for $p \in \{p_i\}$ do**

  3: Trouver le voisinage N_r à l'échelle r

  4: Calculer les mesures de saillance basées sur k_1, k_2, SI, C, H, K

  5: Vérifier si p est saillant avec le critère (SC ∩ HK) à l'échelle r

6: **end for**

7: Sélectionner les PIs de position p saillante en termes de (FQ ou C ou Conf) par rapport à son voisinage sur les rayons r_k, r_k -1 et r_k +1 et ayant une mesure de saillance maximale à l'échelle r_k

8: Regrouper les PIs sous forme de composantes connexes et sélectionner pour chaque composante les PIs de courbure maximale C

ou

8: Trier les PIs selon leur saillance (FQ ou C ou Conf) puis appliquer un clustering

9: **end for**

Dans les tests effectués, nous n'avons considéré que la méthode avec les composantes connexes de l'étape 8. Par contrainte de temps, nous n'avons pas pu mener à bout cette version multi-échelles et exploiter ces résultats. Nous nous sommes contentés de quelques résultats visuels qui seront présentés ultérieurement.

3.2. Eléments impactant sur les mesures de courbure

3.2.1 Artefacts et bruit

En optique, une surface claire réfléchit plus de lumière qu'une surface sombre. C'est ce qui fait que pour les scanners optiques certaines zones de la scène comme les cheveux noirs ou des miroirs ne sont pas correctement détectés. Les réflexions multiples de la surface donnent des points aberrants (outliers). De plus, les limitations physiques des capteurs peuvent ajouter du bruit ou des artéfacts et affecte l'échantillonnage et les valeurs de profondeurs des points. Le type de bruit des scanners 3D est typiquement gaussien. Similairement au bruit 2D qui est convolué avec l'intensité des pixels, le bruit 3D perturbe les coordonnées des points du nuage. Un exemple de l'impact de bruit sur le calcul des normales est présenté dans la Figure 3-13.

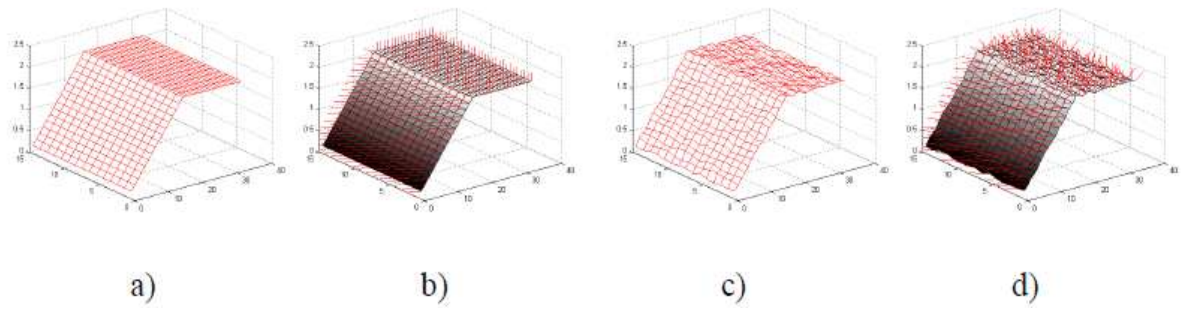


Figure 3-13 a) Surface 3D artificielle sans bruit b) Directions des normales sur la surface non bruitée c) Surface 3D artificielle avec bruit. d) Directions des normales sur la surface bruitée (Bozkurt, et al., 2009)

La surface de l'objet scanné étant inconnue, il est très difficile de décider de la validité ou non-validité des points. Le bruit peut être partiellement enlevé avec par exemple un filtre bilatéral ou un lissage Laplacian. Le filtre bilatéral diminue le bruit tout en préservant les détails de la forme 3D contrairement à un lissage Laplacian qui ne lisse pas seulement le bruit mais aussi les caractéristiques de la forme. Une comparaison entre le lissage Laplacian (implémenté dans la bibliothèque VTK) et le filtrage bilatéral gaussien (Fleishman, et al., 2003) est présenté dans la Figure 3-14. Nous notons la perte des détails de la forme autour de l'œil de bunny sur le modèle lissé par rapport au modèle dé-bruité.

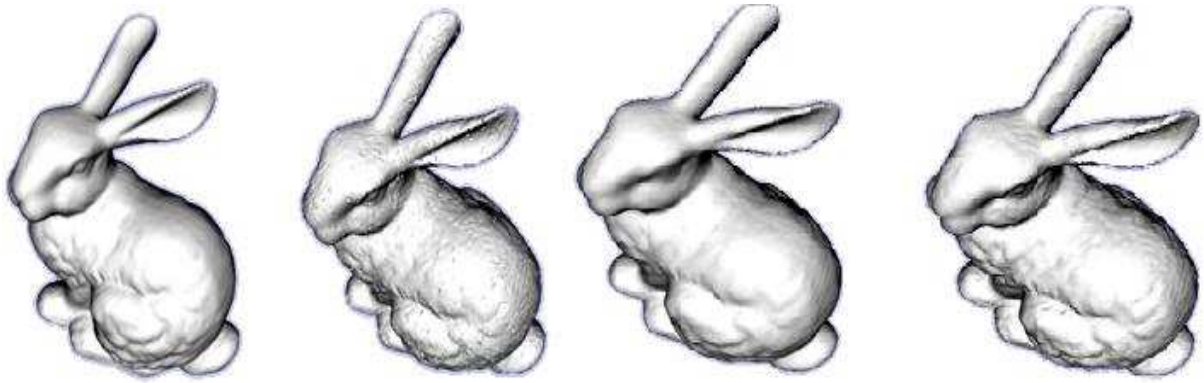


Figure 3-14 De gauche à droite : Modèle Original – Ajout de bruit Artificiel - Modèle lissé par un lissage Laplacien – Modèle dé-bruité par un filtrage bilatéral

Après cette comparaison, nous proposons d'appliquer, en prétraitement, un filtrage bilatéral sur les données des bases **bruitées**. Dans la Figure 3-15, nous présentons le pseudo-code d'une itération de l'algorithme de filtrage bilatéral gaussien de (Fleishman, et al., 2003) que nous a allons utiliser dans notre travail. L'idée de ce filtre sélectif consiste à faire une moyenne gaussienne sur les points les proches dans le voisinage. Le choix des paramètres comporte essentiellement le nombre d'itérations et la taille du voisinage.

```
DenoisePoint(Vertex v, Normal n)
  {q_i} = neighborhood(v)
  K = |{q_i}|
  sum = 0
  normalizer = 0
  for i := 1 to K
    t = ||v − q_i||
    h = ⟨n, v − q_i⟩
    w_c = exp(−t^2/(2σ_c^2))
    w_s = exp(−h^2/(2σ_s^2))
    sum += (w_c · w_s) · h
    normalizer += w_c · w_s
  end
  return Vertex v̂ = v + n · (sum/normalizer)
```

Figure 3-15 Pseudo-code de l'algorithme de filtrage bilatéral (Fleishman, et al., 2003)

3.2.2 Invariance aux transformations, échelle et résolution

Les valeurs des courbures principales d'une surface sont indépendantes des rotations 3D. C'est une propriété très importante pour la reconnaissance d'objets quand la capture des objets se fait selon des orientations arbitraires. Les primitives obtenues à partir des courbures principales sont alors invariantes à l'orientation. De ce fait, les valeurs H, K, S et C sont indépendantes de la translation et de la rotation dans l'espace 3D. Cependant, la variation d'échelle modifie la différence finie des valeurs échantillonnées des modèles et impacte le calcul des dérivées partielles discrètes d'une même surface. De ce fait, les courbures principales κ_1 et κ_2, calculées à partir d'une Hessienne ou des dérivées partielles de deuxième ordre d'une surface, sont dépendantes de l'échelle et varient proportionnellement à un changement d'échelle de la surface dans l'espace 3D. De plus, avec le caractère discret des surfaces obtenues avec les scans 3D, le ratio de l'échantillonnage est un deuxième facteur impactant sur la mesure des courbures. Ainsi les valeurs de H, K et C ne sont pas invariantes par changement d'échelle ou de résolution. Par contre, S est indépendant de l'échelle et de la résolution d'après sa formulation. L'indépendance de l'indice de forme à la résolution est illustrée dans la Figure 3-16, dans laquelle, différents niveaux d'échantillonnage de la même forme sont considérés. La remarque est que la résolution de l'image SI diminue certes, mais les valeurs de SI (intensité en niveau de gris) ne sont pas affectées.

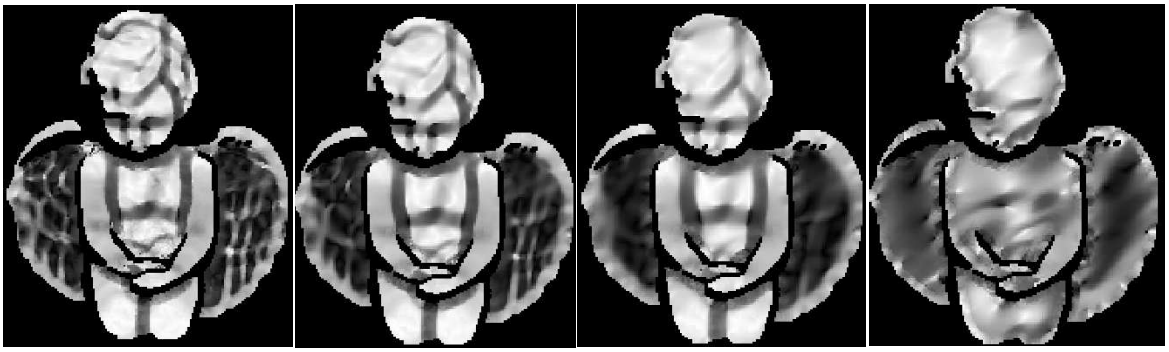


Figure 3-16 Images de l'indice de forme pour différentes résolutions de la forme initiale

3.2.3 Maillage et voisinage

Rappelons que l'objectif de cette thèse est de reconnaitre un objet à partir de quelques vues 2.5D et les mettre en correspondance après extraction des primitives.

- *Maillage :* La subdivision du nuage de points en des sous-régions locales (patchs) permet d'avoir des mesures d'une façon indépendante dans chaque portion du modèle. Nous exploitons la structure régulière en (i, j) de l'image 2D pour construire un maillage du nuage de points. Nous générons ce maillage de la façon suivante: nous cherchons les coordonnées des points minimums et maximums sur l'axe des x et des y pour former la boite englobante de l'objet. Ces coordonnées servent à calculer la taille approximative de la diagonale de l'objet. Ensuite, nous calculons la distance moyenne des points contenus dans cette boite. Puis, nous parcourons la grille des points dans l'ordre donnée par l'image de profondeur pour former, finalement, des facettes triangulaires à partir de points dont la distance entre eux est inférieure à la distance moyenne du nuage.
- Voisinage : Pour la détection des points d'intérêt, nous mesurons la saillance de chaque point. L'unité de mesure est le couple des courbures principales qui se calcule sur un voisinage d'un point. Le choix de la taille de ce voisinage est critique. Dans notre cas cette

taille est déterminée soit par le nombre d'anneaux autour du point soit par la valeur du rayon r d'une sphère centré au point. Dans le cas où les triangles du maillage ont la même taille (tessellation uniforme), nous pouvons considérer un nombre constant d'anneaux ou un rayon r fixe. Par contre, pour un maillage irrégulier et complexe, ces méthodes n'approximent pas le voisinage adéquatement. Une approche adaptative permet de sélectionner la taille du voisinage selon la tessellation autour du point.

- Considérons l'objet comme un graphe G (V, E), où V est l'ensemble des sommets du maillage et E est l'ensemble des arêtes d'adjacence dans l'objet. Pour un point v ∈V, un k-ring autour de v correspond à l'ensemble des points dont la longueur du chemin le plus court à partir de v est k. Une façon adaptative de choisir ce nombre est de prendre une proportion du nombre des facettes dans l'objet.
- Dans le cas où le voisinage des points est donné par une région sphérique autour du point, le rayon de cette sphère est pris proportionnellement à la diagonale de la boite englobante pour assurer une invariance aux changements d'échelles et à l'échantillonnage.
- Nous évitons également dans le calcul des courbures les points situés sur les bords et dans les zones de transition de visibilité. Ces points de discontinuité sont détectés par la recherche des lignes perpendiculaires à l'angle de vue et par leur faible densité (voisinage).

3.3. Résultats expérimentaux

3.3.1 Outils de développements et bases de données

Estimer de manière précise et fiable les courbures principales en tout point est une étape essentielle dans notre détection. Par souci de simplicité, l'approche adoptée est fondée sur l'ajustement local d'une surface polynomiale de second degré 2-jet. L'application d'une Analyse en Composantes Principales (ACP) sur le voisinage des points, suivi d'un ajustement polynomial permet de déduire la base et les coefficients de Monge (Figure 3-17). La bibliothèque CGAL (CGA10) est utilisée pour l'estimation des quantités différentielles d'une surface lissée d'un maillage triangulaire.

Dans le but d'accélérer le temps de calcul, nous avons pu exploiter également, l'implémentation donnée par (Tombari, et al., 2010) pour définir un système de référence local (RF) utilisé pour le calcul du descripteur SHOT. La normale est donnée par le vecteur propre d'une matrice de covariance formée par une combinaison linéaire des distances des points du voisinage. Nous notons ce deuxième algorithme par **RF.**

La bibliothèque VTK (VTK10) est utilisée pour stocker la structure du maillage, et l'application des transformations d'ajout de bruit et de décimation des données.

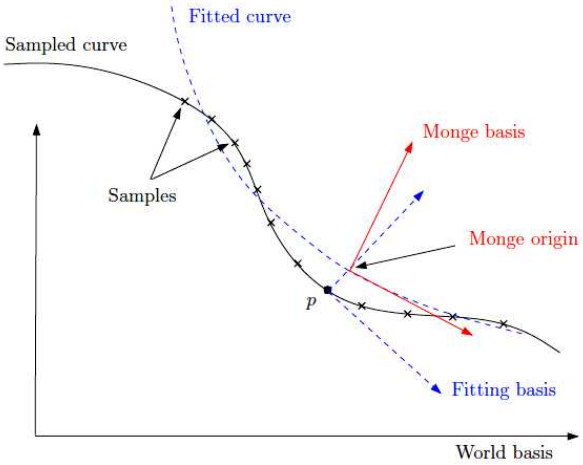


Figure 3-17 Les trois bases orthonormées utilisées pour l'estimation des courbures

Dans nos expérimentations, nous avons utilisé des images de profondeurs de modèles, présentant une variation en angle de prise de vue et d'échelle. Le degré de similarité entre les formes d'objets varie selon le type de la base de données. Des bases différentes en termes de taille, qualité, densité et type d'objets, générées par différents capteurs, sont exploitées dans nos tests. L'évaluation quantitative des détecteurs sera donnée uniquement sur les données de la base Minolta.

L'ensemble de ces bases sera exploité essentiellement dans le chapitre reconnaissance où chaque jeu de données est séparé en deux groupes : un ensemble pour la phase d'apprentissage et un autre pour la requête ou le test.

La base Stuttgart et la base Minolta

La base de Stuttgart University Range Image (Stu10) est composée de 42 objets. Ces images de profondeurs sont générées synthétiquement pour avoir un grand jeu de données offrant une variation de l'angle de vue. A chaque objet correspond 66 poses d'apprentissage et 258 poses pour le test. Ce qui donne un total de 66x42=2772 poses en apprentissage et 258x42=10836 poses en test. Les angles de vues sont pris sur les différents angles de la sphère avec un écart de 23-26° dans la base d'apprentissage et un décalage de 11.5-13° pour les requêtes de test. L'exemple de la Figure 3-18 représente l'ensemble d'apprentissage pour l'objet "machine".

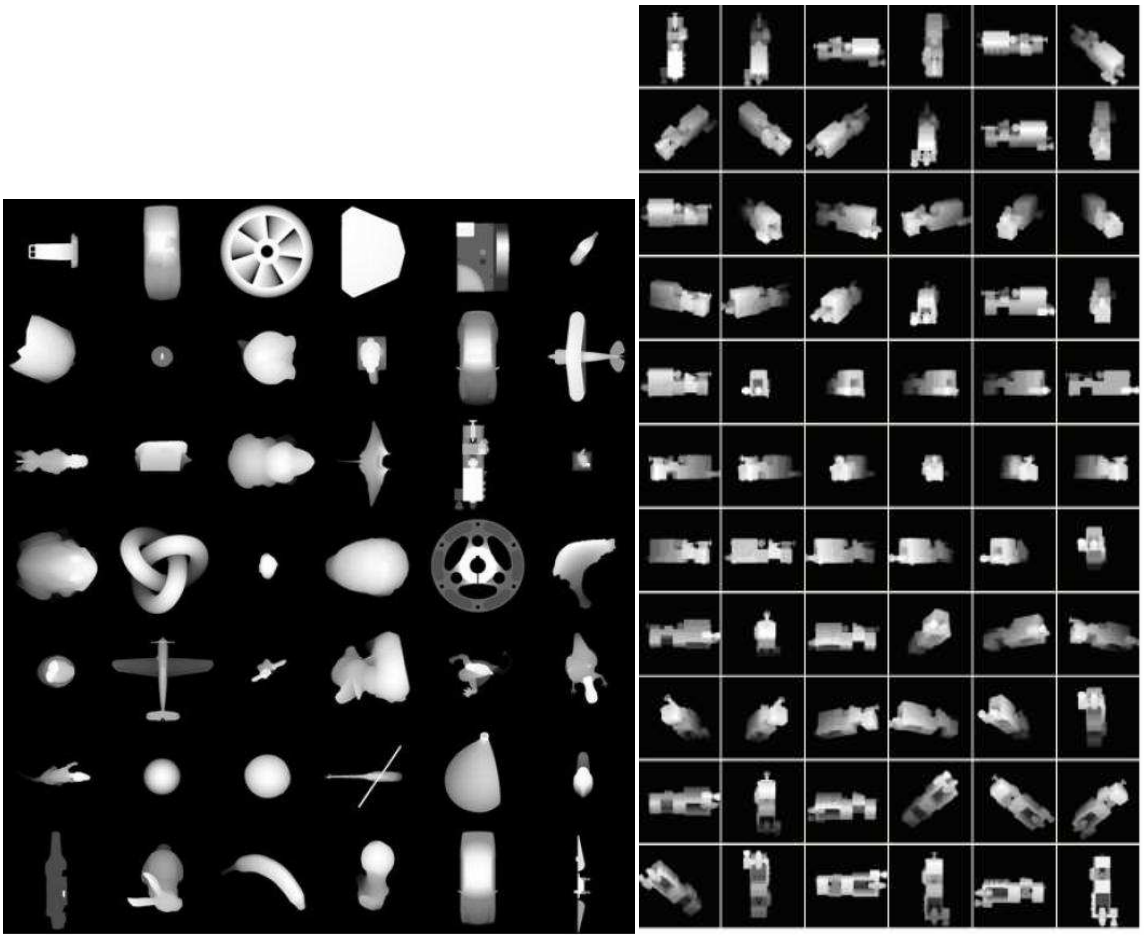


Figure 3-18 A gauche- Les 42 objets de la base de Stuttgart –A droite- L'ensemble des images de la base d'apprentissage pour l'objet « machine »

Une deuxième base d'images de profondeur est la base publique Minolta acquise par un scanner laser (Minolta Vivid 900 laser range scanner) (Min11) par l'université Ohio State University. Un grand jeu de données est fourni avec plus de 30 objets et présente des variations d'angles de 1°, 20° et 36°. Pour les chapitres détection et description, nous avons choisi de travailler avec 9 objets (Figure 3-19) de cette base avec un total de 162 vues d'écart de 20°. Cette sélection d'objet a été proposée par (Chen, et al., 2004). Nous avons remplacé l'objet 9 orangedino de cette base par l'objet bunny.

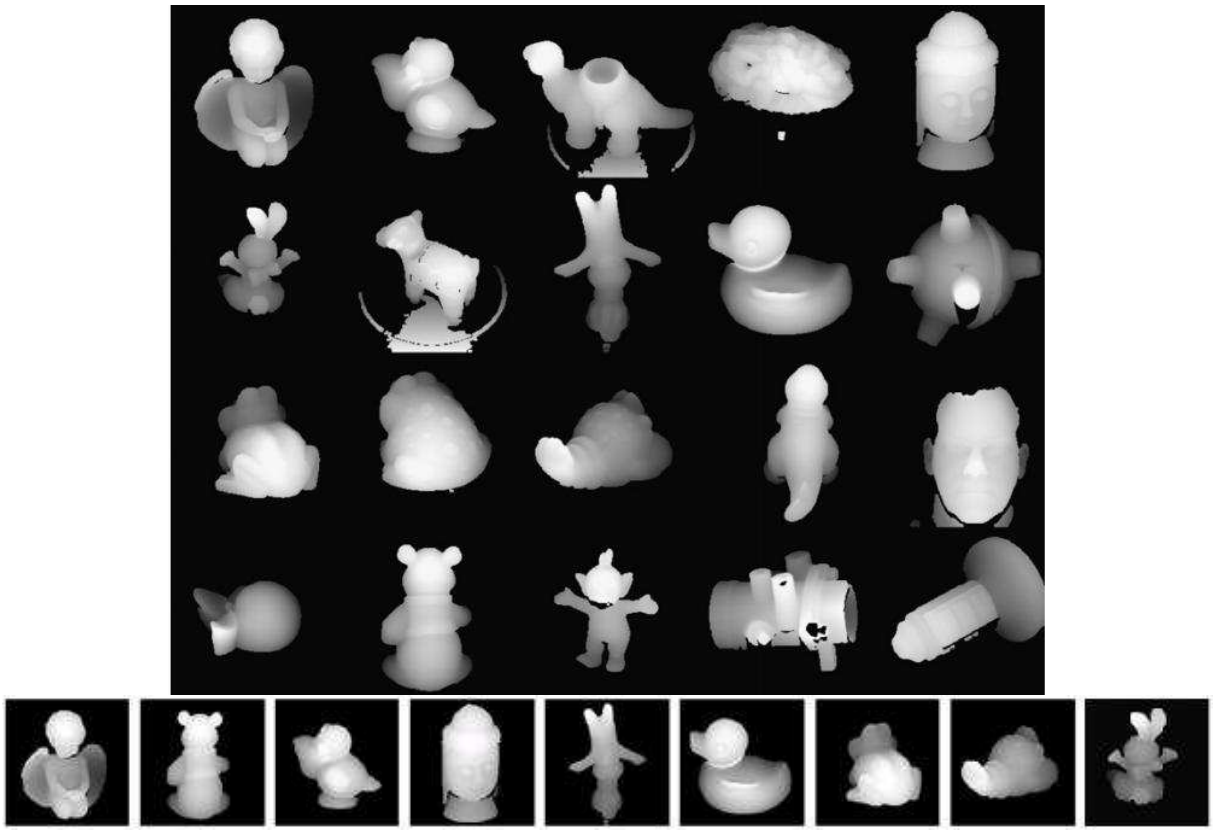


Figure 3-19 En haut- 20 objets de la base Minolta, en bas-Images de profondeurs des 9 objets de la base Minolta utilisés pour notre évaluation (Chen, et al., 2007)

Dans la base Minolta, se présentent, également, des vues prises sous différentes échelles de 2 à 4 échelles pour quelques objets. Dans ce jeu de données issues d'acquisitions réelles de scènes, les objets sont extraits par un processus de segmentation qui laisse des bouts du fond sur quelques vues. Nous avons pu récupérer 7 objets dont les vues ne présentent quasiment pas de problèmes de segmentation sauf pour quelques vues (exemple sur la ligne 1 et 2 de la Figure 3-20). Ces objets sont comme suit avec le nombre d'échelles par objet: bluedino (3 échelles), brain (3 échelles), facesimages (3 échelles), gc_bottle (4 échelles), reddino (3 échelles), valve (4 échelles) et yellowhorn (3 échelles). Pour chaque objet, nous avons sélectionné 11 vues. Le nombre de points dans des vues prises sous le même angle augmente en passant d'une échelle à une échelle plus proche du capteur. Par exemple, pour la vue 0 du modèle Rick des facesimages, nous avons 15523 points et 29920 facettes à l'échelle 0, contre 17883 points et 34813 facettes à l'échelle 1, et 18483 points et 35896 facettes à l'échelle 2.

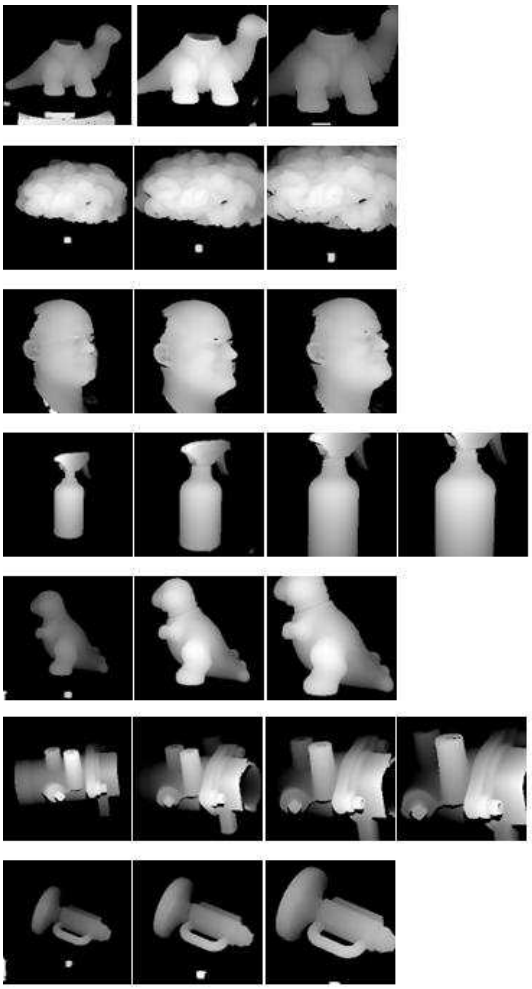


Figure 3-20 Les 7 objets présentant une variation d'échelle. Sur chaque ligne, nous illustrons les vues 0° de chaque échelle

La base RGB-D Object

La base d'objet "RGB-D Object Dataset" (RGB12) est, à notre connaissance, la plus grande base publique de données RGB-D et est composée de 300 objets communs « domestiques » (household). Ces objets sont organisés en 51 catégories rangées selon le système de relations WordNet hypernym-hyponym (similaire à ImageNet). Par exemple la catégorie soda est divisée en des instances comme Pepsi Can et Mountain Dew Can. Les captures sont faites avec la caméra 3D Kinect qui synchronise et aligne des images RGB 640x480 avec des images de profondeurs à une fréquence de 30 HZ. Chaque objet est placé sur un plateau tournant et une capture est prise pour toutes les rotations. La caméra est placée à plusieurs hauteurs pour avoir différents angles de prise de vues.

La Figure 3-21 représente 46 exemples d'objets segmentés de la base RGB-D, qui sont respectivement : apple_1, ball_1, banana_1, bell_peper_1, binder_1, calculator_1, camera_1, cap_1, cell_phone_1, cereal_box_3, coffee_mug_1, comb_1, flashlight_1, food_bag_1, food_box_1, food_can_1, food_cup_1, garlic_1, greens_1, hand_towel_1, instant_noodles_1, keyboard_1, Kleenex_1, lemon_1, lightbulb_1, lime_1, marker_1, mushroom_1, notebook_1, onion_1, orange_1,

peach_1, pear_1, pitcher_1, plate_1, potato_1, rubber_eraser_1, scissors_1, shampoo_1, soda_can_1, sponge_1, stapler_1, tomato_1, toothbrush_1 and watter_bottle_1.

Figure 3-21 Les 46 exemples d'objets segmentés de la base RGB-D (RGB12)

Figure 3-22 Deux vues du même objet pour la base RGB-D Object

Nous énumérons à présent quelques caractéristiques de la Kinect :

- Portée: ~ 50 cm à 5 m.
- Résolution horizontale: 640 x 480. Pour des géométries simples, la résolution est de ~ 0.75 mm par pixel en x sur y à une distance 50 cm, et de ~ 3 mm par pixel en x sur y à une distance de 2 m.
- FOV: 45° en vertical et 58° en horizontal
- Résolution de la profondeur: ~ 1.5 mm pour 50 cm et environ 5 cm à 5 m.
- Bruit: Environ +-1 DN pour toutes les distances, mais ce bruit est non-linéaire. Ce qui veut dire qu'il est égale à +-1 mm pour des distances faibles, et +- 5 cm pour les plus éloignés.

Base « Carotte » du laboratoire

Cette base a été construite dans notre laboratoire dans le cadre du projet CAROTTE « CArtographie par RObo T d'un Territoire » dans lequel l'une des tâches d'un robot mobile était de reconnaitre les

objets de son environnement. La caméra 3D Kinect est utilisée pour capturer quelques prises de vues (3 à 10) de 20 objets différents : ballon rouge, botte fourrée, botte en plastique verte, bouteille mauve, casier de rangement, chaise bleue à roulette, chaise beige à quatre pieds, cylindre vert, disque rose, boite d'archive bleu, gros bidon, livre, marteau, parallélépipède mauve, boite d'archive blanc, petit bidon noir, poubelle blanche, prisme jaune, robot Pioneer quatre roues, ventilateur (Figure 3-23). Ces objets sont de tailles différentes allant de 10cm à 1m. La distance moyenne de capture est de 1m50. L'avantage de ce type de caméra est que l'information RGB est stockée en plus de l'information de profondeur.

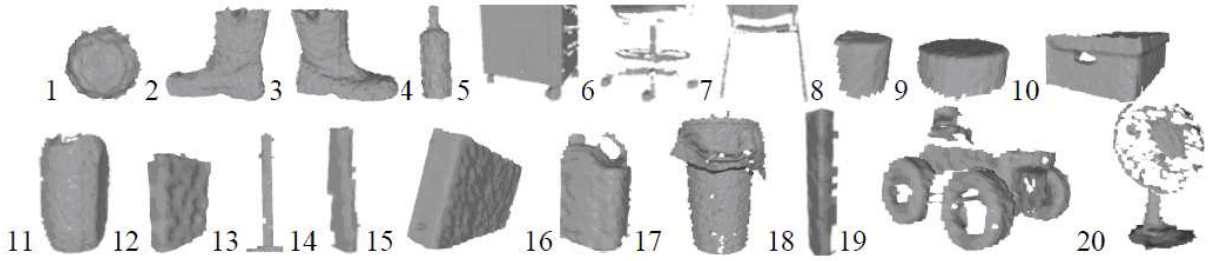


Figure 3-23 Les 20 objets de la base carotte du laboratoire

3.3.2 Evaluation des détecteurs

3.3.2.1 Evaluation visuelle

Nous proposons de commencer cette partie par une brève évaluation visuelle de la performance des 10 détecteurs SC_HK_FQ, SC_HK_C, SC_HK_C_Conf, SC_HK_C_Con, Harris, SI, HK, SC et SURF. Le détecteur FQ, ne fera pas parti des détecteurs testés. Par raison de son temps de calcul et sa faible performance, il a été éliminé.

Nous présentons quelques résultats de ces détecteurs pour les bases: Minolta (Figure 3-24 et Figure 3-25), Carotte (Figure 3-26), Stuttgart (Figure 3-27), et RGB-D (Figure 3-28). Nous allons mettre l'accent sur des aspects particuliers pour classer ces résultats.

❖ **Avantage du processus de combinaison:**

Par le croisement des deux espaces de classification SC et HK, nous profitons des avantages de l'un et de l'autre. En effet, la combinaison des deux estimateurs de courbure SC et HK est robuste aux variations d'angles (Figure 3-25) et élimine les points qui ne sont pas classés pareillement par les deux classificateurs de surface (Figure 3-24). Les points éliminés se caractérisent par une faible variation locale de la forme de l'objet ou peuvent correspondre à du bruit ou à un artéfact. En outre, le filtrage basé sur un tri des mesures de courbure (C ou FQ ou Conf) des PIs renforce cette saillance en prenant les points avec la plus forte variation de courbure par rapport à leur voisinage. La dispersion des PIs des détecteurs SC, HK, SC_HK_C, SC_HK_Conf et SC_HK_FQ, sur la surface de l'objet, est due à l'opération de regroupement basée sur l'écart spatial entre les PIs triés.

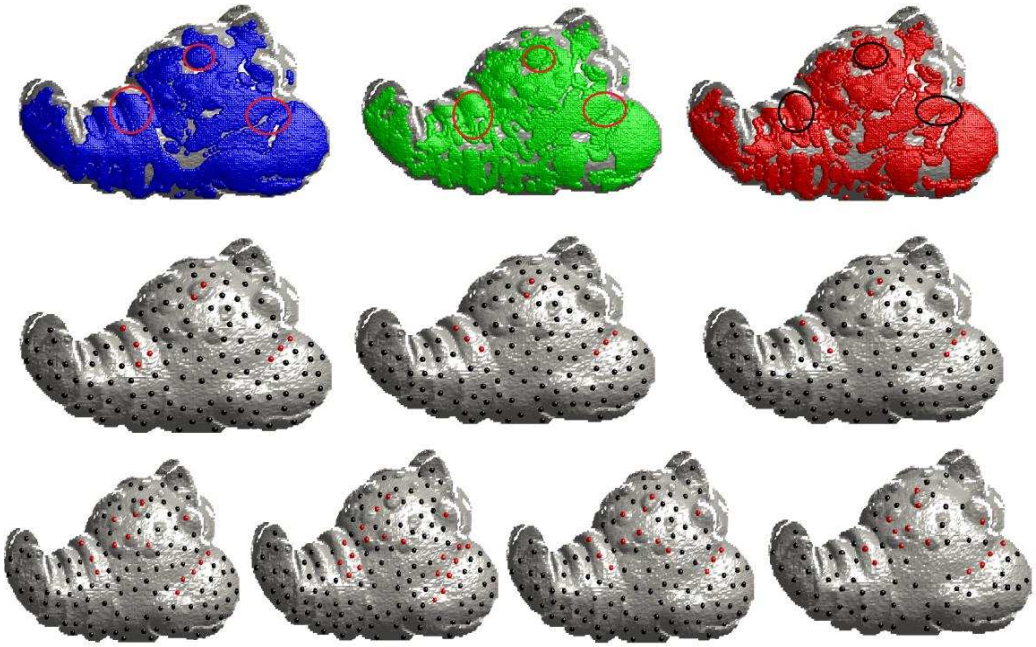


Figure 3-24 Résultat d'extraction des PIs pour le processus de combinaison.
Les détecteurs sont les suivants : -de gauche à droite- Ligne 1: HK sans filtrage (6695), SC sans filtrage (5462), SC_HK sans filtrage (4950). Ligne 2: HK avec filtrage (121), SC avec filtrage (121) et SC_HK_Conf (115). Ligne 3: SC_HK_C (121), SC_HK_FQ (131), SC_HK_Conf (115) et SC_HK_Con (81). Le nombre de PIs est donné entre parenthèses. Les points colorés en rouge dans la ligne 2 correspondent aux points retenus parmi les points de la zone entourée dans la ligne 1. Dans la ligne 3, les positions en rouge montrent des points extraits sur les mêmes zones de l'objet sur les 4 détecteurs.

Entre les trois méthodes de sélection combinée SC_HK_C, SC_HK_Conf et SC_HK_FQ basées sur un tri final sur les valeurs de C ou FQ ou Conf, la différence n'est pas très notable visuellement (quasiment les même positions de PIs sur la Figure 3-24). L'explication est que ces trois estimateurs mesurent la même information de degré de déviation de la courbure par rapport à un voisinage. Ce voisinage intervient, plus fortement, dans les mesures FQ et Conf puisque des notions d'extremum ou de moyenne, sur les valeurs de courbure des voisins, sont utilisées.
D'autre part, la version combinée SC_HK_C_Con, utilisant le processus de composantes connexes, sélectionne moins de points (81 PIs) que les autres détecteurs combinés (~120 PIs). En effet, la recherche de la connexité, entre les points en se basant sur le couple (SC-HK) de classification de la surface, réduit dans un premier temps l'espace de sélection des PIs. Ensuite dans un deuxième temps, l'étendu de chaque composante connexe sur l'objet est représenté par un seul PI avec la valeur de C maximale. Pour cette raison, les positions des PIs de ce dernier sont plus proches des positions du détecteur SC_HK_C que celles des autres versions combinées.

❖ **Comparaison avec des détecteurs de l'état de l'art:**

Nous présentons le résultat des détecteurs Harris_fract, Harris_clust, SURF, SI et SC_HK_FQ pour 4 vues de l'objet Ange dans la Figure 3-25. La particularité de cet objet est le fort degré de détails

présents sur sa forme. Pour les deux versions de Harris : avec la sélection d'une fraction de PIs parmi les plus grandes réponses de Harris et avec le regroupement par clustering, le résultat est très différent. Pour le premier, le Harris_fract, les positions sont condensées et situées sur des zones de transitions locales (presque au niveau des contours de la forme). Quant au Harris_clust, les PIs sont répartis sur toute la surface. Néanmoins, la répétabilité de ces deux détecteurs entre les différentes vues est notable et les positions des PIs sont saillantes avec une variation locale distincte de la forme.

Dans notre cas, nous avons privilégié avoir des PIs dispersés sur la surface, que des PIs condensés et collés les uns aux autres. Cette dispersion garantie la prise en compte d'un maximum de détails sur la forme et réduit la redondance dans la représentation à sauvegarder. Il en résulte la nette stabilité des PIs du détecteur combiné SC_HK_FQ sur les différentes vues. En outre, la nature du type de l'objet testé est adéquate pour le critère de notre détection (caractère courbée). En effet, la surface de l'Ange présente des courbures prononcées, ce qui augmente le nombre de PIs détectés.

Par ailleurs, en dépit du faible nombre de PIs détectés par SI, ces positions restent relativement répétables. Pareillement que notre détecteur, ce dernier extrait des extremums en mesure de courbure (indice de forme).

La remarque à donner concernant le détecteur SURF est que quelques positions de PIs se situent à l'extérieur de l'objet. La raison est que ce dernier utilise, contrairement aux autres détecteurs, une information volumique (à base de voxels obtenus par la répartition de l'intersection d'un cube avec la surface de l'objet). Cependant, l'avantage de ce détecteur est l'attribution d'une échelle (représentée par le rayon de la sphère sur la figure) aux PIs.

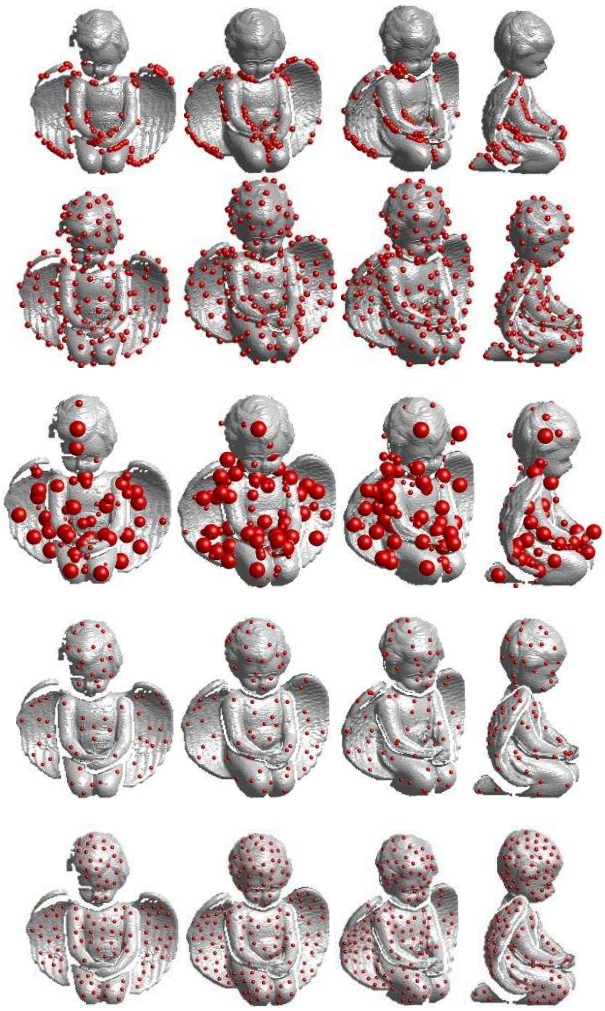


Figure 3-25 Positions des PIs détectés par respectivement de haut vers les bas : Harris_fract, Harris_clust, SURF, SI et SC_HK_FQ pour les angles (0°, 20°, 40° et 80°)

❖ **Présence de bruit**:

Les résultats sur la base de Carotte présente un autre type d'objets avec des conditions plus réelles. Malgré le bruit existant dans les données de la Kinect, nos algorithmes combinés réussissent à extraire des PIs situés sur des les zones caractéristiques de la forme de l'objet. Par exemple sur la Figure 3-26 des objets de la base Carotte, les PIs sont concentrés sur les plis au niveau du sac de la poubelle, sur les coins, les bords et le bouchon pour l'objet bidon, et sur le support et les pales du ventilateur. Ces positions saillantes sont relativement stables et se répètent sur les différentes vues du ventilateur.

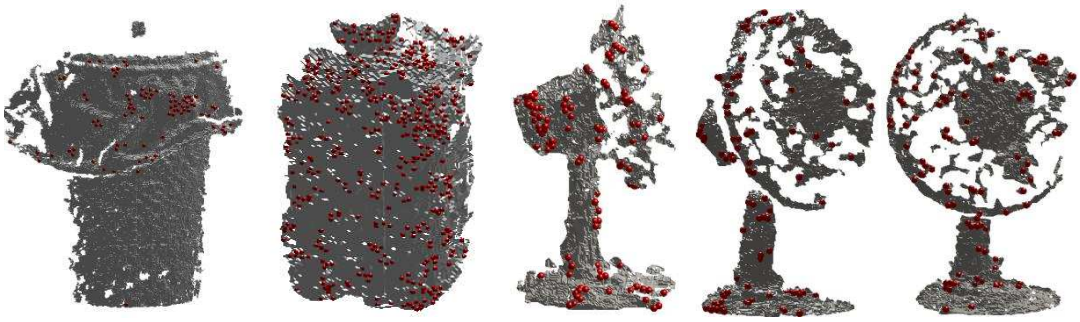


Figure 3-26 Résultat de détection des PIs sur les objets de la base Carotte avec respectivement de gauche à droite: les détecteurs SC_HK_FQ (poubelle), SC_HK_C (bidon) et SC_HK_Conf (ventilateur)

❖ **Objets mécaniques:**

L'illustration de la Figure 3-27 montre le résultat de détection de l'algorithme combiné SC_HK_C sur des objets de types mécaniques pour les modèles agfa, machine et auto de la base Stuttgart. Ces surfaces sont à dominance de surfaces planes. Comme la sélection basée sur les critères SC et HK ne prend pas en compte ces zones, nous avons rajouté le type plan dans les types de surfaces sélectionnées. N'ayant pas un critère de tri sur ce type de surface, néanmoins, le filtrage utilisant une condition sur le seuil de la distance spatiale entre les PIs permet d'avoir des PIs répartis sur la surface de l'objet. L'inconvénient de cet approche est que le nombre de PIs peut être grand (~400 PIs) pour des modèles à grande densité de points comme le cas des objets de la base Stuttgart. L'information prélevée sur ces PIs est redondante et le risque de confondre ces PIs entre eux augmente, au moment mise en correspondance.

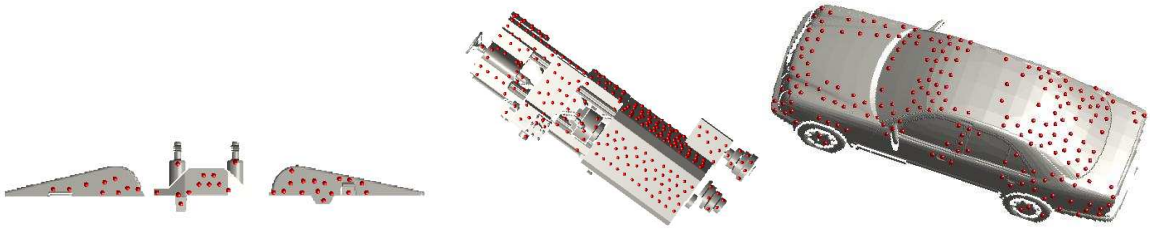


Figure 3-27 Positions des PIs détectés par le détecteur SC_HK_C sur des objets de type mécanique pour les modèles agfa, machine et auto de la base Stuttgart

❖ **Similarité entre les objets:**

Parmi les contraintes d'une méthode locale, est que l'extraction de PIs sur des objets de forme très similaire donne les mêmes positions physiques. L'illustration de la Figure 3-28 appuie cette remarque. En effet, les vues partielles des objets oignon, tomate et orange se ressemblent fortement. Le bruit des données est réparti de la même façon sur la surface des objets et renforce cette similitude. Cette limitation du détecteur peut être surpassée par une définition d'un descripteur distinguant ces points. Par exemple en utilisant la texture de la surface, nous parvenons à différencier ces objets comme le cas de la couleur des modèles oignon, tomate et orange.

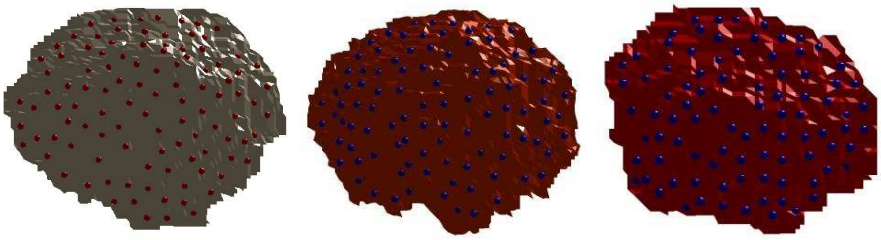


Figure 3-28 Les PIs détectés sur les objets: oignon, orange et tomate de la base RGB-D montrant la forte similarité de la forme. Une couleur de la surface est rajoutée pour différencier ces objets.

❖ **Version multi-échelles :**

Nous avons pu entamer l'implémentation de l'approche multi-échelle décrite précédemment. Nous présentons quelques figures (Figure 3-29, Figure 3-30 et Figure 3-31) illustrant des résultats de ce détecteur. D'après la Figure 3-29, l'etendu des composantes connexes dans la version multi - échelles est réduit par rapport à la version mono-échelle et le nombre des PIs finaux est largement plus petit.

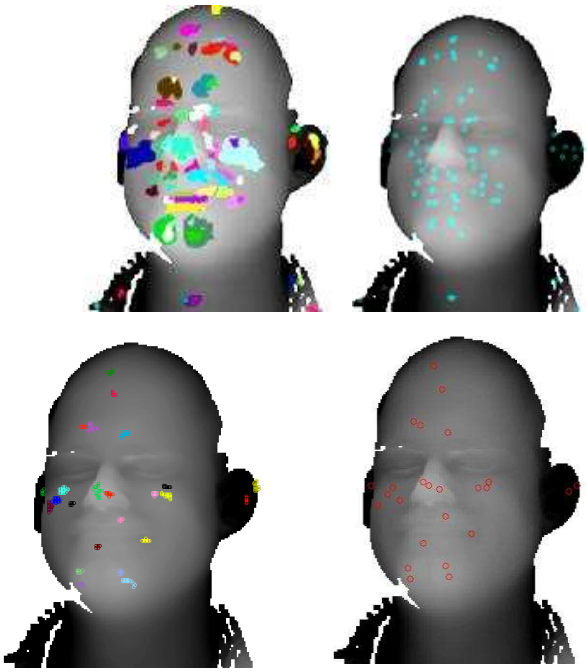


Figure 3-29 Résultat de la détection en échelle fixe (première ligne) et en multi échelles (deuxième ligne) pour le détecteur SC_HK_Con. La première colonne représente l'étendu des composantes connexes et la deuxième les positions des PIs finaux.

D'après la Figure 3-32 illustrant, sur des images de l'indice de forme, le résultat du processus de détection par l'algorithme SC_HK_Con sur trois tailles croissantes de voisinage. Les mesures de l'indice de forme représentées sont calculées sur ces trois nivaux de voisinages. Nous remarquons que les PIs détectés sont concentrés sur les mêmes régions de l'objet de visage (la bouche, le nez, les yeux et les oreilles). Outre cette stabilité à l'échelle, nous avons constaté que les figures (Figure 3-30, Figure 3-31) mettent l'accent sur une forte répétabilté des PIs pour des variations d'angle de vues. Par

exemple, nous avons compté 18 PIs qui sont répétables pour l'approche SC_HK_Con muti-échelle par rapport à 15 PIs répétables pour l'approche de Sparse points introduite par (Castellani, et al., 2008).

Bien que prometteuse pour le peu de tests effectués, l'évaluation de cette méthode n'a pas pu être poussée plus loin, par contrainte de temps.

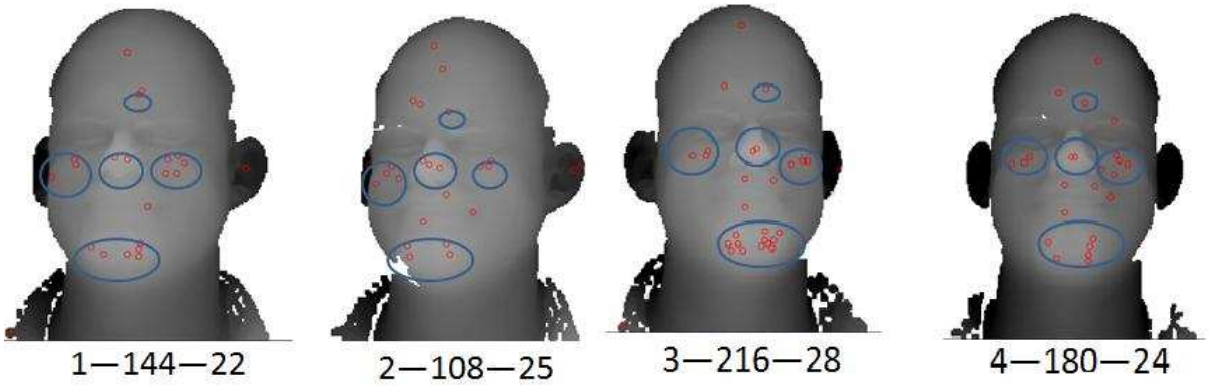


Figure 3-30 Invariance aux angles de vues pour l'approche SC_HK_Con multi-échelle. Les nombres représentent respectivement : l'ordre de la vue, le numéro de la vue dans la base Minolta et le nombre des PIs.

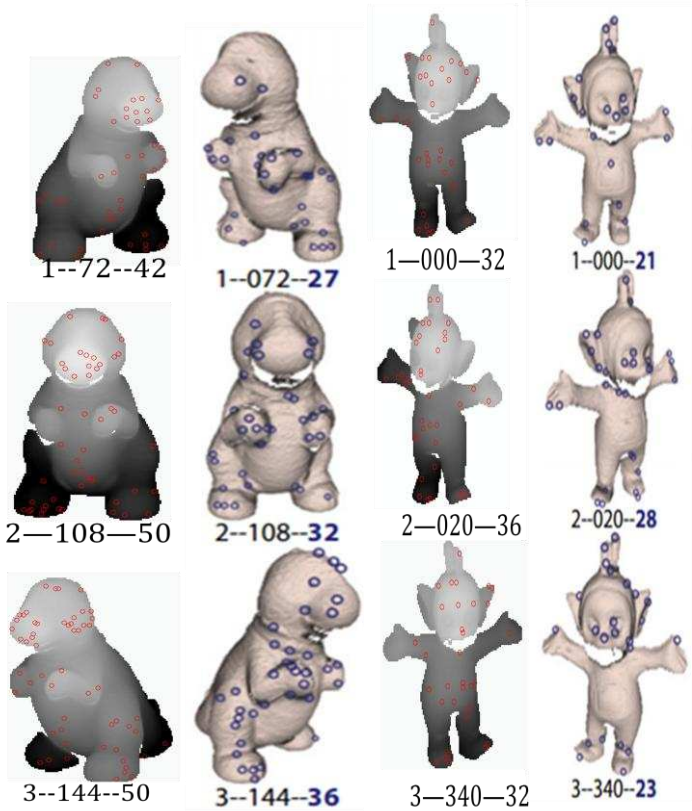


Figure 3-31 18 Résultats de comparaison de l'approche SC_HK_Con muti-échelle par rapport à l'approche de (Castellani, et al., 2008)

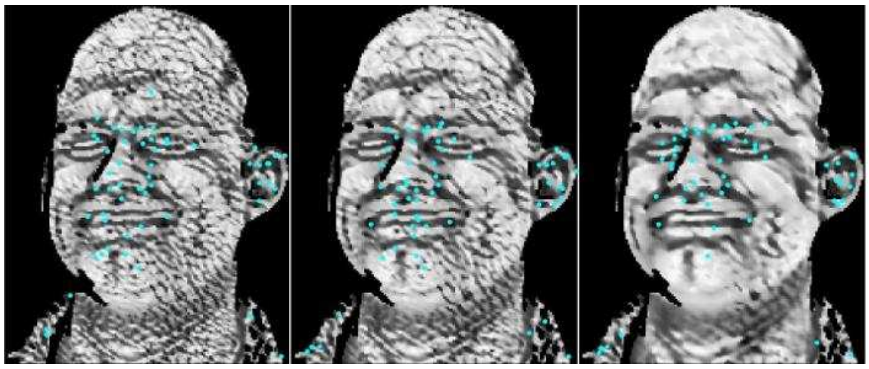


Figure 3-32 Représentation du résultat de détection en utilisant trois niveaux de voisinages sur les images d'indice de forme.

La comparaison visuelle de cette partie reste approximative et n'évalue pas la réelle performance des détecteurs. Pour le paragraphe suivant, nous introduisons une analyse quantitative de la répétabilité de ces détecteurs aux transformations géométriques de rotation et de changement d'échelle, et de leur robustesse au bruit et à l'échantillonnage. Cette étude comparative est faite uniquement sur les objets de la base Minolta.

3.3.2.2 Stabilité aux changements d'angles

❖ **Description du protocole de test** :

Le critère d'évaluation des expérimentations est la répétabilité des points d'intérêt. Etant donné un objet O et une fonction de transformation T, qui peut être une translation, une rotation, un changement d'échelle ou un ajout de bruit, T(O) est l'objet après l'application de la transformation. Soit aussi PO l'ensemble des points d'intérêt extraits sur O. Ainsi, la répétabilité est définie par:

$$R_{O,T(O)} = \frac{|P_O \cap P_{T(O)}|}{|P_O|}$$

Un seuil sur la distance entre les points détectés sur T(O) et les points détectés sur O après les avoir transformés avec T permet de compter le nombre de points répétables. La quantification de cette répétabilité se fait en comparant la distance physique des points appariés à un seuil de distance. La définition de ce seuil décide de la correspondance ou pas entre les positions des points: si la distance entre deux points est inférieure au seuil fixé, le nombre de points répétables est incrémenté.

Une première façon d'évaluer la répétabilité est de fixer le seuil de la distance et calculer la répétabilité entre le modèle original et l'objet transformé, et puis faire la moyenne pour chaque objet de la collection. Finalement, la répétabilité correspond à la valeur médiane de toutes les valeurs de répétabilité obtenues pour tous les objets de la base.

Une deuxième évaluation est faite en traçant la courbe médiane des différentes mesures de répétabilité en fonction d'une variation du seuil de distance.

Pour avoir une normalisation dans la comparaison des différents détecteurs, nous pouvons soit faire le rapport entre le premier et le deuxième plus proche voisin, soit, le rapport entre la distance du premier plus proche voisin et une distance maximale calculée sur toute la base et pour tous les détecteurs. Nous avons suggéré que la deuxième normalisation est plus adéquate pour notre comparaison, vu qu'elle permet d'avoir la même normalisation pour tous les détecteurs et surtout ne pénalise pas les détecteurs qui sortent des points dispersés sur la surface de l'objet.

Les paramètres de chaque méthode sont choisis empiriquement d'une façon à avoir des nombres similaires de PIs et pouvoir effectuer des comparaisons de répétabilité au même ordre. Par exemple pour le détecteur SI, nous jouons sur les paramètres alpha et beta pour varier le nombre de PIs. Pareillement, nous ajustons les entrées et les seuils des algorithmes des détecteurs Harris et SURF.
Les seuils H_{zero}, K_{zero} et C_{zero} sont choisis tel que $K_{zero} = H^2_{zero}$ et $C_{zero} = K_{zero}$ comme proposé par (Akagündüz, 2011). Nous avons sélectionné empiriquement pour ces seuils respectivement les valeurs 0.01, 0.0001 et 0.01.
Pour les expérimentations sur le détecteur de Harris, les valeurs du paramètre de Harris est K = 0.04 et la valeur du paramètre de sélection finale est égale à 0.01 pour la méthode de Harris_fract et égale à 0.006 pour la méthode de Harris_clust.
Sur les 9 objets de la base Minolta, nous appliquons les détecteurs sur les 18 vues de chaque objet. Nous avons au total 17 transformations. La courbe de répétabilité des PIs entre les différentes vues pour les 10 détecteurs : SC_HK_FQ, SC_HK_C, SC_HK_Conf, SC_HK_Con, Harris_fract, Harris_clust, SI, SC, HK et SURF, est présentée dans la Figure 3-33. Seulement la partie gauche des graphes obtenus est représentée (plus importante car ce qui nous intéresse est l'existence de points homologues très proches entre vue originale et vue transformée).

❖ **Résultats** :

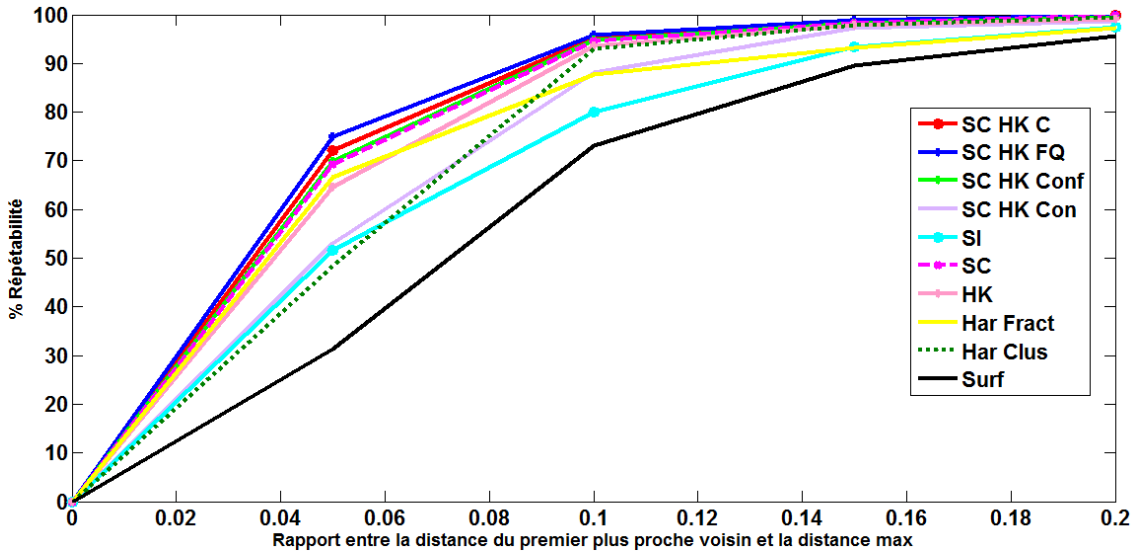


Figure 3-33 Répétabilté des PIs pour les 10 détecteurs : SC_HK_FQ, SC_HK_C, SC_HK_Conf, SC_HK_C_con, Harris_fract, Harris_clust, SI, SC, HK et SURF sur les 9 objets de la base de Minolta.

La performance de la répétabilité des trois détecteurs combinés SC_HK_C, SC_HK_Conf et SC_HK_FQ est comparable pour une variation d'angles de vue. La courbe de SC_HK_FQ est légèrement au dessus de celles des autres détecteurs. Les explications données, dans la partie évaluation visuelle, sont valables pour justifier ce comportement.
L'exception de la courbe du SURF qui est bien au dessous des autres est à noter. Nous pensons que peut être le fait d'ajuster les paramètres pour avoir un nombre de PIs comparables, a pu faire baisser la performance de ce détecteur.
Le Harris dans sa version de sélection par fraction est meilleur que la version Harris_clust pour les faibles valeurs du seuil. Ce comportement s'inverse à partir de la valeur 0.09 du seuil.

D'après le Tableau 3-1, pour une valeur du seuil fixée à 0.16, les cinq premières répétabilités se présentent comme suit : 98.8% pour SC_HK_FQ, 98.5% pour SC_HK_C, 98.2% pour SC, 98.2% pour SC_HK_Conf et 97.9% pour Harris_clust.
L'invariance des mesures S, C, H et K à la rotation explique la haute répétabilité des détecteurs basés sur ces estimateurs. Le processus de combinaison de critères a augmenté considérablement cette répétabilité en comparant aux résultats des détecteurs basés sur uniquement un de ces critères. Les PIs regroupés par composantes connexes sont moins stables que ceux ordonnés et regroupés par clustering. La répartition de ces points sur toute la forme par la méthode de clustering, par rapport à une concentration sur certaine zones obtenue par les composantes connexes, peut expliquer ce léger avantage. Un comportement inverse est observé avec le détecteur Harris. La sélection d'une fraction des points après les avoir ordonnés selon leur réponse d'Harris est plus stable qu'une sélection de PIs par clustering du moins pour les petites valeurs du seuil.
D'autre part, la version combinée SC_HK_C_Con a une répétabilité largement plus faible que les autres détecteurs combinés (pour des seuils de distances petits). Ce détecteur ne sera pas pris en compte dans le reste des évaluations.

Seuil	0.06	0.16
SI	51.7%	93.3%
SC_HK_FQ	**74.9%**	**98.8%**
SC_HK_C	**72.0%**	**98.5%**
SC_HK_Conf	**69.8%**	**98.2%**
SC_HK_Con	54.0%	97.2%
HK	64.6%	97.8%
SC	**69.3%**	**98.2%**
Harris_fract	66.6%	93.2%
Harris_clust	48.4%	97.9%
SURF	31.4%	89.6%

Tableau 3-1 Valeurs de répétabilité, à deux seuils de distance, entre les angles de vues pour les 10 détecteurs

3.3.2.3 Stabilité à l'échelle

❖ **Description du protocole de test** :

Base à variation *artificielle* d'échelles

Dans un premier temps de nos expérimentations, nous proposons de générer des vues avec une variation d'échelles faite manuellement. Cette transformation correspond à une variation de la densité des points sur le maillage. Pour ce faire, nous appliquons une transformation de décimation (qui réduit

le nombre de sommet sur le modèle) avec un facteur de réduction f égal à 0.3 sur les 18 vues de chacun des 9 objets Minolta. Nous disposons alors de 18 nouvelles vues pour chaque objet. Une troisième échelle est obtenue par une réduction de la densité des points dans les modèles d'origines avec un taux de 50% du nombre de points initial. Au final, nous constituons notre base avec trois variations d'échelle : les vues originales (échelle 0), les vues de densité réduite de 30% (échelle 1) et les vues de densité réduite de 50% (échelle 2). Un exemple de décimation est donné dans la Figure 3-34. Pour chaque objet, nous disposons de 54 vues au total.

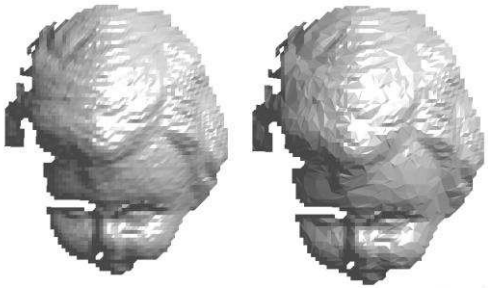


Figure 3-34 Exemple de décimation avec un facteur f = 50% -à gauche le modèle original et à droite le modèle après décimation

Sur les 9 objets de la base Minolta, nous appliquons les détecteurs sur les 18 vues de chaque objet, et à chaque échelle. Nous avons considéré les 2 transformations de changement d'échelles suivantes : la première, **Ech_1,** s'applique aux 18 vues de l'échelle 1 vers les 18 vues de l'échelle 0 et la deuxième, **Ech_2,** transforme les PIs des 18 vues de l'échelle 2 vers les vues de l'échelle 1. Les courbes de répétabilité des PIs entre ces deux transformations pour les 9 détecteurs : SC_HK_FQ, SC_HK_C, SC_HK_Conf, Harris_fract, Harris_clust, SI, SC, HK et SURF, est présentées dans les figures Figure 3-35 et Figure 3-36. Pour mieux visualiser l'écart entre ces courbes, nous avons normalisé la distance des PIs, cette fois ci, par la distance moyenne de toutes les distances de la base et des détecteurs (au lieu de prendre la distance maximale).

Base à variation réelle d'échelles

Dans un deuxième temps de notre évaluation, les données des 7 objets de la base Minolta, présentant une variation de vues et d'échelles sont utilisées. Pour chaque objet, nous considérons les 11 vues sur 2 échelles choisies parmi les trois ou quatre échelles de cet objet.

Nous appliquons les 9 détecteurs sur les 11 vues de chaque échelle pour tous les objets. La courbe de répétabilité des PIs entre ces échelles pour ces 9 détecteurs est illustrée dans la Figure 3-37.

La Figure 3-38 montre les positions des PIs extraits par les détecteurs Harris_fract, Harris_clust, SC_HK_C et Surf pour la même vue et sur deux échelles différentes de l'objet gc_bottle.

❖ **Résultats** :

Les remarques générales sur ces trois courbes sont comme suit :

- Entre les deux transformations artificielles de l'échelle, nous constatons la même allure des courbes des détecteurs. L'échelle 2, où les points sont réduits à 50%, simule une distance d'acquisition plus éloignée que celle de l'échelle 1, dont la décimation est de 30% du nombre initial des points. De ce fait, la répétabilité de tous les détecteurs baisse entre la transformation de l'échelle la plus proche (Ech_1) vers la transformation la plus éloignée (Ech_2).
- Entre les transformations artificielles et la transformation réelle de l'échelle, le comportement des détecteurs SURF et Harris_clust change. Au plus faible seuil de distance, le Surf et le

Harris_fract, suivis par SI ont la meilleure performance sur les transformations artificielles. Sur la variation réelle d'échelle, c'est le Harris_fract, Harris_clust et SI qui l'emportent.

- Etonnamment, sur le changement d'échelle des données réelles, les résultats montrent que le Surf, qui est une approche multi-échelle, a clairement une faible répétabilité. Nous pensons que le faible nombre de points détectés sur les vues peut justifier ce résultat. Un paramétrage adéquat pour augmenter le nombre de PI donnera probablement un meilleur résultat. Cependant, la réelle valeur de ce détecteur peut résider dans le caractère descriptif exprimé par le voisinage des PIs (grâce à l'attribution d'une échelle adéquate à chaque PI), que nous étudierons dans la phase de description et de reconnaissance qui suivent.
- En variation réelle d'échelles, les PIs des détecteurs SC_HK_C, SC_HK_FQ, SC et HK se répètent de la même façon. Nous remarquons une légère supériorité de SC_HK_C par rapport aux deux autres détecteurs combinés pour les plus faibles seuils de distance. Le processus de combinaison, n'améliore pas la performance des détecteurs simples, en changement d'échelles.

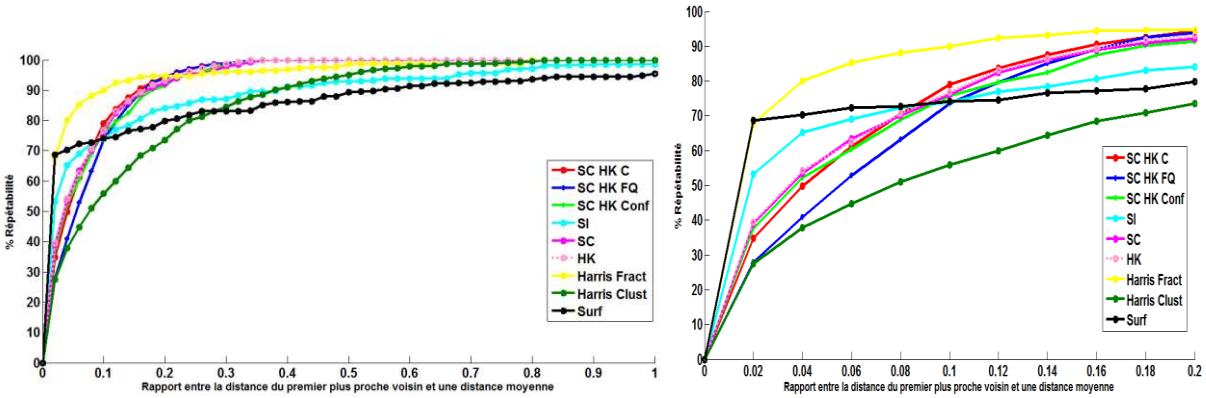


Figure 3-35 Répétabilité des 9 détecteurs pour la transformation d'échelle Ech_1- A droite-Zoom sur la partie gauche de la première courbe

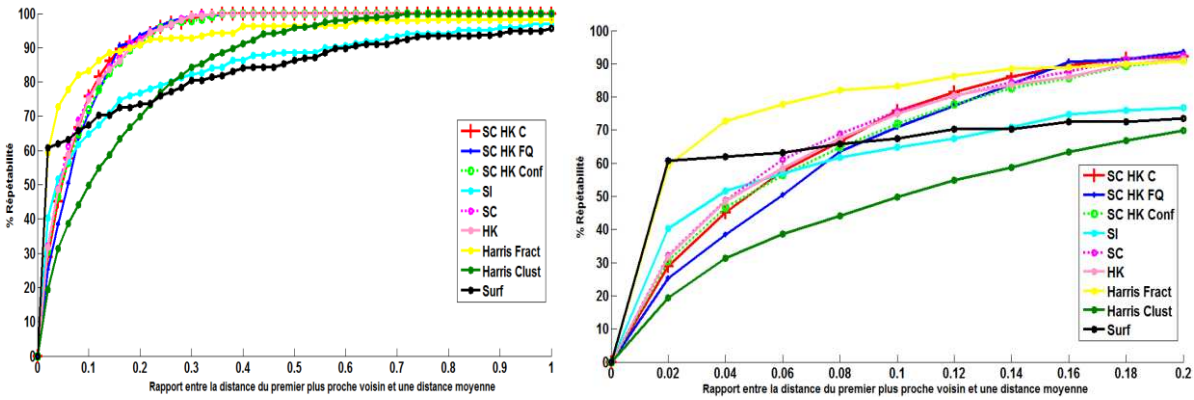


Figure 3-36 Répétabilité des 9 détecteurs pour la transformation d'échelle Ech_2 -A droite-Zoom sur la partie gauche de la première courbe

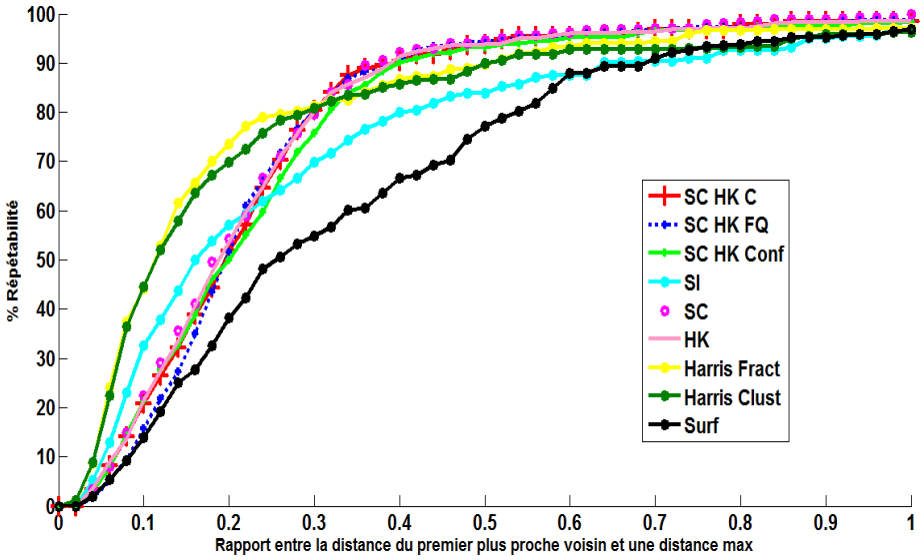


Figure 3-37 Répétabilité des 9 détecteurs SC_HK_FQ, SC_HK_C, SC_HK_Conf, SI, SC, HK, Harris_fract, Harris_clust et SURF sur 11 vues de deux échelles réelles

En variation d'échelles, à un seuil fixé à la valeur 0.4, où la majorité des détecteurs ont une performance supérieure à 80%, nous notons les performances de stabilité suivantes : 89.7% pour SC_HK_FQ, 89.7% pour SC_HK_C, 87.8% pour SC_HK_Conf, 83.9% pour Harris_fract ,81.9% pour Harris_clust, 78.1% pour SI et 62.0% pour SURF. Ces résultats nous emmènent à conclure, concernant l'invariance à l'échelle, sur la stabilité des détecteurs proposés et leur performance qui reste comparable à celle des détecteurs de l'état de l'art.

Le choix adaptatif du voisinage (avec une valeur proportionnelle à la taille de l'objet) nécessaire pour le calcul des courbures et des mesures de saillance est l'une des raisons de cette stabilité. Ce paramètre est le seul facteur, prenant en compte la robustesse aux changements d'échelles dans la version mono-échelle de notre algorithme. Nous pensons que la performance de nos détecteurs pourrait éventuellement s'améliorer avec la version multi-échelle.

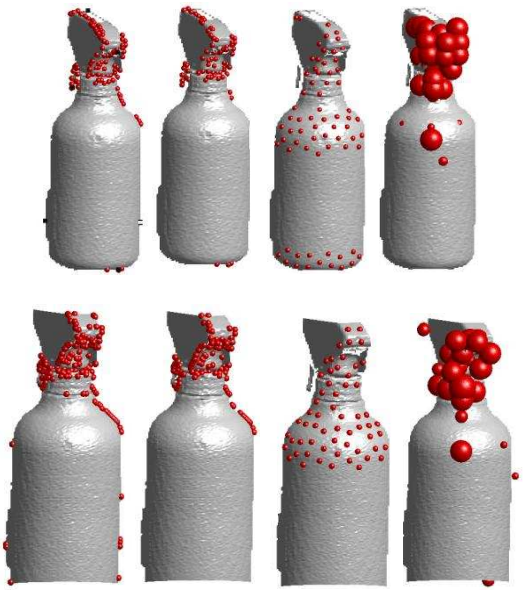


Figure 3-38 Répétabilité des PIs extraits par les détecteurs Harris_fract, Harris_clust, SC_HK_C et Surf dans l'ordre de gauche à droite, pour la même vue sur deux échelles : échelle 1 en première ligne et échelle 2 en deuxième ligne.

Dans notre évaluation de la répétabilité, nous cherchons des homologues des PIs des vues de l'échelle 1 se trouvant dans les vues de l'échelle 2 qui est plus proche du capteur. Or, comme le montre la Figure 3-38, la vue de l'échelle 2 ne présente qu'une portion de la surface de l'objet. Notre détecteur, SC_HK_C, détectent des points sur la vue de l'échelle 1 qui n'ont pas de correspondants réels sur la vue de l'échelle 2. La distance d'appariement de ces points est alors grande, c'est qui justifie l'infériorité de la performance de notre détecteur par rapport à celle du Harris_fract, par exemple. Nous pensons, que la transformation inverse, c'est-à-dire de l'échelle 2 l'échelle 1, fera augmenter la répétabilité pour nos détecteurs. Point de vue pratique, dans notre appariement des PIs pour la reconnaissance d'objet, nous suggérons de prendre en considération cette remarque. Nous proposons de fixer une distance seuil (sur la distance physique des PIs) à partir de laquelle nous décidons de valider ou pas la mise en correspondance d'une paire de PIs.

3.3.2.4 Stabilité au bruit

❖ **<u>Description du protocole de test</u>** :

L'évaluation de l'impact du bruit sur le résultat de détection se fait par l'ajout d'un bruit gaussien au maillage initial. Puis, nous mesurons la répétabilité des PIs détectés sur le nouveau maillage bruité par rapport à ceux détectés dans le maillage initial. Pareillement qu'en rotation, la comparaison des répétabilités des détecteurs se fait par la variation d'un seuil de la distance des PIs normalisée par une distance maximale calculée sur toute la base. Un bruit gaussien de variance 0.2 x MR (résolution du maillage) est appliqué à 5 vues de chacun des 9 objets Minolta. Nous présentons la Figure 3-39, les courbes de la médiane des valeurs de répétabilité de chacun des 9 détecteurs: SC_HK_FQ, SC_HK_C, SC_HK_Conf, Harris_fract, Harris_clust, SI, SC, HK et SURF. Dans le Tableau 3-2, nous montrons le pourcentage de répétabilité pour quelques détecteurs en fixant deux seuils du rapport des distances entre les PIs.

❖ **Résultats** :

Le bruit fausse les valeurs initiales des profondeurs des points et par conséquent la mesure de courbure est bruitée, surtout si la taille du voisinage est petite. Pour les deux détecteurs de la Figure 3-40, le nombre de PIs augmente sur les vues bruitées par rapport aux vues initiales.

Le Harris_fract donne nettement la meilleure répétabilité pour les petites valeurs du seuil des distances. Nous expliquons ce résultat par la forte concentration de ses PIs sur certaines zones de la surface. La distance entre ces PIs est, de ce fait, très faible que celle pour un détecteur qui extrait des points éparpillés sur la surface.

En général, pour l'ajout d'un bruit, les détecteurs, utilisant les espaces SC et/ou HK, ont une stabilité comparable. Nous notons, particulièrement, que la répétabilité de nos détecteurs combinés est rapidement au dessus de 90%, dès la valeur 0.06 du seuil correspondant au ratio des distances des PIs sur la distance maximale. A la valeur 0.12% du seuil, les trois détecteurs combinés ont 100% de répétabilité. La performance légèrement supérieure de SC_HK_FQ peut s'expliquer par l'implication des valeurs de courbures de son voisinage dans le calcul de son facteur de qualité dans la phase du tri. Il en résulte que ce facteur renseigne sur une plus large zone spatiale que celle par exemple des PIs SC_HK_C pour lesquels la saillance implique la mesure de courbure C du PI uniquement. La sélection finale par regroupement permet également d'obtenir des positions réparties sur toute la forme de l'objet et augmente la possibilité d'approcher plus de points sur la surface de l'objet.

Par ailleurs, la stabilité de ces détecteurs est fortement liée à la robustesse du calcul des quantités différentielles (normale et courbure) sur des données bruitée. En effet, dans notre cas, pour chaque point du nuage de point, nous calculons ces quantités en appliquant la méthode de Monge sur le voisinage de ce point. Ainsi, la mesure sur le point porte la moyenne des contributions des points qui l'entourent. De ce fait, la taille du voisinage intervient dans cette stabilité. Pour un choix d'une taille faible du voisinage, les courbures seront sensibles aux perturbations des données. Par ailleurs, l'inconvénient majeur de cette méthode est le temps de calcul qui reste linéaire avec le nombre de points dans le nuage.

Pour des bases de données qui sont naturellement bruitées, nous suggérons d'appliquer un prétraitement de dé-bruitage avant d'appliquer nos algorithmes de détection, tel le cas des bases de la Kinect.

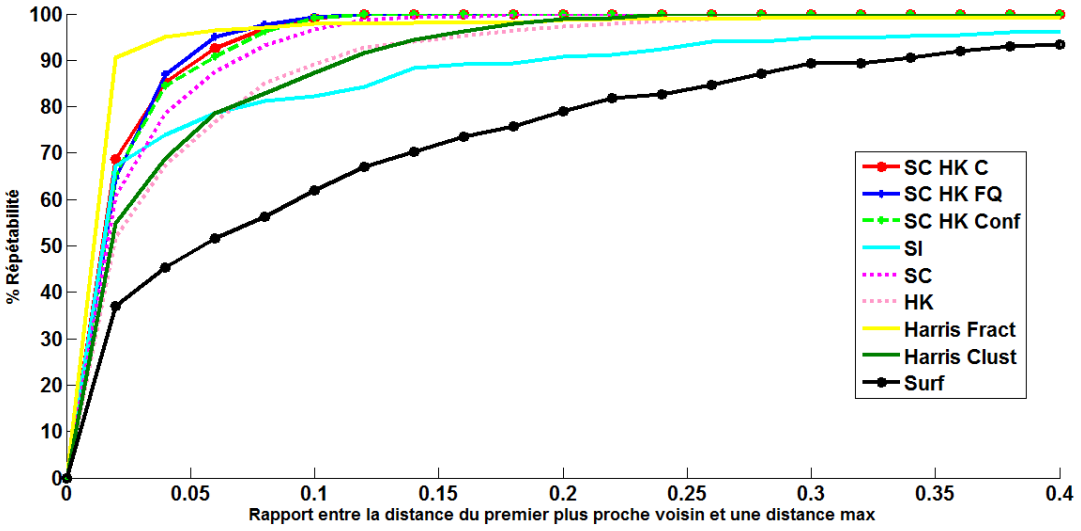


Figure 3-39 Courbes de répétabilité des 9 détecteurs testés sur les 9 objets Minolta et pour un bruit de variance 0.2 MR

Seuil	0.06	0.12
SI	78.7%	84.3%
SC_HK_FQ	**94.9%**	**100%**
SC_HK_C	**92.6%**	**100%**
SC_HK_Conf	90.8%	**100%**
Harris_fract	**96.4%**	97.9%
Harris_clust	78.5%	91.6%
SURF	51.5%	67.0%

Tableau 3-2 Valeurs de répétabilité pour un ajout de bruit gaussien de variance 0.2 MR au maillage

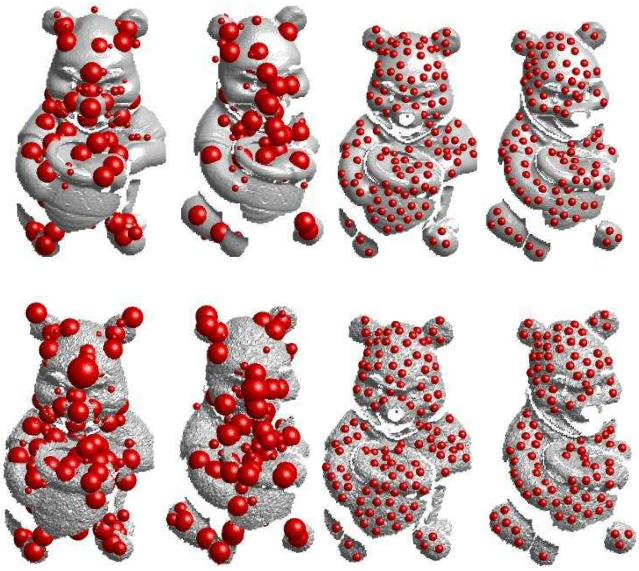


Figure 3-40 Répétabilité des PIs entre les deux vues 0° et 20° initiales (première ligne) et leur version bruitée (deuxième ligne) pour les détecteurs Surf (à gauche) et SC_HK_FQ (à droite)

3.3.2.5 Temps de calcul

Parmi les critères de la faisabilité de ces méthodes de détection et de représentation de la forme pour des applications temps réel, le cout du temps de calcul est principalement décisif. Dans notre cas, nous n'avons pas optimisé au maximum le cout d'exécution de nos algorithmes par contrainte de temps. Cependant, notre performance de calcul reste acceptable pour des applications s'exécutant en temps quasi réel et pour des données 3D pas très denses. Cependant, la précision et les détails de la forme peuvent manquer pour ce genre de données. Nous constatons que la tendance de la plupart des applications du domaine est d'utiliser des caméras 3D de faible résolution comme la Kinect par rapport à des scanners laser plus précis. Dans le Tableau 3-3, nous présentons le temps d'exécution des deux méthodes Monge et RF utilisées pour le calcul de courbures et des normales. L'évaluation est faite sur deux types de données (synthétique et réelle) de densité de points différente.

L'exécutable fournie par les auteurs du SURF 3D exécute la détection et la description en même temps. Dès lors, nous ne pouvons pas estimer le temps de calcul du détecteur SURF. Cette information sera présentée dans le chapitre descripteur. L'exécution du Harris se fait sous le système lunix et nos algorithmes tournent sous le système Windows.

Les propriétés système utilisées correspondent à un processeur Intel CORE i7, 2.8GHz et 4Go de mémoire RAM.

❖ **Résultats** :

Il est à expliquer que dans nos algorithmes, nous calculons, dans l'étape de détection toutes les informations de quantités différentielles qui vont nous servir pour l'étape de détection et description à la fois. Les deux détecteurs Harris_Fract et Harris_Clust, fournissent juste les coordonnées des PIs et une information sur la réponse de Harris en ces points.

Densité	Algo	SC_HK_C	SC_HK_FQ	SC_HK_Conf	SI	Harris_Fract	Harris_Clust
Stuttgart	**Monge**	18.08	32.53	33.9	35.9	0.48	0 .61
10948 points	**RF**	3.5	16.1	16.5	18.06		
Carotte	**Monge**	**9.2**	17.2	17	18.05	0.22	0.24
4685 points	**RF**	1.9	13.8	14.2	16.7		

Tableau 3-3 Temps de calcul des détecteurs en (s)

La remarque est que nos détecteurs sont gourmands en temps de calcul par rapport au Harris. L'exécution de la méthode de Monge (la plus précise en performance) est linéaire avec le nombre de points dans le nuage de points. Le SC_HK_C est le plus rapide des trois détecteurs implémentés puisque les voisins ne sont pas revisités, contrairement au cas de SC_HK_FQ et SC_HK_Conf. L'approche RF est nettement moins couteuse en temps de calcul et peut, du coup, s'exécuter en temps quasi-réel. Cependant, sa précision dans le calcul des normales et courbures est moins bonne.

Il est clair que la linéarité de l'approche de Monge reste handicapante pour une exécution en temps quasi réel. Néanmoins, nous avons pu tester une alternative à cette contrainte, en sous-échantillonnant le nuage de points en des sous régions (patchs). Ainsi, les mesures des courbures se font sur des patchs au lieu des points. Le traitement de recherche de PIs saillants est remplacé par une recherche de patchs saillants (en considérant la classe de la surface du patch (sur les espaces SC et HK) et un tri sur son indice de courbure (C ou FQ ou Conf)). Un PI est sélectionné à la position du point le plus proche du barycentre du patch. Par conséquent, le temps de calcul devient proportionnel au nombre de patch et non plus au nombre de points. Une légère perte d'information sur les détails de la forme, est détectée avec cette approche puisque nous décimons les points (seulement une proportion sera prise en compte (1 point sur 4). Cette perte reste négligeable pour des données denses et pour un pas d'échantillonnage pas très grand. En revanche, nous n'avons pas eu le temps de faire une évaluation de la répétabilité et la robustesse des détecteurs avec les résultats de cette décimation.

3.4. Synthèse

Dans ce chapitre, nous avons, tout d'abord, présenté les principes de quelques détecteurs de points d'intérêt 3D sélectionnés de l'état de l'art. Ce choix n'est pas fait au hasard mais selon, d'un coté, l'idée d'avoir des détecteurs connus (du moins, à notre connaissance) pour pouvoir s'y comparer, et d'un autre coté, la considération des travaux qui nous ont inspiré dans notre démarche. Ensuite, nous avons exposé notre contribution, qui propose de combiner des critères de saillance adoptés dans ces méthodes. Avec quatre façons de sélectionner les PIs finaux après une première extraction, nous avons mis en place nos quatre nouveaux détecteurs combinés: SC_HK_C, SC_HK_FQ, SC_HK_Conf et SC_HK_Con. Nous avons enchainé avec une étude comparative de la performance de nos nouveaux détecteurs par rapport aux approches présentées. Suite aux expérimentations menées, le processus de combinaison des critères (SC_HK) a réussit à surpasser la performance des versions séparées SC et HK. Selon le critère d'évaluation, nos méthodes ont un comportement soit supérieure soit assez proche de celui des méthodes de l'état de l'art étudiées (Harris_3D, SI et SURF_3D).

En résumé, les résultats et les conclusions de cette évaluation sont comme suit :

- Invariance à l'angle de vue
  - Nette supériorité de la répétabilité des détecteurs combinés, en particulier celle de SC_HK_FQ et SC_HK_C
- Invariance à l échelle
  - Performance de répétabilité de nos détecteurs supérieure au SURF et au Harris_Clust, mais inférieure au Harris_Fract
  - Proposition de fixer une distance seuil sur la distance physique des PIs mis en correspondance lors du processus de reconnaissance d'objet
- Robustesse au bruit
  - performance légèrement supérieure du Harris_fract et une meilleure stabilité de nos détecteurs combinés par rapport au reste des détecteurs
  - nécessité d'appliquer un prétraitement de dé-bruitage pour une meilleure performance sur les bases de la Kinect qui sont bruitées
- Temps de calcul de nos algorithmes encore à optimiser pour le cas de données très denses, surtout avec la technique la plus précise ; suggestion de l'idée du sous échantillonnage en considérant des patchs comme base des mesures de saillance, qui permettrait de trouver un compromis entre perte des détails de la forme et optimisation du temps

- Proposition d'une version multi-échelle des détecteurs proposés. L'évaluation de cette extension est donnée en perspective pour des éventuels prochains travaux.

Il est à noter que ces résultats et conclusions sont donnés sur un ensemble de 9 objets de la base Minolta. Bien que le nombre de vues par objet (18 vues) soit acceptable pour notre évaluation, la taille de cet ensemble de test reste petite pour pouvoir faire des généralisations. En outre, la spécificité des objets de cette base est leur caractère naturel (animaux ou peluches). En effet, ces objets se distinguent par leurs détails de la forme et leurs courbures bien accentuées. Dès lors, le comportement et la performance des différents détecteurs est susceptible de changer pour de objets de type mécanique avec la dominance des formes géométriques planes (comme dans l'exemple de la Figure 3-27 illustrant la détection sur quelques objets de la base Stuttgart). A ce niveau, nous pouvons conclure

que ces résultats ont permis de constituer une première idée sur la performance des détecteurs proposés relativement à quelques détecteurs de l'état de l'art.

Certes, les critères d'évaluation, adoptés dans ce chapitre, ont permis de confirmer la stabilité et la robustesse des détecteurs proposés, néanmoins, la détection n'est que la première étape de la chaine d'un système de reconnaissance. La réelle valeur de ces détecteurs se confirmera avec le complément d'une description du voisinage des PIs extraits. La continuité de ce traitement de description sera au cœur du prochain chapitre.

4. CHAPITRE : DESCRIPTEURS

Comme expliqué précédemment, la phase de description de la zone locale (le voisinage) autour des PIs est une étape cruciale pour un système de reconnaissance. La représentation locale donnée par ces descripteurs constitue l'entrée de la phase de mise en correspondance des modèles. Dans ce chapitre, nous présentons quelques descripteurs 3D de la littérature et nous introduisons le principe de nos deux nouveaux descripteurs, inspirés de ces derniers. L'évaluation comparative de ces descripteurs est effectuée sur les aspects suivants : invariance aux transformations géométriques de rotation et l'échelle, pouvoir descriptif, robustesse au bruit et à l'échantillonnage et temps de calcul.

4.1. Présentation des descripteurs testés

Les descripteurs 3D peuvent être divisés en deux catégories selon Tombari et al. (Tombari, et al., 2010): Signature et Histogramme. Une Signature décrit chaque point du voisinage d'un point donné par une ou plusieurs mesures géométriques exprimées en coordonnées locales relatives à un système local de référence invariant (Reference Frame, RF). L'utilisation de ces informations spatiales bien localisées augmente le pouvoir descriptif des descripteurs à base de signature.
Les méthodes basées sur les histogrammes décrivent la région du support par un cumul de mesures géométriques ou topologiques locales (ex. nombre de points, superficie d'un maillage) dans des histogrammes par rapport à un domaine spécifique (ex. coordonnées du point, courbures) et qui nécessite la définition soit d'un axe de référence (Reference Axis, RA), soit d'une RF locale. En général, les descripteurs à base d'histogrammes sont robustes au bruit et à l'échantillonnage grâce au processus de regroupement de l'information de la forme dans des cases d'intervalles. Les approches que nous allons présenter appartiennent à l'une de ces types de catégories: histogramme ou signature ou les deux ensembles.

4.1.1 Descripteur Spin Image

Comme mentionné précédemment dans le chapitre 2, le Spin Image (Johnson, et al., 1999) a été proposé pour décrire des points d'intérêt par Li et Guskov (Li, et al., 2007). Ce descripteur traduit les propriétés locales de la surface dans un système de coordonnées orienté, fixe et lié à l'objet. Ce système est indépendant du changement de vue, contrairement à un système de coordonnées lié à la caméra qui lui dépend de l'angle de vue. Le spin est définie en un point orienté désigné par sa position 3D (p) et sa direction associée (n la normale de la surface locale). Une base 2D de coordonnées locales est formée en utilisant le plan tangent P en p, orienté perpendiculairement à la normale n, et la ligne L passant par p parallèlement à n. Un système de coordonnées cylindriques (α, β) du point p est alors déduit, où α est la coordonnée radiale définissant la distance (non négative) perpendiculaire à L et β est la coordonnée d'élévation qui définit la distance signée perpendiculaire (positive ou négative) à P. La carte des mesures du Spin peut être exprimée comme une fonction de projection des points 3D

de l'objet en coordonnées 2D associés dans la base du point orienté. La projection est comme suit (Johnson, et al., 1998) :

$$S_O: R^3 \rightarrow R^2$$

$$S_O(x) \rightarrow (a, \beta) = \left(\sqrt{\|x-p\|^2 - (n \cdot (x-p)^2}, n \cdot (x-p)\right)$$

Équation 4.1

La Figure 4-1 explique la formation de ce système de coordonnées. Une fenêtre locale de voisinage appelée support est définie pour chaque PI à décrire, et tout point de ce support est exprimé dans la base du PI selon la fonction (Équation 4.1). Les coordonnées cylindriques (α, β) de ces points sont cumulées par la suite dans des cases discrètes d'un tableau 2D selon les équations (Équation 4.3). Pour chaque couple de coordonnées est incrémenté la case correspondante, ainsi que les cases qui l'entourent dans la table pour réduire l'effet du bruit (Équation 4.4). Pour ce faire, la contribution du point 2D est dispersée sur les quatre cases voisines du tableau 2D en utilisant une interpolation bilinéaire. La création de la représentation sous forme de tableau 2D de la Spin Image est illustrée dans la Figure 4-2. Un exemple de spins images générés pour trois points orientés est donné dans la Figure 4-3.

Les paramètres : taille du tableau et taille maximale de l'objet agissent sur le pouvoir descriptif de ce descripteur. La taille maximale de l'intervalle du tableau correspond aux maximums des valeurs de coordonnées $(\alpha_{max}, \beta_{max})$ dans toutes les bases générées des points orientés de l'objet. La taille du bin noté b doit être ajustée de façon à ce que l'image du spin ne soit pas très grande tout en gardant élevé le pouvoir descriptif. La taille de l'image finale (i_{max}, j_{max}) dépend de la valeur de la taille du bin et est calculée comme suit (Johnson & Hebert, 1998):

$$i_{max} = \frac{2\beta_{max}}{b} + 1 \qquad j_{max} = \frac{a_{max}}{b} + 1$$

Équation 4.2

D'après Johnson et Hebert (Johnson, et al., 1999), un meilleur résultat de l'appariement des descripteurs Spin, indépendant de la taille et de la résolution du modèle (densité des points), est obtenu pour une taille du bin multiple de la résolution du maillage.

Le cumul des valeurs 2D (α, β) des points dans des valeurs discrètes correspondant aux bins de l'histogramme se fait selon les équations suivantes :

$$i = \left\lfloor \frac{\beta_{max} - \beta}{b} \right\rfloor \qquad j = \left\lfloor \frac{a}{b} \right\rfloor$$

Équation 4.3

Les poids bilinéaires des attributs servant à l'incrémentation des bins sont :

$$\mathrm{a} = \alpha - ib \qquad b = \beta - jb$$

Équation 4.4

Avec cette définition, le pouvoir descriptif des Spins Images permet d'apparier convenablement deux surfaces différentes d'un même objet. Pour une variation de l'échantillonnage des deux surfaces ou pour une présence de bruit, les deux spins images sont légèrement différents. Cependant avec un échantillonnage uniforme de la surface (les arêtes des facettes ont presque la même longueur) les spins images des points des deux surfaces sont linéairement corrélés.

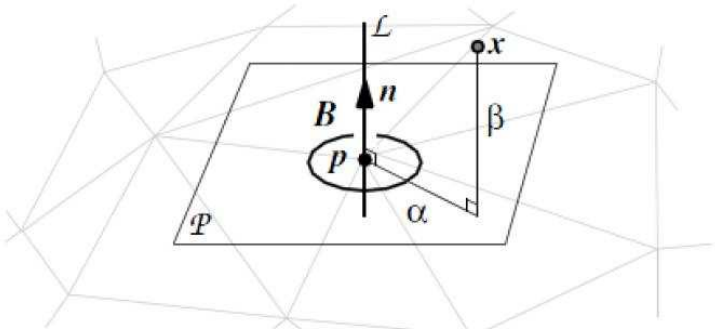


Figure 4-1 Formation du système de coordonnées cylindriques pour un point orienté (Johnson, et al., 1998)

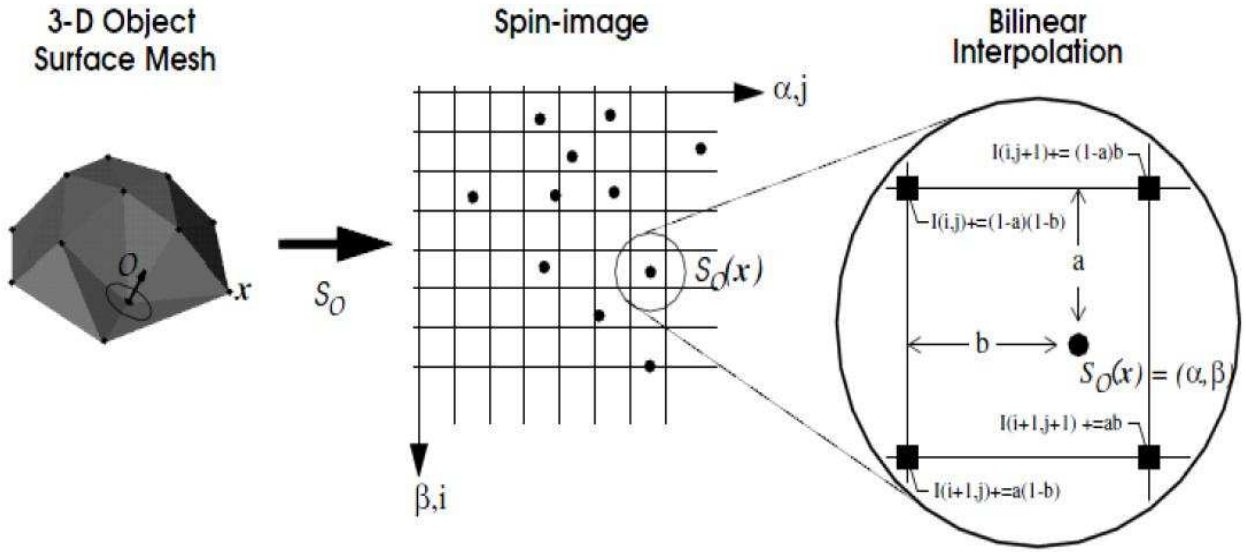


Figure 4-2 Processus de création de la représentation 2D du spin image (Johnson, et al., 1998)

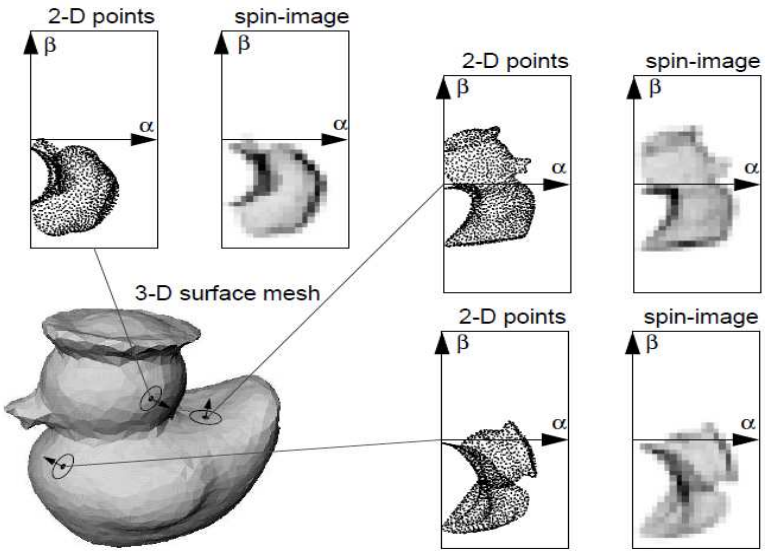


Figure 4-3 Représentation des Spin-images pour deux valeurs de support pour trois points orientés de la surface du modèle de Duck. Dans l'accumulateur, les zones foncées correspondent à plus de points projetés dans les cases (Johnson, et al., 1998)

4.1.2 Descripteur Thrift

Le deuxième descripteur considéré dans notre étude comparative est le Thrift, proposé par A. Flint et al. (Flint, et al., 2007) comme extension du SIFT 2D. Ce dernier se compose d'un histogramme cumulant, pour un voisinage du PI, le cosinus de l'angle entre deux normales calculées sur deux supports. En effet, pour chaque point d'intérêt, un support des points voisins se situant à une distance inférieur à une distance seuil σ du point référence, et est défini par :

$$Support(\mathbf{z}) = \{\mathbf{y} \in \mathcal{X} : \|\mathbf{y} - \mathbf{x}\| \leq \sigma\}$$

Équation 4.5

Pour chaque point y du support deux fenêtres W_1 et W_2 de points correspondant à deux valeurs de distances w_{small} et w_{large}, et sont définies par :

$$W_1 = \{\mathbf{p} \in \mathcal{X} : \|\mathbf{p} - \mathbf{y}\| \leq \omega_{small}\}$$
$$W_2 = \{\mathbf{p} \in \mathcal{X} : \|\mathbf{p} - \mathbf{y}\| \leq \omega_{large}\}$$

Équation 4.6

En pratique, les normales de surface sont approximées avec des plans de moindre carrées. Soit, deux plans P_1 et P_2 correspondent respectivement aux fenêtres W_1 et W_2, et n_{small} et n_{large} leurs normales respectives (Figure 4-4), qui peuvent être interprétées comme les courbures principales de la carte de densité (présentée dans le détecteur Thrift, section 2.3.1.2). Les distances w_{small} et w_{large} sont constantes pour tout les points d'intérêt et proportionnelles au rayon de référence du point. Rappelons qu'une échelle σ est donnée en sortie du détecteur, les valeurs utilisées, dans les expérimentations de (Flint, et al., 2007), sont: $w_{small} = 0.3 * \sigma$ et $w_{large} = 0.8 * \sigma$.

Le descripteur au point z est donné par l'histogramme des valeurs des cosinus de l'angle entre les n_{small} et n_{large} des points y appartenant au support :

$$\cos(\theta) = \frac{\mathbf{n_{small}} \cdot \mathbf{n_{large}}}{\|\mathbf{n_{small}}\| \|\mathbf{n_{large}}\|}$$

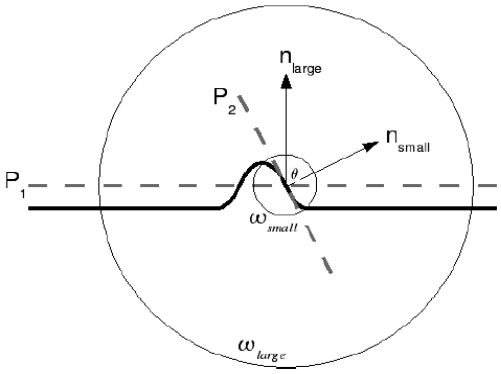


Figure 4-4 Deux plans par moindre carrés P_1 et P_2 et leurs normales (Flint, et al., 2007)

L'intervalle des valeurs du cosinus est [0, 1] pour une variation d'angle entre 0° et 90° et le cumul de ces valeurs dans les cases de l'histogramme est normalisé. La dimension du descripteur final dépend du nombre de cases donné en paramètre.

Le descripteur Thrift se caractérise par son invariance aux rotations 3D, due à son principe de comparer des normales estimées sur deux échelles. De plus, l'utilisation des plans des moindres carrés pour l'estimation des normales assure l'invariance à la densité des points du maillage.

4.1.3 Local surface patches (LSP)

Le descripteur LSP "Local Surface Patches", introduit par Chen et Bhanu (Chen, et al., 2004), est formé d'un histogramme 2D qui cumule, à la fois, l'information des cosinus de l'angle entre les normales des points du support et la normale référence du PIs, et la valeur de l'indice de forme de ces points. La région des voisins N désigne le support et est définie par l'Équation 4.7 :

$$N = \{pixels\ \ N, \|N - P\| \leqslant \epsilon_1\} \text{ and } acos(n_p \bullet n_n) < A.$$

Équation 4.7

Dans cet équation, le • désigne le produit scalaire entre la normale n_p au point référence p, et la normale n_n d'un voisin appartenant à N. Le cosinus est égal au produit scalaire entre les deux normales et acos est la fonction de cosinus inverse. Les paramètres $\varepsilon_1$et A contribuent à la performance du descripteur en terme de pouvoir discriminant de la représentation locale du patch. Pour chaque point N_i du voisinage N, l'indice de forme S_i et le cosinus de l'angle entre sa normale et celle du point référence sont calculés. Un histogramme 2D est formé en cumulant les points dans des cases de deux axes relatifs à la valeur de l'indice de forme et à l'angle des normales selon l'Équation 4.8. L'axe des indices de forme varie entre 0 et 1 et l'axe du cosinus prend ses valeurs dans [- 1, 1]. Dans l'Équation 4.8, l'opérateur de crochet correspond à l'opérateur d'arrondissement des valeurs réelles (float) en la valeur de l'entier le plus proche et de plus petite valeur. Dans cette équation, h_x est l'indice sur l'axe

horizontal et v_y est l'indice sur l'axe vertical. b_h et b_y correspondent aux bins des intervalles sur les deux axes. Une interpolation bilinéaire est utilisée pour réduire l'effet du bruit dans le calcul de l'histogramme 2D. L'exemple de la Figure 4-5 montre un histogramme LSP sous forme d'image en niveau de gris. Les pixels clairs correspondent aux bins dont le nombre de points cumulés est grand.

$$h_x = \left\lfloor \frac{S_i}{b_h} \right\rfloor, \quad v_y = \left\lfloor \frac{\cos\theta + 1}{b_y} \right\rfloor$$

Équation 4.8

En plus de l'histogramme 2D, la représentation LSP inclut le type T_p obtenu par les mesures des courbures gaussiennes et moyennes de la surface locale (une valeur de 1 à 9) et le centroïde du support. L'ensemble est utilisé pour la phase de mise en correspondance des PIs.

Dans ce chapitre, lors de notre implémentation et évaluation de ce descripteur, nous limitons la représentation LSP à l'histogramme 2D uniquement. Notons que dans la formulation de ce descripteur, la notion de RA unique et non ambigüe est utilisée par le biais du vecteur normal du point référence.

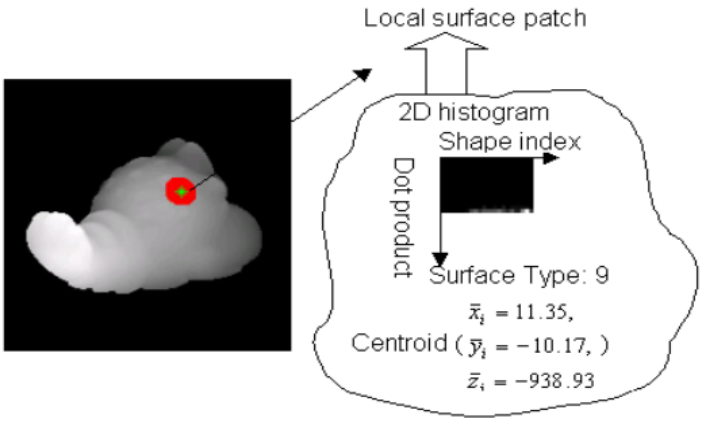


Figure 4-5 Histogramme 2D des indices de formes vs le produit scalaire entre la normale du point référence et ceux de ses voisins dans la représentation LSP (Chen, et al., 2007)

4.1.4 IndThrift

A l'issu de notre étude de ces deux derniers descripteurs, nous avons proposé un nouveau descripteur qui mixe la formulation du descripteur LSP avec celle du Thrift : une variante du descripteur Thrift sous forme d'histogramme 2D. Le nouveau design s'inspire du formalisme de l'histogramme 2D du LSP et remplace la mesure du cosinus entre la normales référence et les normales du voisinage par celle proposée dans le Thrift. Nous appelons ce descripteur: IndThrift (Indice de forme + Thrift).

Plus clairement, pour chaque point du support du PI, deux rayons sont définis et deux normales sont associées à chacune des deux fenêtres correspondantes à ces deux rayons. La formation de l'histogramme 2D se fait comme suit: Pour chaque point v du voisinage du PI, s'incrémente une case de l'histogramme correspondant au couple (indice de forme de v, angle entre les deux normales de v). L'axe du cosinus cumule les valeurs des cosinus de l'angle entre les deux normales de chaque point du support du PI. L'axe de l'indice de forme cumule les valeurs de l'indice de forme des points appartenant au voisinage. Un passage par une interpolation bilinéaire sur les deux axes apporte plus de

robustesse et de stabilité à ce nouveau descripteur.). Notons que ce nouveau descripteur ne définit pas une base locale référence associée au PI (une propriété héritée du descripteur Thrift). Dans notre étude comparative, nous suggérons d'implémenter, puis d'évaluer ce descripteur en termes de pouvoir descriptif, invariance aux rotations et à l'échelle, et robustesse au bruit et à la densité des points.

4.1.5 Descripteur SHOT

Un troisième descripteur qui a inspiré notre démarche de description est le SHOT (Signature of Histograms of OrienTations) proposé par Tombari et al. (Tombari, et al., 2010). Les atouts de ce descripteur de forme regroupent, d'un coté, la définition d'un système de référence 3D unique et non ambiguë (RF), et d'un autre coté l'utilisation d'une information géométrique sur la position des points dans le support. Le pouvoir descriptif et la robustesse au bruit, occultation, translations et rotations, ont été démontré par (Tombari, et al., 2010) dans une étude comparative.

L'estimation des RFs locales se fait par le calcul des directions des vecteurs propres (EVD : Eigen Vector Directions) d'une nouvelle matrice de covariance inspirée de la méthode du Total Least Squares (TLS) pour l'estimation des directions des normales. Cette matrice M est définie comme combinaison linéaire des distances des points appartenant à un support sphérique du PI et donnée par :

$$\mathbf{M} = \frac{1}{\sum_{i:d_i \leq R} (R-d_i)} \sum_{i:d_i \leq R} (R-d_i)(\mathbf{p}_i - \mathbf{p})(\mathbf{p}_i - \mathbf{p})^T$$

$$avec\ \ d_i = \|p_i - p\|_2$$

Équation 4.9

Dans cette expression, un poids faible est attribué aux points les plus éloignés et tous les points contenus dans un support sphérique de rayon R sont utilisés pour le calcul de M. Le centroide du support est remplacé par le point référence p. Les vecteurs propres de cette matrice forment des directions orthogonales, répétables et robustes au bruit. La normale correspond au vecteur propre associé à la plus faible valeur propre. A ce niveau pour avoir une RF répétable, les vecteurs propres doivent être non ambigus. L'idée est de réorienter le signe de chaque vecteur propre d'une façon à ce qu'il soit cohérent avec le signe de la majorité des vecteurs qu'il représente. Le signe des axes x et z est déterminé avec ce principe et l'axe y est obtenu par le produit vectoriel z × x.

Le descripteur est formé par des histogrammes locaux définis sur des volumes 3D dans une grille sphérique 3D superposée sur le support des points à la manière du SIFT. Cette grille partitionne l'espace selon les axes radial, azimut et élévation. Dans la formulation du descripteur SHOT, le nombre de bins spatial est fixé à 32 qui est le résultat de 8 divisions en azimut, 2 divisions en élévation et 2 divisions en radial, représentées dans la Figure 4-6.

Pour chaque histogramme local, les points sont cumulés en fonction de l'angle, θ_i, entre la normale de chaque point n_{vi} du support et la normale n_u du PI. La fonction $\cos\theta_i = n_u \cdot n_{vi}$ est déduite pour chaque couple de normales. La répartition de la taille des cases ne se fait pas uniformément sur l'intervalle des valeurs du cosinus. Un pas grossier est donné pour les cases correspondant aux directions proches de la normale référence et un pas plus fin est utilisé pour les cases ou les directions sont orthogonales. En effet, les directions les plus orthogonales à la direction de la normale référence coïncident avec des zones de forte variation locale de la forme (une zone de forte courbure) et sont alors plus riche en

informations. Ce choix permet également de réduire l'impact du bruit qui peut subsister sur des régions quasi planes et qui sont moins descriptives. Sur ces régions, le cumul des cosinus se fait sur un petit nombre de cases de l'histogramme, ce qui permet de réduire la différence entre les histogrammes bruités lors de la mise en correspondance.

La généralisation du design du SHOT est présentée dans ce qui suit.

Notons par $SH_{G,f}$ (P) la signature générique des histogrammes sur le support sphérique autour d'un point P. L'entité G est le vecteur des valeurs des propriétés du sommet, et l'entité f est la métrique utilisée pour comparer ces propriétés. Pour calculer un histogramme de signature, la fonction f est appliquée sur toutes les paires (G_P, G_Q), avec Q représentant un sommet générique appartenant au support sphérique autour de P. Dans la formulation du SHOT, G correspond à l'estimation de la normale, N, et f(.) est le produit scalaire noté par p(.) ;

$$f\left(G_P, G_Q\right) = p\left(N_P, N_Q\right) = N_P \cdot N_Q$$

Équation 4.10

Après le calcul de m signatures d'histogrammes relatives à différentes paires (propriété, métrique) sur la sphère, ces histogrammes sont juxtaposés (Équation 4.11) pour former le descripteur final D(P) au point P.

$$D(P) = \bigcup_{i=1}^{m} SH^{i}_{(G,f)}\left(P\right)$$

Équation 4.11

Dans la structure du SHOT, un effet de bord peut apparaitre pour une perturbation dans le calcul de la RF locale. Pour réduire l'impact d'une telle perturbation, une interpolation quadra-linéaire est faite, premièrement, sur les cases voisines dans l'histogramme local, et deuxièmement, sur les cases ayant le même indice dans l'histogramme et qui correspondent aux volumes voisins dans la grille spatiale. Plus précisément, pour chaque nouvelle entrée l'incrémentation se fait par une multiplication d'un poids égale à 1 – d dans chaque dimension. Au niveau de l'interpolation dans l'histogramme, d est la distance de l'entrée actuelle par rapport à la valeur centrale de la case. Dans la grille spatiale, d est la distance angulaire de l'entrée par rapport à la valeur du centre du volume, pour les deux dimensions relatives à l'élévation et à l'azimut. Et concernant la dimension radiale, d est la distance euclidienne de l'entrée par rapport à la valeur du centre du volume. Selon la dimension en question, la distance d est normalisée par la distance entre les voisins des cases de l'histogramme ou par la distance entre les volumes de la répartition spatiale. La robustesse aux variations de l'échantillonnage est réalisée par une normalisation du descripteur. En effet, les différences locales sont la source de l'information discriminante du descripteur. De plus, cette normalisation assume que la densité de l'échantillonnage des points varie d'une façon globale, contrairement à la méthode de Frome et al. (Frome, et al., 2004). L'approche pointée par (Johnson, et al., 1999) pour éliminer l'effet de bord est également utilisée. Ce formalisme a permis d'augmenter la répétabilité et la robustesse du descripteur tout en optimisant le temps de calcul. Les paramètres influant sur la performance du SHOT sont essentiellement le rayon du support et le nombre de bins dans l'histogramme.

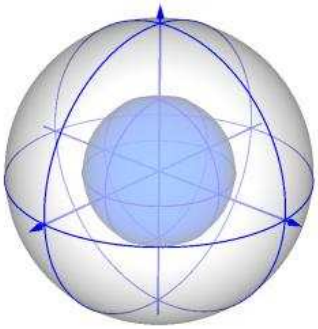


Figure 4-6 Structure spatiale du descripteur SHOT: 8 divisions en azimut, 2 divisions en élévation et 2 divisions en radial (par soucis de clarté seulement 4 divisions azimut sont représentées) (Tombari, et al., 2010)

4.1.6 CSHOT

La formulation du descripteur SHOT a été étendue pour ajouter l'information de la texture à la représentation de la forme et a donné le descripteur Color-SHOT (CSHOT) (Tombari, et al., 2011). Ce descripteur est composé d'une juxtaposition des histogrammes des normales et des histogrammes de couleurs sur le support de la grille sphérique. Une meilleure performance que celle du SHOT est réalisée avec cette version combinée sur des bases de données publiques.

D'après la formulation généralisée (Équation 4.11) du SHOT, le paramètre m prend la valeur 2 dans le cas particulier du CSHOT pour exprimer la combinaison de deux signatures : celle des histogrammes de mesures de forme et celle des mesures de texture. La première signature correspond exactement à la formulation originale du SHOT. Pour la définition de la deuxième signature, un vecteur des intensités des triplets RGB désigné par R représente les propriétés de la texture et la métrique L_1 (Équation 4.12) est utilisée pour comparer les triplets RGB entre eux. Au final, l'histogramme de la texture est formé par les valeurs de la somme absolue des différences entre les valeurs des triplets RGB du point référence et ceux des points voisins dans le support.

$$l\left(R_P, R_Q\right) = \sum_{i=1}^{3} \left|R_P(i) - R_Q(i)\right|$$

Équation 4.12

Le descripteur CSHOT hérite des paramètres du SHOT à savoir le rayon du support et le nombre de bins dans chaque histogramme. Un nombre de bin différent pour chaque type d'histogramme est utilisé étant donné la différence dans la nature des deux histogrammes de signatures.

Certes le pouvoir descriptif du CSHOT a amélioré la performance du SHOT, néanmoins, l'information de la texture n'étant pas donnée par tous les capteurs 3D, l'utilisation de ce descripteur est restreinte.

4.1.7 IndSHOT

Inspiré des formulations des deux descripteurs décrits plus haut le LSP et CSHOT, nous avons proposé de définir un nouveau descripteur qui combine les principes de ces deux derniers pour les raisons suivantes :

- Le succès du descripteur CSHOT revient, d'une part, à la définition d'une RF robuste et invariante aux transformations géométriques et, d'une autre part, à la façon de juxtaposer et normaliser les signatures de deux histogrammes. En comparaison à l'état de l'art, l'hybridité signature-histogramme de ce descripteur a permis, grâce à la propriété des signatures, de hausser le pouvoir descriptif et d'augmenter la robustesse au bruit et à l'échantillonnage, une caractéristique héritée de celles des histogrammes. Comme l'attestent les résultats dans (Tombari, et al., 2011), le CSHOT a surpassé la performance de la forme originale SHOT avec l'ajout de l'information de texture (couleur). Cependant, cet apport n'est exploitable qu'en présence de données de texture (couleur). L'idée, à ce niveau, est de proposer une autre mesure qui caractérise la forme au lieu de la texture, et donc de substituer à la couleur cette mesure dans la formulation du CSHOT.
- Concernant le descripteur LSP, l'histogramme 2D proposé combine deux mesures de la forme locale sur le support: la première est le cosinus des normales qui est invariant aux transformations géométriques 3D et la deuxième est la valeur de l'indice de forme qui est invariante à l'échelle et à la résolution.

Suite à ces constations, notre suggestion est d'utiliser l'histogramme de mesures de l'indice de forme comme deuxième signature d'histogramme à la place de l'histogramme des mesures de la texture utilisé dans le CSHOT. Dans le design généralisé du SHOT, nous aurons, pareillement que pour le CSHOT, une première signature d'histogrammes des mesures du cosinus de l'angle entre la normale référence et les normales des points du support, une deuxième signature d'histogrammes de mesures des différences entre la valeur de l'indice de forme du point référence et celle de son voisinage. Pour cette deuxième signature, G est l'indice de forme noté par IF et f est la norme L_1 notée par t(.) dans l'Équation 4.13.

$$t(IF_P, IF_Q) = |IF_P - IF_Q|$$

Équation 4.13

Dans notre implémentation de ce descripteur, la normale d'un sommet d'une facette est calculée par l'ajustement d'un plan aux points qui lui sont connectés dans la facette et la normale du point référence (le PI) correspond à la RF. Nous notons ce nouveau descripteur par IndSHOT (Indice de forme + SHOT). Une illustration du design du IndSHOT est donnée dans la Figure 4-7.

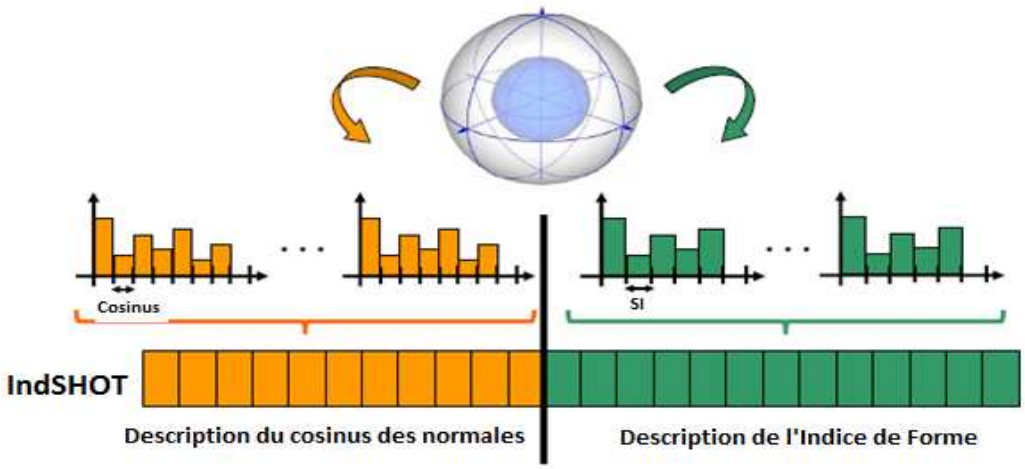


Figure 4-7 Représentation du descripteur IndSHOT

Un résumé des descripteurs présentés concernant leur catégorie et leur définition d'une RF locale unique et non ambigüe est donné dans le Tableau 4-1. Dans la deuxième section de ce chapitre, nous allons évaluer la performance de ces descripteurs.

Méthode	Catégorie	Définition d'une référence
Spin Image (Johnson, et al., 1999)	Histogramme 2D des coordonnées radiale vs les coordonnées d'élévation des points voisins au PI	Axe de référence
LSP (Chen, et al., 2007)	Histogramme 2D des cosinus entre la normale du PI et celle des points voisins vs leurs indices de forme	Axe de référence
Thrift (Flint, et al., 2007)	Histogramme 1D des cosinus entre les normales des points voisins, calculées sur deux rayons	_
IndThrift	Histogramme 2D des cosinus entre les normales des points voisins calculées sur deux rayons vs leurs indices de forme	_
SHOT (Tombari, et al., 2010)	Histogramme 1D des cosinus entre la normale du PI et celle des points voisins	RF Locale
IndSHOT	Histogramme 1D joignant les cosinus entre la normale du PI et celle des points voisins, et les différences de leurs indices de forme	RF Locale

Tableau 4-1 Catégorisation des cinq descripteurs

4.2. Evaluation Expérimentale

Dans cette partie, la distance choisie pour mesurer la similarité entre les histogrammes des descripteurs est la distance euclidienne comme utilisée par (Tombari, et al., 2010) dans leur étude comparative.

4.2.1 Paramètres

Une analyse quantitative des descripteurs est menée pour déterminer les valeurs des paramètres des descripteurs d'une façon qui augmente leur performance et optimise le temps de calcul pour la phase de mise en correspondance.

Dans les expérimentations de cette partie, la base utilisée correspond aux 9 modèles de la base Minolta introduite dans le chapitre Détection. L'évaluation de l'influence de chaque paramètre sur le résultat de la mise en correspondance est donnée par un graphe généré de la façon suivante:

- Pour une valeur du paramètre étudié, le descripteur est calculé sur m positions de points sélectionnés aléatoirement sur le modèle original. Puis, un lissage Laplacien de facteur de relaxation f= 0.15 x MR (résolution du maillage), est appliqué sur les données du modèle original et le descripteur est recalculé sur les m mêmes positions du modèle transformé (lissé).
- Par la suite, chaque descripteur obtenu du modèle transformé est comparé aux autres descripteurs du modèle d'origine pour trouver les deux premiers plus proches voisins en termes de distance euclidienne entre les histogrammes des descripteurs.
- Nous avons une mise en correspondance entre une paire de descripteurs si le ratio entre la distance euclidienne du premier plus proche voisin et la distance euclidienne du deuxième plus proche voisin est inférieure à un seuil fixé à 0.75. Une paire de descripteurs appariés est dite correcte si elle décrit la même position spatiale dans les modèles où sont calculés les deux descripteurs.
- Ce processus est répété pour tous les descripteurs calculés sur les vues de chaque modèle de la base Minolta. Pour chaque type de descripteur, la moyenne des rappels et des précisions est déduite des vues de chaque objet. Les résultats correspondent à la médiane des taux de rappels et de précisions obtenus sur tous les 9 objets.

Le rappel calcule la fraction du nombre des correspondances correctes sur le nombre total des descripteurs en test. La précision est la fraction du nombre des correspondances correctes sur le nombre total des correspondances. Chaque courbe du graphe correspond à un type de descripteur et indique l'évolution du rappel/précision en fonction du paramètre étudié.

En général, selon le type du descripteur, des traitements spécifiques peuvent améliorer le résultat de la mise en correspondance. Par exemple, dans le cas d'un histogramme avec des « creux » (sparse), la distance euclidienne se calcule sur l'ensemble des paires des cases non vides et de même indice. Nous ne tenons pas compte de ce genre de traitements pour cette partie du manuscrit. Le paramètre m varie entre 50 et 150 points, selon la densité de points des vues, pour toutes les expérimentations de cette section.

Le paramétrage se fait pour les descripteurs Spin, SHOT, LSP, Thrift, IndSHOT et IndThrift. Notons que pour les descripteurs (LSP, IndSHOT et IndThrift), le calcul de l'indice de forme n'est pas effectué dans l'étape de description mais est récupéré de la phase de détection. Le temps de calcul présenté dans cette partie n'inclue pas ce calcul.

4.2.1.1 Choix du nombre de Bins

❖ **<u>Description</u>**

La taille des Bins est un paramètre important dans la génération des descripteurs à base d'histogrammes et correspond à la largeur géométrique des cases dans les histogrammes. En effet, ce paramètre détermine la marge de l'espacement dans les cases de l'histogramme nécessaire pour répartir et moyenner les valeurs de la variable cumulée et éviter le cas des points isolés. Un espacement grossier ou plus fin impacte directement sur le pouvoir de description.

La taille du bin dépend du nombre de cases (ou dimension) choisi pour l'histogramme et vice-versa. Le produit de la taille du bin avec le nombre de bins donne la taille de l'intervalle de variation des valeurs, qui correspond à la différence entre la valeur maximale et la valeur minimale de la mesure cumulée.

Une analyse quantitative de l'impact du nombre de bins sur la similarité entre descripteurs appariés est donnée dans la Figure 4-8. Pour simplifier cette étude, nous prenons n comme nombre de bins pour la description du cosinus des normales et n-2 comme nombre de bins pour la description de l'indice de forme et des coordonnées cylindriques dans le Spin. Ce qui nous donne une dimension de $(n-2)^2$ pour le descripteur Spin, n (n-2) pour les descripteurs LSP et IndThrift, n pour le Thrift, 32 (n+1) pour le descripteur SHOT et 32((n-1)+ (n+1)) pour le descripteur IndSHOT. Le 32 correspond au nombre de bins spatial. En variant la valeur du nombre de bins, nous traçons la courbe précision et la courbe rappel. La performance des descripteurs est déduite du comportement des ces courbes: plus la valeur du rappel et de la précision est grande plus le descripteur est performant.

❖ **Résultats**

A ce niveau, nous cherchons la valeur n des bins qui puisse maximiser le rappel et précision pour la majorité des descripteurs. D'après ces courbes, d'une manière générale, en augmentant le nombre de bin, la précision et le rappel augmentent légèrement et se stabilisent pour le SHOT et IndSHOT, de même pour le comportement de la courbe de précision du LSP, Spin et IndThrift. Quant au rappel, il augmente puis diminue pour le LSP, Spin et IndThrift. Le Thrift continue à augmenter pour les deux courbes.

Par ailleurs, nous remarquons que les valeurs du rappel et de précision pour le IndSHOT sont les plus élevés et celle du Thrift s'écarte largement des autres courbes lui inférant une faible performance.

En outre, pour une valeur fixe du nombre de bin, ces courbes renseignent aussi sur le degré de robustesse des descripteurs pour un lissage de la structure du maillage. Nous allons voir cette propriété plus tard dans ce chapitre.

Pour le cas des petites valeurs du nombre de bin, plusieurs valeurs sont moyennées pour correspondre à une seule case et le pouvoir discriminant du descripteur diminue. En effet, une structure très condensée de l'information ne traduit plus la distribution réelle de la forme. Le résultat est que plusieurs histogrammes auront la même forme sans pour autant correspondre à des régions similaires de la surface. Par ailleurs, pour des grandes valeurs du nombre de bins, la distribution devient éparse (avec beaucoup de cases vides) et l'information locale est dispersée sur plusieurs cases en donnant des points singuliers (isolés). Certes cette description est la plus proche de la distribution réelle de la mesure cumulée, mais nous perdons la compacité et la robustesse au bruit et à la densité des points. Comme l'illustre la courbe de la Figure 4-8, plus le nombre de bins augmente, plus le rappel décroit traduisant la diminution des correspondances correctes. Pour des valeurs du bin dans l'intervalle 10 à 20, et mis à part le Thrift, la précision des descripteurs est au dessus de 0.9 et le rappel est quasiment à son maximum.

Le temps de calcul des descripteurs n'est pas affecté par le choix du nombre de bins. Cependant, ce paramètre influe sur le temps de calcul pour la phase de mise en correspondance entre les descripteurs.

- Le choix du nombre de bins adéquat doit se faire sur la valeur qui satisfait le compromis entre, d'un coté, la compacité et la robustesse (une faible valeur du bin) du descripteur et d'un autre coté, son pouvoir descriptif (une grande valeur du bin). Notre choix est porté sur la valeur **n = 12**.

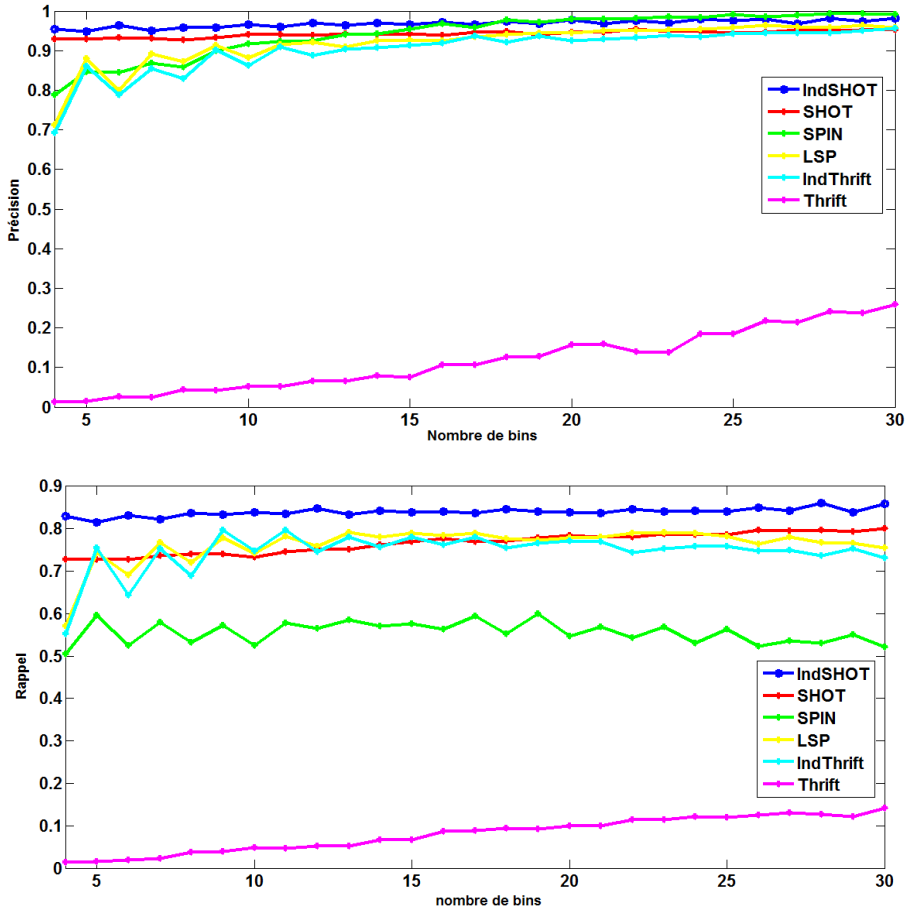


Figure 4-8 Courbe rappel et courbe précision en fonction du nombre de bins

4.2.1.2 Choix du rayon du support

❖ **Description**

Dans ce paragraphe, nous nous intéressons à la valeur du deuxième paramètre: rayon du support r utilisé dans le calcul des descripteurs des PIs. Du fait que nous travaillons à une échelle fixe, la phase de détection ne nous fournit pas une échelle de description. Le rayon du support utilisé pour le calcul des descripteurs est le même pour tous les PIs et est fixé en paramètre par l'utilisateur. Sa valeur est proportionnelle à la taille de la diagonale de la boite englobante de l'objet pour assurer l'invariance à la densité des points et à l'échelle. Notons que le rayon du support dans le Spin Image est le produit de la taille du bin par le nombre de bins dans chaque dimension de la Spin Image.

Dans la Figure 4-9, nous traçons la courbe rappel et la courbe précision en fonction du rayon du support. Et dans la Figure 4-10, nous montrons l'évolution du temps de calcul des descripteurs en fonction de la taille du rayon, pour une vue de l'objet Duck.

❖ **Résultats**

Le rayon du support du PI influe sur le distance de mise en correspondance de la façon suivante: D'une part, plus le rayon est grand plus le voisinage est grand et plus les mesures calculées (indice de forme ou cosinus entre les angles de normale ou les coordonnées cylindriques) sur ce support sont lissées et stables. D'autre part, pour un rayon de support plus grand, des nouveaux voisins plus éloignés sont ajouté et des nouvelles mesures sont cumulées dans l'histogramme. L'allure de

l'histogramme obtenu est modifiée par rapport à celle d'un histogramme résultant d'un rayon de plus petite taille. Le type de la forme locale autour du PI intervient aussi sur la stabilité du descripteur : une zone plane sur un rayon r est plus stable pour un changement de rayon qu'une zone sous forme de dôme, par exemple.

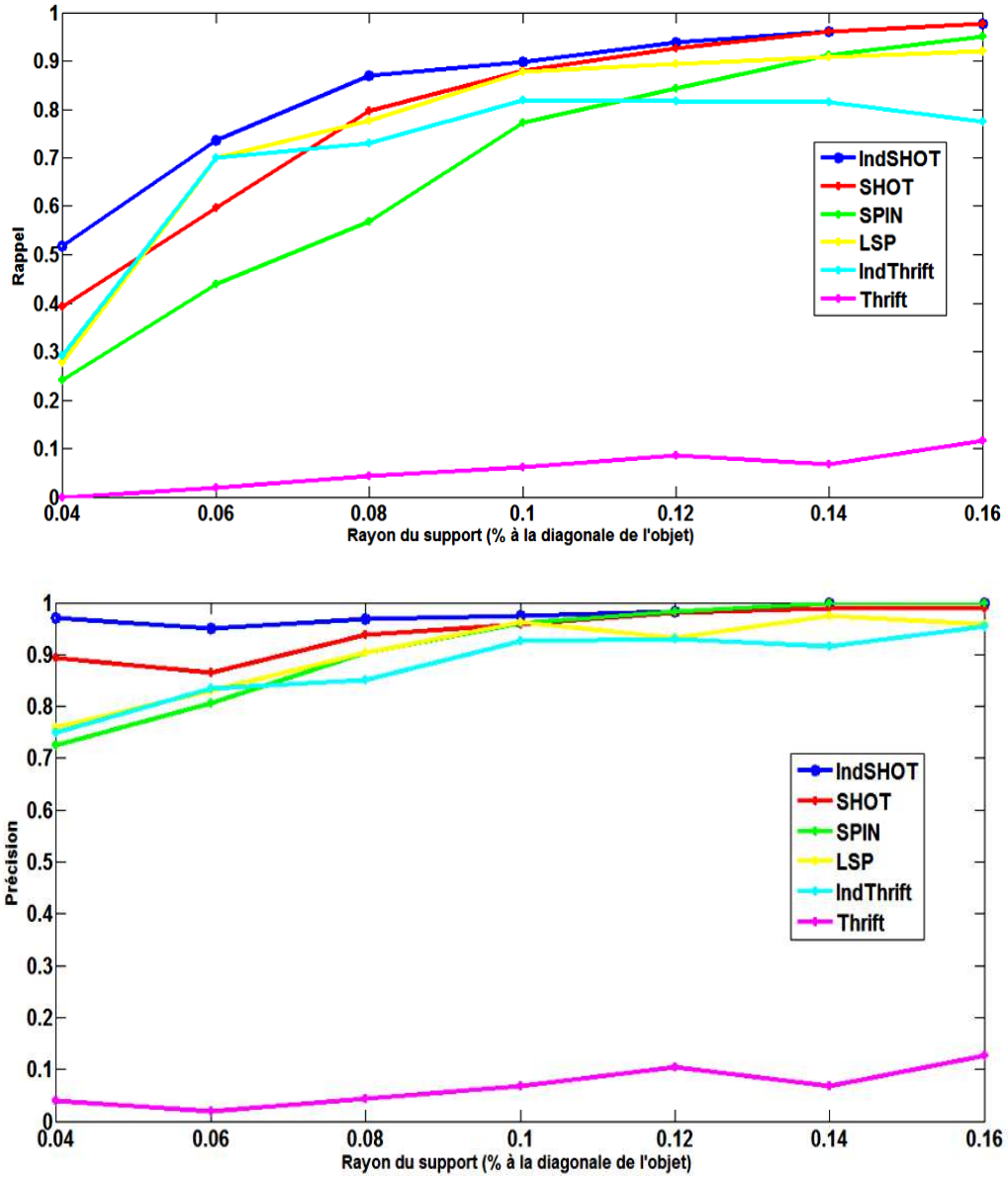


Figure 4-9 Courbe précision et courbe rappel en fonction du rayon du support

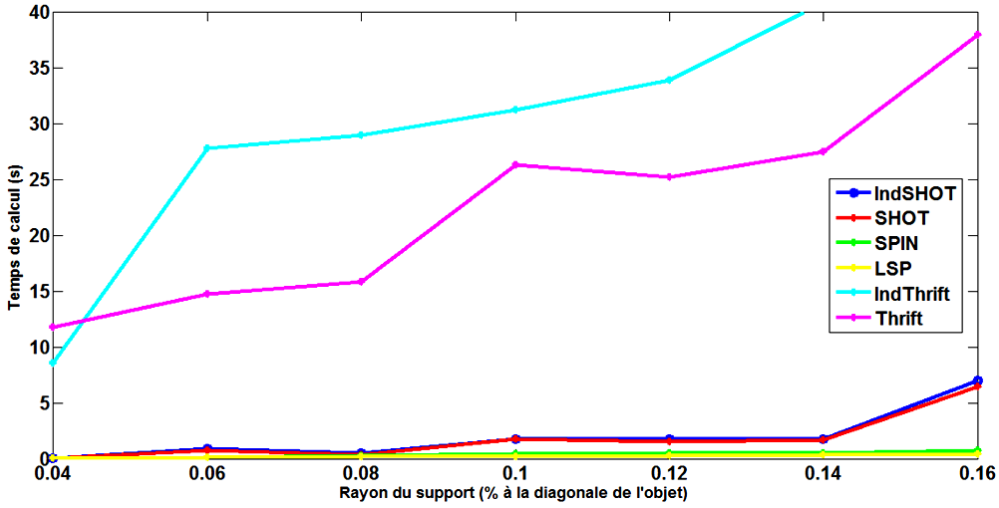


Figure 4-10 courbe temps de calcul en fonction du rayon du support

Ces remarques sont consolidées par le résultat de la Figure 4-9 où l'appariement est plus correct pour des plus grandes valeurs du rayon (entre 0.08 et 0.16) pour tous les descripteurs. Le rappel augmente puis décroit, pour le IndThrift entre les valeurs 0.08 et 0.12 du rayon. Nous pensons, qu'à partir d'une certaine valeur du support, le descripteur devient stable puis sa performance décroit (tel le cas pour le IndThrift). Cette valeur représente la taille limite et maximale du support pour laquelle la structure locale décrite change de forme et par conséquent, le descripteur change d'allure. D'après les deux courbes rappel et précision, cette limite pourrait être atteinte à partir de la valeur 0.08 x diagonale du support. Or, pour obtenir un meilleur pouvoir descriptif, le support doit avoir une taille proportionnelle à une structure de petite taille pour pouvoir moyenner suffisamment de détails localement sur la surface. Les structures de grandes tailles nous ramènent à une description de l'ordre des propriétés globales des modèles et nous sortons d'une approche locale.

Par ailleurs, la contrainte majeure est le temps de calcul qui clairement croît avec l'augmentation de la taille du support. Nous notons que le temps de description du IndThrift et du Thrift, pour les faibles tailles du rayon est d'environ 10 fois plus important que celui des autres descripteurs.

- Le choix de la valeur du rayon du support doit alors répondre au compromis efficacité-performance des descripteurs. D'après les courbes obtenues, une proportion de 0.08 de la taille de la diagonale de l'objet parait convenable pour satisfaire ces deux contraintes.

4.2.1.3 Choix de l'angle du support

Une autre condition à vérifier pour choisir les points du voisinage du point référence est l'angle du support de ce point. Cet angle correspond à l'angle maximum formé par la normale du point référence et la normale d'un point du voisin. Cet angle est introduit pour minimiser l'impact des occultations et du bruit. Pour un point référence **A** de position et normale ($\mathbf{p_a}$, $\mathbf{n_a}$), un point **B** de position et normale ($\mathbf{p_b}$, $\mathbf{n_b}$) est dans le support de **A** si :

$$\text{Acos}\,(\mathbf{n_a}, \mathbf{n_b}) < \text{Angle_ support}$$

Le choix d'une faible valeur de cet angle peut ne pas être suffisant pour décrire la forme locale du voisinage du point référence (sauf dans le cas des zones planes). Néanmoins, un angle faible de support permet d'assurer une robustesse au bruit et aux occultations. En effet, en général, la direction d'une normale d'un point aberrant s'écarte largement de celle des normales des points de la surface locale du support. Par conséquent, un très grand angle de support (>150°), inclue un grand nombre de voisins qui peuvent correspondre à des points bruités et faussent les valeurs cumulées dans les histogrammes. D'un autres coté, plus l'angle est grand, plus la direction des normales voisines est orthogonale à la direction de la normale du point référence, ce qui permet de représenter une information plus riche sur la forme locale de la surface autour du point référence (cas de zone sous forme dôme ou selle). En outre, les angles faibles correspondent à des points situés à peu près sur la même zone plane que le point référence et ne représente pas forcement une information très descriptive de la forme locale.

Face à ce compromis, notre choix final de la valeur de l'angle se fait sur la valeur qui réalise la balance entre le pouvoir descriptif et la robustesse du descripteur. Pour les tests de ce chapitre nous avons choisi la valeur 90°.

4.2.2 Comparaison des descripteurs

Cette évaluation suit un principe similaire à celui utilisé pour les paramètres dans le paragraphe précédent. Pour ce faire, les descripteurs sont calculés sur le modèle d'origine et sur le modèle transformé (modèle test) pour les mêmes positions spatiales des PIs. Pour avoir une comparaison la plus juste possible entre les descripteurs et pour que la performance des descripteurs ne soit pas affectée par les erreurs apportées par les détecteurs, nous utilisons le même « détecteur » de PIs pour tous les algorithmes : nous extrayons aléatoirement un ensemble PIs sur chaque modèle (un nombre entre 50 et 150), puis nous extrayons les points correspondants physiquement sur le modèle transformé. La mesure de distance euclidienne est utilisée pour la phase de mise en correspondance entre chaque descripteur du modèle test et tous les descripteurs du modèle original. Pour décider de l'appariement, le ratio entre la distance du plus proche voisin et le deuxième plus proche voisin, est utilisé : si le ratio est inférieur à un seuil th_d, la correspondance est établie entre le descripteur test (du modèle transformé) et le descripteur du premier plus proche voisin du modèle d'origine. La correspondance des descripteurs est validée si les positions spatiales des deux points de calcul se correspondent.

Comme méthodologie d'évaluation des descripteurs 2D, la courbe ROC, représentant le taux des vrais positifs versus le taux des faux positifs, est généralement utilisée pour comparer les descripteurs. Néanmoins, l'ambiguïté dans le calcul du taux des faux positifs défavorise cette méthode. Nous adoptons alors dans notre évaluation l'approche proposée par Mikolajczyk et Schmid dans (Mikolajczyk, et al., 2005) et nous présentons le résultat en termes de courbes Rappel versus 1-Précision. Pour obtenir ces courbes, nous varions la valeur du seuil th_d. Nous commençons par l'évaluation de l'invariance des descripteurs pour des transformations géométriques rigides (rotation et translation).

4.2.2.1 Invariance aux rotations

❖ **Description**

Parmi les objectifs de notre travail, la caractérisation de la forme de l'objet d'une façon invariante à un changement d'angles de vue est très importante. En effet, en test, un même objet doit être reconnu pour une prise de vue différente de celles stockées dans la base de données. L'étape de description doit satisfaire la propriété d'invariance aux transformations géométriques.
Un changement de l'angle de vue peut se présenter soit par une rotation du capteur autour de l'objet soit par la rotation de l'objet autour d'un capteur fixe. Dans notre expérimentation, nous nous situons dans le deuxième cas. Dans la base Minolta, nous sélectionnons 5 vues, avec un écart d'angle de 20°, pour chacun des 9 objets. L'idée, à ce niveau, est de considérer pour chaque objet 4 transformations de rotation autour de l'axe y entre les vues adjacentes (exemple une transformation de la vue 0° vers la vue 20°). Le processus de calcul des descripteurs et de mise en correspondance, décrit plus haut, est appliqué pour l'ensemble des vues des 9 objets.

Une illustration du résultat du processus de mise en correspondance des descripteurs IndSHOT, sur un modèle d'Ange, entre la vue initiale située à 100° et la vue située à 120°, est donnée dans la Figure 4-11. Une évaluation quantitative est illustrée, sous forme de courbes: rappel vs 1-précision, dans la Figure 4-12. Nous illustrons, dans la Figure 4-13, les images des histogrammes des six descripteurs calculés au même point physique (le point n°1 sur les deux vues de la Figure 4-11) entre deux vues.

❖ **Résultats**

- La Figure 4-11 montre que le descripteur IndSHOT a correctement apparié 34 points parmi 65 points mis en correspondance sur les 71 points de calcul sur les deux vues, ce qui donne un rappel de 47% et une précision de 52%. Un résultat qui annonce une forte stabilité de ce descripteur pour une variation d'angles de vue.

- Sur les histogrammes de la Figure 4-13, nous remarquons la forte similarité entre les deux images des histogrammes des descripteurs IndSHOT et IndThrift. D'ailleurs, lors du processus de mise en correspondance, ceux deux points sont correctement appariés entre eux pour ces deux descripteurs. Bien que les deux histogrammes du Thrift se ressemblent, la remarque est que ce descripteur a la même allure quasiment sur tous les points de la surface de l'objet, ce qui augmente le nombre faux homologues. En outre, les images du Spin et du LSP, représentent des matrices avec plusieurs cases vides qui n'ont pas la même position sur les deux images des deux vues. Or, la mesure de similarité utilisée dans notre cas est la distance euclidienne qui compare les histogrammes bin à bin. Il en résulte que la distance de mise en correspondance est grande pour ces descripteurs et les points ne sont pas appariés.

- D'une manière générale, en augmentant le seuil de distance pour la mise en correspondance des descripteurs, le nombre total des correspondances augmente, et nous nous attendons à ce que la précision (TP/ TP+FP) diminue et le rappel (TP/ TP+FN) augmente.

    - Les résultats de la Figure 4-12 montrent que le IndSHOT est de très loin plus robuste à une rotation des données initiales. En effet, la courbe des valeurs du rappel vs la précision de ce descripteur est nettement au-dessus des autres courbes. Ce descripteur a amélioré la stabilité du descripteur d'origine SHOT.
    - Le pouvoir d'invariance porté par le IndThrift est inférieur aux deux précédents mais finit par rattraper le SHOT à partir d'une certaine valeur du seuil du ratio (th_d >0.65) des distances.

- Cette invariance aux rotations est due essentiellement à l'utilisation des bases locales (RF) dans la formulation de ces descripteurs. La description de la variation locale de la forme autour du PI relativement à une RF (SHOT et IndSHOT) est invariante par rotation. De plus, une normale calculée en un point est la même quelques soit l'angle de la prise de vue. Par ailleurs, une similitude conserve les angles géométriques (les mesures d'angles non orientés). En outre, l'indice de forme est constitué des courbures principales qui correspondent à un calcul différentiel de second ordre de la surface (présenté dans le chapitre 1) et sont invariant aux orientations. Avec ces caractéristiques, les descripteurs SHOT, IndThrift et IndSHOT cumulant les cosinus des normales et/ou les valeurs de l'indice de forme, sont discriminants et invariants aux rotations.

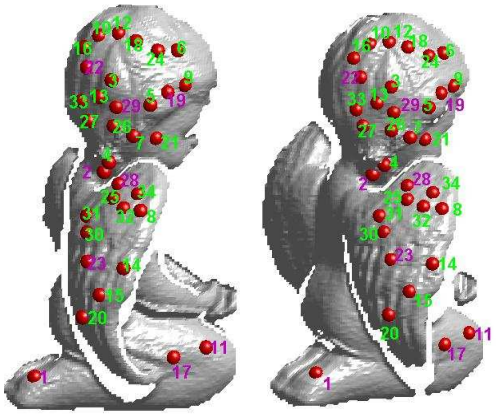


Figure 4-11 Résultat de la mise en correspondance après rotation des PIs entre la vue initiale à 100° et la vue à 120° des descripteurs IndSHOT

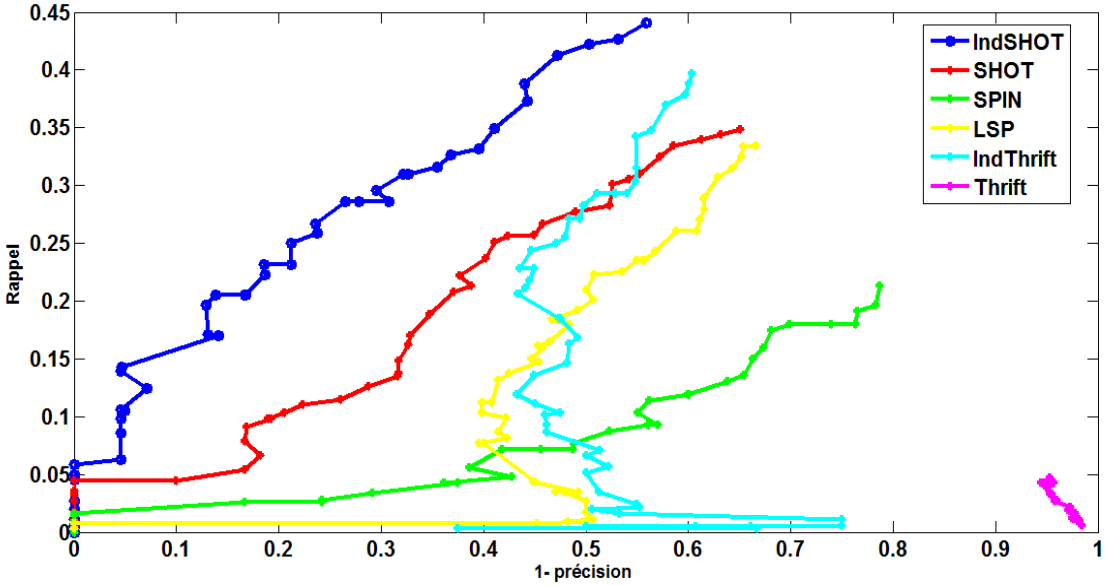


Figure 4-12 Courbes Rappel vs 1-Précision pour un changement d'angles entre les modèles Minolta

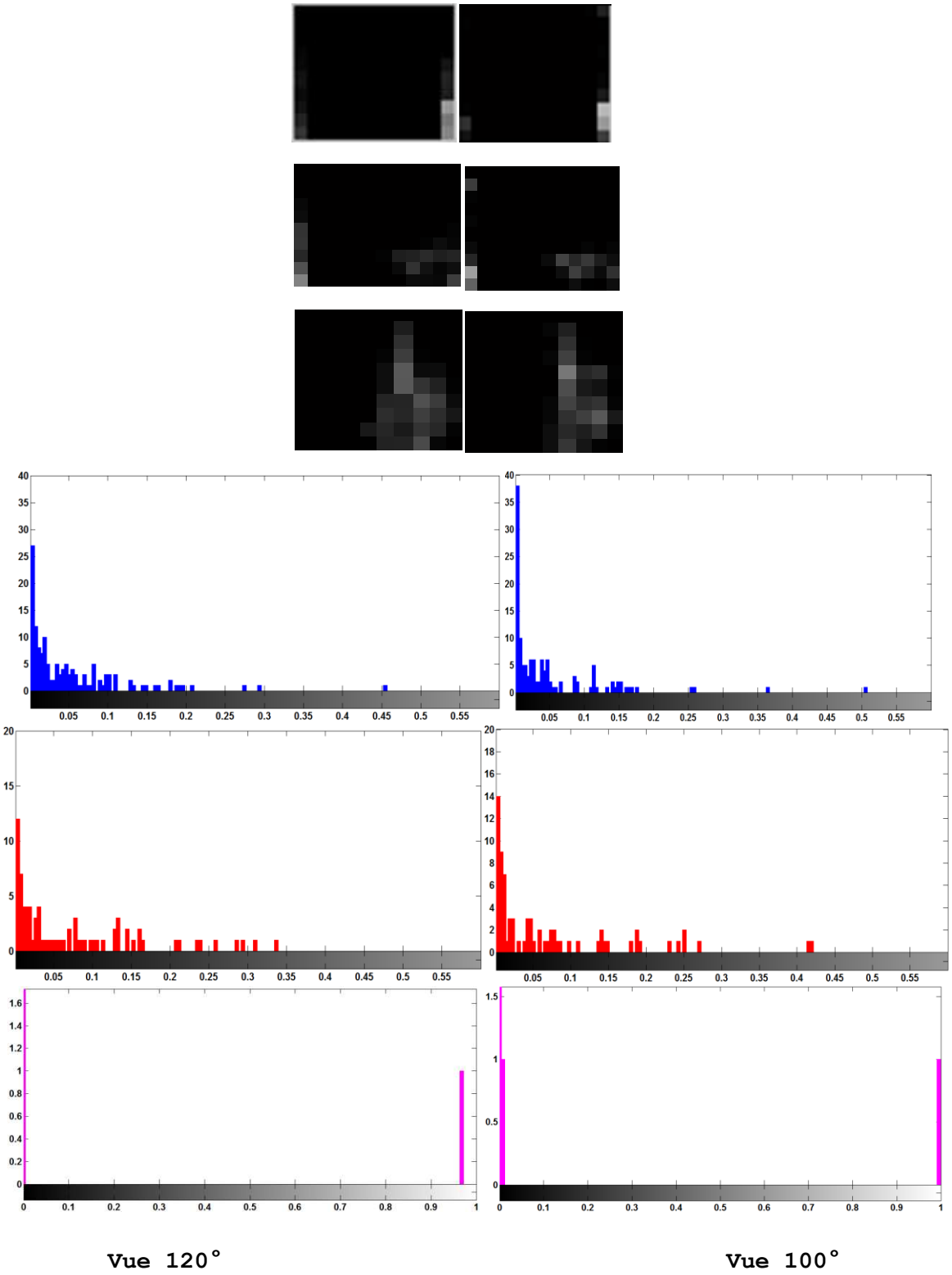


Figure 4-13 Représentations, au même point physique situés sur deux vues (100° et 120°), des descripteurs: IndThrift, LSP, Spin, IndSHOT, SHOT et Thrift, dans l'ordre de haut vers le bas

- Le LSP a un comportement assez proche de celui du IndThrift, puisqu'ils encodent la forme locale avec la même information (normales et indice de forme).
- Le Spin, en dépit de sa propriété d'invariance aux rotations, montre une très faible stabilité. Rappelons que pour un point orienté sur une surface 3D, le Spin image est

décrite par deux paramètres: distance du plan tangent du point orienté à ses voisins et la distance du point orienté à son vecteur normale. Sachant qu'une rotation est une similitude, les distances sont conservées. Les deux mesures de distance du Spin sont alors invariantes par rotation. Or, dans notre évaluation, nous considérons un changement d'angles dans des conditions d'acquisition réelle. Ce qui implique que pour une même région de la surface vue sous deux angles différents, la densité des points et l'amplitude du bruit varient. La sensibilité du Spin à ces variations justifie le résultat obtenu.

- Le Thrift a la plus faible performance. En effet, bien que les mesures cumulées correspondent à des mesures de cosinus de normales, la première lacune de ce descripteur est l'absence d'une RF ou d'une RA. De plus, les cosinus correspondent aux cosinus de l'angle entre deux normales du même point calculées sur deux supports de tailles différentes. Ces angles sont très faibles et varient sur un petit intervalle de l'histogramme (le cosinus est entre 0.8 et 1) et ne renseignent pas suffisamment sur la forme locale. Avec la formulation du Thrift, plusieurs descripteurs ont la même allure et se confondent lors de la mise en correspondance, ce qui baisse le taux de rappel. Par ailleurs, l'ajout de l'indice de forme à ce descripteur, avec la définition du IndThrift, a nettement haussé sa performance.

La conclusion pour cette évaluation est l'importance de l'utilisation d'une base de référence pour la description locale du support. Le IndSHOT l'emporte pour l'invariance à la rotation sur les autres descripteurs. Pour notre évaluation le changement d'angles se fait sur des données réelles. Ce qui nous rapproche d'un scénario de mise en correspondance de vues partielles d'objets pour une tache de reconnaissance. D'après ces résultats, nous estimons que ce descripteur réussira à correctement apparier les différentes vues des objets.

4.2.2.2 Invariance à l'échelle et à la densité de points

❖ **Description**

Dans ce paragraphe, nous traitons la problématique de l'invariance des descripteurs pour un changement d'échelle et pour une variation de la densité du point du maillage. Géométriquement, une transformation de translation ou un changement d'échelle correspond à une homothétie dans l'espace.

Dans un premier temps, nous considérons une décimation du maillage qui consiste à réduire le nombre de facettes et de points dans le maillage initial de chaque objet. Dans la Figure 4-14, Nous traçons les courbes rappel vs précision des 6 descripteurs, correspondant au processus de décimation du maillage initial avec un facteur de 20% de réduction sur les 9 modèles de Minolta.

Dans un deuxième temps, nous considérons la base des 7 objets Minolta présentant une variation d'échelles. Dans notre évaluation, nous avons choisi deux échelles pour chaque objet avec 11 vues sur chaque échelle. Le meme processus de génération des courbes rappel-précision est appliqué pour la transformation de changement d'échelle. Ces courbes sont données pour seulement les descripteurs SHOT, IndSHOT et IndThrift dans la Figure 4-15.

❖ **Résultats**

Variation de la densité des points :

- Après décimation, le maillage obtenu a une valeur de résolution (moyenne des distances entre les sommets du maillage) plus grand et les détails de la forme sont moins visibles. Par exemple, si nous prenons un maillage de 100 facettes, avec une résolution de 0.63 et une décimation de facteur 20%, le maillage transformé aura 80 facettes et une résolution du maillage de 0.72. En augmentant le niveau de décimation, le nombre de points apparié diminue. Nous remarquons nettement qu'avec la décimation, la performance des descripteurs a chuté.
- En variation de densité de points, le IndSHOT a une meilleure performance pour les premières valeurs du seuil sur le ratio des distances des deux premiers plus proche voisins. Les descripteurs LSP et IndThrift ont à peu près le meme comportement. A partir du seuil (0.6), la performance de ces deux derniers devient meilleure que celle du IndSHOT.

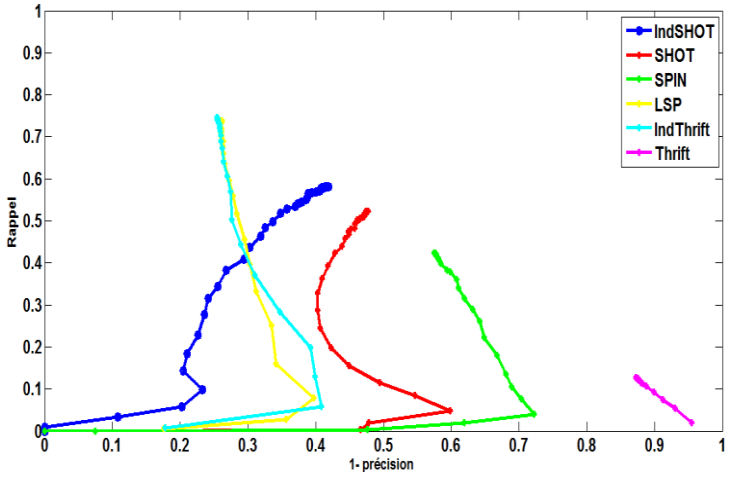


Figure 4-14 Courbe rappel-précision pour une décimation de 20% du maillage initial

Le descripteur Thrift a une très faible invariance à la densité des points avec un rappel <0.2. Les mêmes explications données pour l'invariance à la rotation sont valides ici.

La performance du Spin chute considérablement pour une densité plus faible du nuage de point. Pareillemnt que pour le LSP, la base de référence (RA) utilisée pour le calcul des coordonnées cylindrique dépend de la méthode de calcul de la normale en un sommet. Cependant, et sachant qu'une homothétie de rapport k modifie les distances par un facteur |k|, les spins images du meme point sur deux surfaces échantillonées uniformement avec deux échelles différentes, sont linairement corrélés.Une méthode standard pour la mise en correspondance de données corrélées linéairement est le coéfficient de corrélation linéaire. L'utilisation de cette mesure de similarité, à la place de la distance euclidienne, augmenterai la performance de l'appariement pour le descripteur Spin. Cependant, pour une comparaison plus égale entre les descipteurs, nous conservons la meme mesure de similraitéentre les différents descripteurs.

- Il est à noter que ces descripteurs sont sensibles aux perturbations dans le calcul des normales qui est la moyenne des normales des facettes directement adjacentes au point. Une normale

d'une facette est calculée par une méthode basique de produit vectoriel entre les vecteurs des sommets de la facette. Or, en variant la résolution du maillage par une décimation, le même point n'aura plus les mêmes voisins sur les deux maillages et seront plus éloigné spatialement dans le maillage décimé. La normale résultante est, de ce fait, plus au moins différente de la normale initiale.

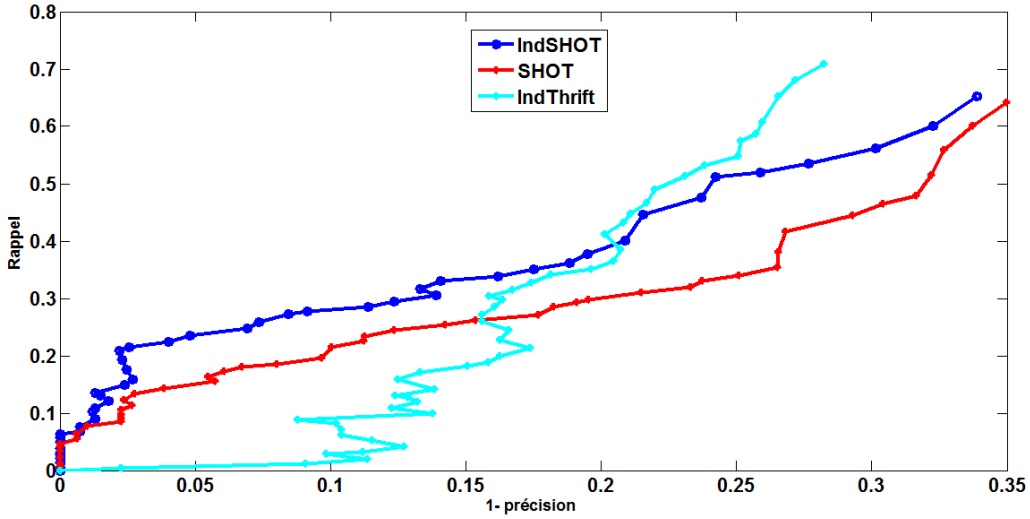


Figure 4-15 Courbe rappel-précision pour un changement réel d'échelle pour les descripteurs SHOT, IndSHOT et IndThrift

Variation d'échelles :

Nous mettons l'accent sur le fait que l'évaluation de la variation d'échelle s'est faite sur des données réelles en prenant des acquisition des objets à différentes distances. En déplaçant l'objet, sa posture et sa pose (par rapport à l'angle de vue) n'est pas exactement la meme. Un décalage dans les positions spatiales de calcul des descripteurs peut fausser notre processus de mise en correspondance. Pour ce fait, nous avons introduit une marge dans la distance seuil (distance moyenne) servant à valider l'appariement des points physiquement.

Nous rappelons que la normalisation dans le formalisme des descripteurs permet d'assurer l'invariance à la densité des points et à l'échelle. Pour une variation d'échelle, d'après les courbes de la Figure 4-15, les descripteurs IndSHOT et SHOTdonnent un meilleur taux de précision-rappel pour les faibles valeurs du ratio de comparaison des distances. Nous savons que les angles sont conservées par une homothétie, et que le SHOT utilise une RF invariante pour l'expression des cosinus des angles, la stabilité des descripteurs SHOT et IndSHOT s'explique alors. D'autre part, la caractéristique d'invariance à l'échelle de l'indice de forme consolide la robustesse du descripteur IndSHOT et IndThrift. Pour les grandes valeurs du seuil (>0.75), le IndThrift devient meilleure que le IndSHOT. Néanmoins, pour les grandes valeurs de précision, qui caractérise typiquement les applications temps réel, IndSHOT a un meilleur degré de Rappel.

En variation d'échelle, un rappel >60% n'est atteint que vers les dernières valeurs du seuil.

- Comme évoqué dans la partie évaluation du chapitre Détecteurs, la décimation des données peut étre assimilée à une variation artificielle d'échelle. Avec les résultats présentées en décimation et en changement d'échelles, le IndSHOT encore une fois a une meilleure performance que les autres descripteurs.
- L'invariance à un changement d'échelles semble être le plus grand défi pour les méthodes de description proposées si le support est fixe (mono-échelle). Dans notre choix de la taille du support, nous avons pris un rayon proportionnel à la diagonale de la taille de l'objet. Pour deux échelles différentes, la taille de l'objet ne varie pas et la zone locale prise en compte dans la description est la même. Si le descripteur est suffisamment discriminant et stable, son invariance à une variation de la densité des points est réalisable pour ce choix.
- Nous prévoyons que l'utilisation d'une échelle appropriée (entre autre le résultat de la version multi-échelle) pour la description de chaque point améliorera cette performance.

4.2.2.3 Robustesse au bruit

❖ **Description**

Pour mesurer l'impact du bruit sur l'étape de description et de mise en correspondance, nous ajoutons un bruit Gaussien sur les données initiales. Nous menons deux expérimentations avec deux valeurs de variance gaussienne σ_1 et σ_2 qui correspondent aux pourcentages respectifs: 10% et 30% de la résolution du maillage de chaque modèle.

Les courbes rappel-précision pour ces deux perturbations du maillage initiales des 9 modèles Minolta (5 vues par objets) sont données dans la Figure 4-16.

❖ **Résultats**

- Sur les deux perturbations de données, le descripteur IndSHOT est le moins sensible au bruit avec des taux de rappel-précision plus élevés que les autres descripteurs. Cette robustesse au bruit est due essentiellement à la robustesse de la RF et au processus de normalisation du descripteur. Cette robustesse est confirmée par l'ajout de l'information de l'indice de forme puisque une meilleure performance, que celle du SHOT, est réalisée avec cette nouvelle combinaison.
- Le Spin parait très sensible au bruit dès la faible perturbation de variance (σ_1= 0.1xMR). En augmentant l'écart type du bruit gaussien, la précision et le rappel décroit et tend vers 0. Ceci est du au fait que le descripteur est très sensible aux petites variations dans l'estimation de la normale. Ce descripteur est basé sur le cumul des coordonnées cylindrique (radiale et élévation) des points du support, exprimés par rapport à la normale du PI (RA). Or, pour une perturbation des coordonnées des points du maillage, le calcul de la normale référence qui se fait d'une façon non robuste au bruit (par une simple approximation de moindre carrée) est perturbé. Ce résultat est cohérant avec la conclusion donnée par (Tombari, et al., 2010).
- La performance du LSP décroit rapidement en augmentant le degré du bruit. La défaillance de la robustesse du LSP est, similairement au Spin, liée à la non robustesse de l'estimation de la RA et des quantités différentielles (normales) au bruit.

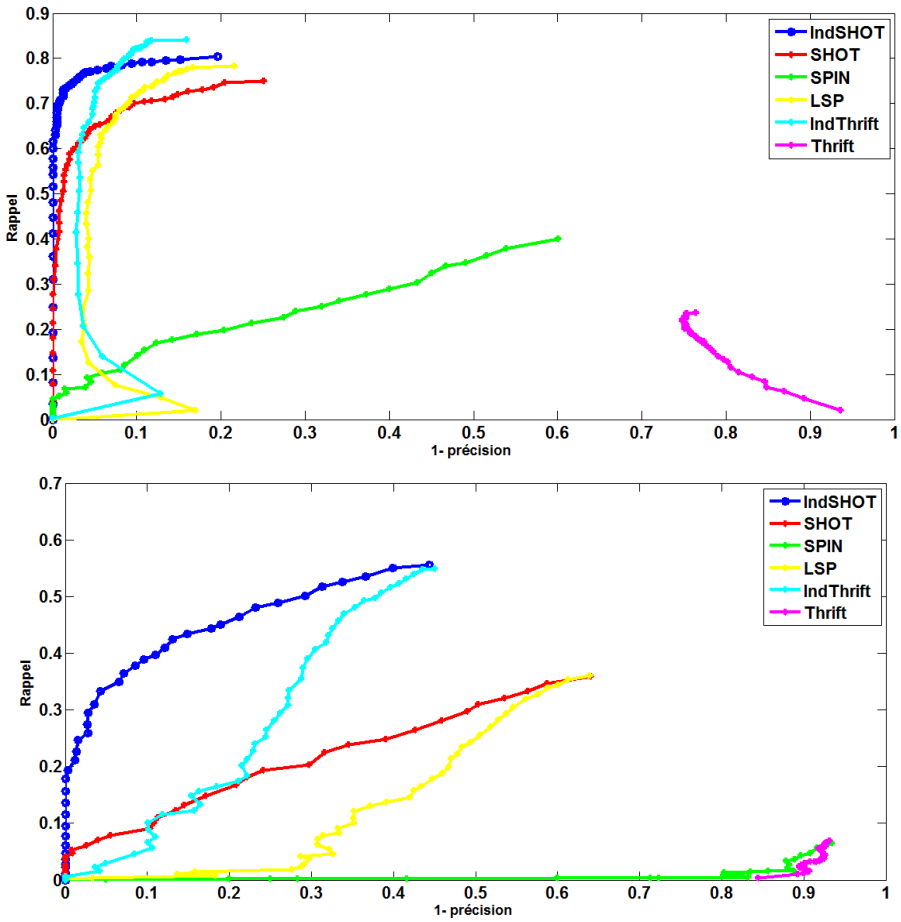


Figure 4-16 Courbe rappel-précision pour un bruit, de σ = 0.1xMR -en haut, et de σ = 0.3xMR -en bas

- Le Thrift, basé uniquement sur les normales avec l'absence de notion de base de référence, reste le plus vulnérable à une perturbation des données. Par ailleurs, le IndThrift se situe en deuxième position concernant la robustesse au bruit. Nous consluons que la forme 2D de ce nouveau descripteur joignant le formalisme du Thrift avec l'indice de forme a permis de créer une description plus stable et plus discriminante.

Ces résultats confirment et consolident les constatations données précédemment.

4.2.2.4 Robustesse à un lissage

❖ **Description**

La modification du maillage par un lissage des données est un autre type de perturbations des données initiales que nous avons étudié dans notre évaluation de la robustesse des descripteurs. Cette modification lisse les détails de la forme et rend difficile la tâche d'extraction d'informations discriminantes, comme l'illustre la Figure 4-17. Nous étudions l'impact d'une transformation de lissage Laplacien des données, avec un facteur de relaxation f de valeur proportionnelle à la résolution du maillage (0.25*MR), sur la performance des descripteurs. Similairement à ce qui a été présenté pour le bruit, nous traçons la courbe Rappel-Précision dans la Figure 4-17 pour une application de cette transformation sur 5 vues de chacun des 9 objets Minolta.

❖ **Résultats**

Le résultat montre clairement la forte stabilité du descripteur IndSHOT par rapport aux autres descripteurs (meilleure rappel et précision). Les courbes LSP, IndThrift et SHOT ont un comportement assez proche les unes des autres. Pareillement au cas d'ajout d'un bruit, le comportement du Spin est clairement sensible au lissage. Les mêmes raisons avancées pour la robustesse au bruit sont valables pour le lissage. La formulation donnée pour ces histogrammes parait sensibles à l'échantillonnage.

Les résultats présentés dans cette partie confirment ceux obtenus pour le bruit en invoquant la stabilité du descripteur IndSHOT grâce à son design de normalisation et son hybridité histogramme-signature pour décrire l'information locale.

- Le pouvoir descriptif et discriminant intervient pour accentuer l'invariance ou la robustesse des descripteurs. Une quantification de ce pouvoir descriptif est présentée dans la partie qui suit.

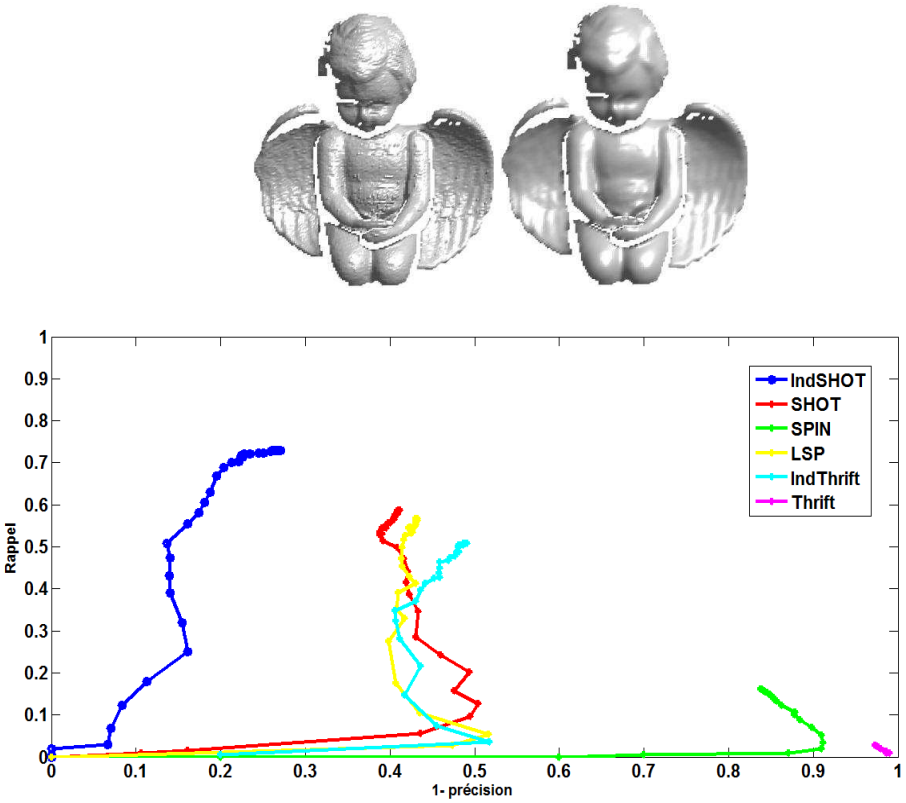


Figure 4-17 En haut-Exemple de lissage du modèle de l'ange -En bas- Courbe rappel précision pour un facteur de relaxation f = 0.25xMR

4.2.2.5 Pouvoir descriptif et stabilité pour un décalage de la position physique

❖ **Pouvoir descriptif**

○ **Quotient R_d** :

Choksuriwong et al. (Choksuriwong, et al., 2005) se sont intéressés à l'évaluation de la performance de différents descripteurs invariants par transformations géométriques et ont proposé, pour quantifier le pouvoir descriptif et discriminant des descripteurs, un ratio de l'écart type des moyennes des valeurs

du descripteur sur les objets de la base sur la moyenne des écarts type du descripteur. Nous notons ce quotient R_d.

En effet, dans le but de reconnaitre un objet qui apparait sous différentes orientations ou échelles, le descripteur utilisé doit donner des valeurs similaires sous ces vues. Soit E_i le vecteur formé par les valeurs de l'écart type du descripteur j obtenu pour l'objet i. Afin d'avoir des descripteurs similaires pour le même objet, la valeur moyenne des E_i doit être minimisée.

D'un autre coté, pour pouvoir discriminer les objets entre eux, la dissimilarité entre les descripteurs doit être grande pour bien séparer ces objets. Soit M_j le vecteur formé par les valeurs de la moyenne du descripteur j obtenu pour un objet i. Pour augmenter le pouvoir discriminant du descripteur entre les objets, l'écart type σ_{M_j} des M_j, doit être maximisé. Par conséquent, le quotient $R_d = \sigma_{M_j} / \overline{E_j}$ présente un critère qui caractérise la performance d'un descripteur.

Dans notre cas, les descripteurs sont sous forme d'histogrammes calculés sur des PIs. Nous proposons de calculer ce ratio de la façon suivante:

- Pour un objet i, et un point k sélectionné sur une vue x, nous cherchons les homologues de k sur les autres n vues de l'objet. Nous avons ainsi n descripteurs décrivant le même point physique k. Pour ces n histogrammes, et pour chaque bin (composante) de l'histogramme, nous calculons la moyenne M_i sur les valeurs des vues de l'objet pour obtenir un vecteur des moyennes des composantes. Puis, nous calculons, pour chaque case et pour toutes n vues, l'écart type par rapport à ce vecteur des moyennes. Nous formons l'écart type total e_i , pour l'objet i, comme la somme de ces écarts types.
- Pour chaque type de descripteur j, répéter ce traitement en prenant deux points par vue et faire la moyenne des écarts E_{ji} par objet. Ensuite, la moyenne sur tous les objets $\overline{E_j}$ est déduite.
- Nous disposons pour chaque objet i d'un vecteur M_i des moyennes des composantes du descripteur calculé au même point physique sur ses n vues. Nous calculons, par la suite, pour chaque vecteur M_i, son écart σ_i avec l'ensemble des vecteurs M_m ($m \neq i$). Le nominateur du quotient σ_{M_j} correspond à la somme des écarts types σ_i sur tous les objets. Pour chacun des six descripteur, nous déduisons la valeur du ratio $R_d = \sigma_{M_j} / \overline{E_j}$. Le dispositif de ces étapes est schématisé dans la Figure 4-18.

- **Quotient C_{dis}**:

Nous introduisons un autre ratio, que nous appelons $\mathbf{C_{dis}}$. Ce ratio calcule pour k points extraits, aux mêmes positions physiques sur une vue initiale et sur sa vue transformée par rotation, le rapport entre la moyenne des distances des descripteurs faussement appariés, et la moyenne des distances des descripteurs correctement appariés. Plus ce ratio est grand plus le pouvoir descriptif est grand.

Les résultats sont donnés dans le Tableau 4-2. D'après ces ratios, il est clair que le IndSHOT a le plus grand pouvoir descriptif. Le SHOT et le IndThrift se positionne à peu près au même niveau. Le ratio R_d du Thrift est élevé, cependant la faible valeur de $\mathbf{C_{di}}$ fait chuter son pouvoir descriptif.

Classés dans un ordre décroissant de performance, nous avons l'IndSHOT, le SHOT, le IndThrift. Ces résultats rejoignent les constatations données dans les paragraphes précédents pour conclure clairement sur la supériorité du descripteur IndSHOT.

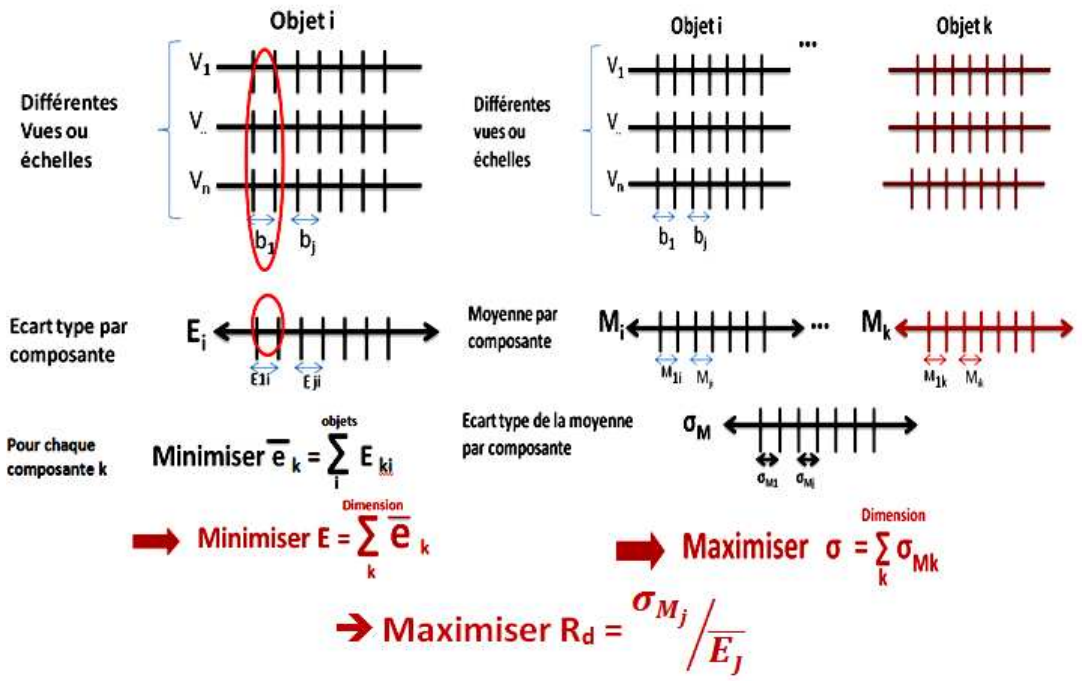


Figure 4-18 Schéma de calcul du quotient R_d utilisé pour l'évaluation du pouvoir descriptif et discriminant d'un descripteur

	SHOT	IndSHOT	Spin	LSP	IndThrift	Thrift
Ratio $\mathbf{C_{dis}}$	1.07	**1.47**	0.79	0.99	1.07	0.25
Ratio $\mathbf{R_d}$	3.5	**4.8**	2.6	2.5	3.09	4.5

Tableau 4-2 Pouvoir descriptif des six descripteurs avec le ratio C_{dis} et le ratio R_d

❖ **Stabilité pour un décalage de la position physique du point de calcul**

L'étape de description des régions locale autour des PIs intervient après la phase de détection. La répétabilité de ces détecteurs est primordiale pour réussir la description et le processus de mise en correspondance. Néanmoins, cette répétabilité reste approximative puisque nous savons, qu'une erreur de décalage dans les positions spatiales des PIs extraits par les détecteurs peut subvenir d'une vue à une autre. Nous proposons dans ce paragraphe, d'étudier le comportement des différents descripteurs pour décrire la même zone décalée avec une marge ε croissante. Pour ce faire, nous sélectionnons un point sur une vue de l'objet d'Ange. Puis, nous cherchons les k plus proche voisins de ce point que nous ordonnons dans un ordre croissant selon leur distance du point référence. Nous calculons pour chacun de ces points les six descripteurs sur la zone qui les entoure avec le même rayon du support.

Nous illustrons dans la Figure 4-19, la distance entre le descripteur du point référence de calcul et les descripteurs de ses k (~ 200) plus proche voisins. L'axe des x représente la distance physique du point référence au $k^{ème}$ plus proche voisin et qui correspond au décalage.

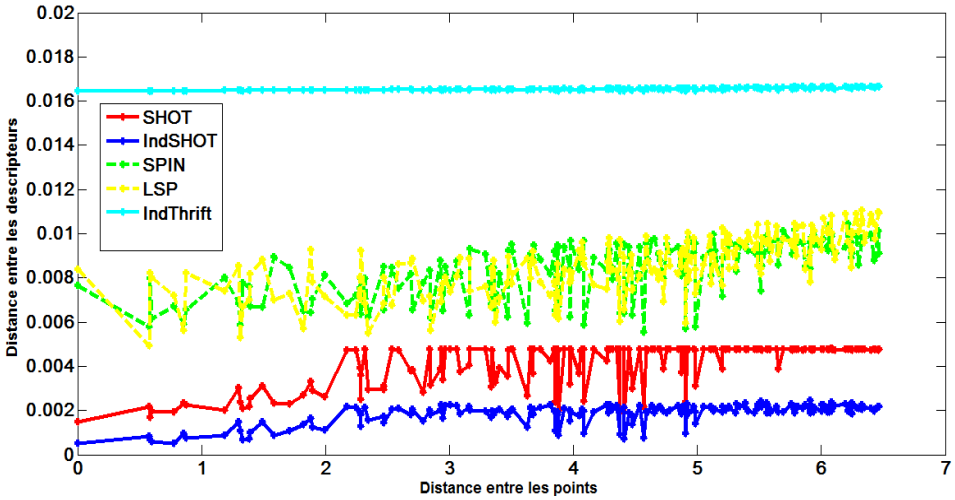


Figure 4-19 Evolution de la distance entre le descripteur d'un point référence et les descripteurs de ses k plus proches voisins (en fonction de la distance physique du $k^{ème}$ plus proche voisin)

Le IndThrift parait le plus stable des descripteurs pour un décalage de la position du calcul. La courbe des distances pour le IndSHOT évolue avec une faible pente et reste relativement stable à partir du 20ème plus proche voisin. La pente et l'amplitude de la courbe du SHOT est légèrement plus grande que celle du IndSHOT. Quant aux courbes du spin et du LSP, elles oscillent avec une amplitude plus importante exprimant une faible stabilité de ces deux derniers. D'un autre coté, la distance calculée entre les descripteurs est largement plus faible pour le IndSHOT.

- Notre conclusion sur cette partie, est que les nouveaux descripteurs proposés, IndSHOT et IndThrift, ont montré une performance qui a amélioré le pouvoir descriptif et la stabilité des descripteurs de l'état de l'art desquels nous nous sommes inspirés.

4.2.2.6 Temps de calcul et outils de développement

Dans notre travail, les algorithmes de description Spin, Thrift, IndThrift et LSP ont été implémentés. Pour le SHOT, l'implémentation originale est utilisée. L'algorithme original du CSHOT a été modifié pour l'adapter au descripteur IndSHOT. Le temps de d'exécution présenté dans cette partie n'inclue pas le temps de calcul des quantités de l'indice de forme qui est donné en sortie du détecteur.

Pour automatiser les calculs, nous avons créé une partie commune pour tous les descripteurs qui lit les données, construit le maillage, extrait les points d'intérêt (PIs) et lance la description. Le code implémenté dans le SHOT introduit un parallélisme dans le traitement des descripteurs en lancant des taches parallèles qui traitent la description des PIs par tranche de taille (nombre des PIs/nombre de tache). Nous utilisons ce même paraléllisme pour tous les algorithmes de descriptions. Les résultats du temps de calcul sont donnés dans le Tableau 4-3. Pour avoir une idée sur la totalité du temps nécessaire pour notre système et sa comparaison avec une approche existante, nous présentons, dans le

Tableau 4-4, le temps de calcul des phases de détection et de description pour le couple (SC_HK_C, IndSHOT) et pour l'approche du SURF.

La bibliothèque graphique VTK est utilisée pour les traitements appliqués au maillage. Les filtres suivants ont été utilisé pour la partie précédante de l'évaluation des descripteurs :

vtkPolyDataNormal est un filtre qui calcule les vecteurs normaux pour un maillage polygonal. Les normales correspondent aux normales des facettes, et la normale d'un sommet est déduite de la moyenne des normales des facettes qui lui sont adjacentes.

vtkSmoothPolydataFilter est un filtre qui lisse un objet de type vtkPolyData en utilisant une méthode de type Laplacienne. Les coordonnées des sommets du maillage sont remplacées par la moyenne des coordonnées des sommets qui lui sont connexes (qui lui sont directement liés par une arête). Un facteur de relaxation contrôle le degré des remplacements des sommets. Ce processus est répété selon le nombre d'itérations donné en paramètre.

vtkDecimatePro est un filtre de décimation qui réduit le nombre de polygones d'un maillage en approximant au mieux la géométrie de la forme originale. L'algorithme procède de la manière suivante : chaque sommet du maillage est inséré dans une file de priorité. Certains points qui ne peuvent pas être supprimés ou triangulés ne sont pas traités. Après chaque suppression d'un point de la file, lui succède une triangulation des trous. Le processus se repètent jusqu'à ce que la file soit vide. Le maillage avec les points qui restent est divisé en des pièces séparées selon une liaison non-manifold et les points sont réinsérés dans la file à nouveau. Puis, le même traitement est répété récursivement jusqu'à atteindre le pourcentage de réduction donné en paramètre.

❖ **Résultats**

Base: Nombre de PIs	**IndSHOT**	**SHOT**	**IndThrift**	**LSP**	**Spin**
Stuttgart : 75 points	0.56	0.57	2.9	0.54	0.52
Carotte : 72 points	0.42	0.41	2.15	0.42	0.42

Tableau 4-3 Temps de calcul (en seconde) des cinq descripteurs

Base : densité de points	Détection+description	
	Surf	SC_HK_C + IndSHOT
Stuttgart : 10948 points	**15.20**	**18.64**
Carotte : 4685 points	**10.13**	**9 .62**

Tableau 4-4 Temps de calcul (en seconde) des phases de détection et de description pour le couple (SC_HK_C, IndSHOT) et pour l'approche du SURF

D'après ces tableaux :

- Le temps d'exécution des descripteurs IndSHOT, SHOT, LSP et Spin est de même ordre.
- Le temps de calcul du descripteur IndThrift est important (allant jusqu'à 7 fois le temps du IndSHOT). L'explication est que ce descripteur se base sur le calcul des normales du voisinage sur deux rayons (le premier w_{Large} est égal au rayon du support et le deuxième w_{Small} est égal à la moitié du rayon du support). Les opérations de parcours de tout les points du support du point référence et le recalcule pour chacun de deux nouvelles normales résultantes des valeurs des normales des points des deux supports (W_{Large} et W_{Small}), sont couteuses.
- La légère supériorité dans le temps de calcul de la combinaison SC_HK_C -IndSHOT par rapport au détecteur-descripteur SURF, revient au temps mis par le détecteur pour calculer les quantités différentielles : courbure et normale. Cette étape reste à optimiser.

4.3. Conclusions

Nous avons implémenté et testé six descripteurs basés sur les histogrammes cumulant une information calculée essentiellement à base de géométrie différentielle. Le pouvoir descriptif, l'invariance aux transformations géométriques et la robustesse au bruit et à l'échantillonnage ont été évaluées et les résultats ont montré :

- Stabilité des descripteurs IndSHOT et IndThrift pour une variation d'angles de vue
- Supériorité nette du descripteur IndSHOT pour variation d'échelles
- Robustesse au bruit des deux descripteurs IndSHOT et IndThrift
- Forte stabilité du IndSHOT pour une transformation de lissage des données
- Pouvoir descriptif élevé du IndSHOT et une forte stabilité du IndThrift suite à un décalage de la position physque
- la supériorité du nouveau descripteur IndThrift par rapport à la version du Thrift. La forme 2D de ce nouveau descripteur en ajoutant l'information de l'indice de forme a haussé le pouvoir descriptif et a renforcé la stabilité et la robustesse du descripteur.
- La supériorité du nouveau descripteur proposé IndSHOT au descripteur d'origine SHOT. En effet, une nette amélioration est réalisée par ce nouveau descripteur qui combine deux informations de la forme locale (indice de forme et cosinus des normales) par rapport à une seule information (cosinus des normales) contenue dans le SHOT.
- Nous avons pu constater l'importance de l'utilisation d'une base locale de référence dans l'expression des descripteurs à travers la comparaison effectuée avec le descripteur proposé IndThrift n'incluant pas de base de référence dans sa formulation. L'étude comparative a confirmé la performance supérieure du IndSHOT utilisant une RF.
- L'évaluation quantitative a montré que le descripteur LSP est moins robuste et moins descriptif en comparaison avec les descripteurs SHOT et IndSHOT et IndThrift. Bien que ce descripteur se base sur la définition d'une RA pour l'expression des cosinus des normales, cette RA n'est pas calculée d'une façon robuste. En effet, la normale référence du PI correspond à la normale du point obtenue par une simple moyenne sur les normales des sommets voisins à la facette. Alors que la RF du SHOT et IndSHOT est obtenue par le calcul des vecteurs propres d'une matrice de covariance pondérée par les distances des points appartenant au support. Ce processus a permis d'augmenter la robustesse et la répétabilité de ces deux derniers. D'un autre coté, l'utilisation des cosinus des normales calculées sur deux

supports de taille différentes (IndThrift) parait plus performant qu'une description par le cosinus entre la normale du point référence et les normales de ses voisins (LSP).

- Un temps de calcul quasi réel pour l'étape de description du IndSHOT contrairement à celui du IndThrift.

Dans le cadre pratique d'une application de reconnaissance d'objets, le chapitre suivant expose une comparaison de la performance des descripteurs présentés en association avec les détecteurs les plus stables et robustes du chapitre précédants. L'évaluation du taux de reconnaissance sur différentes bases de données synthétiques et réelles démontrera la supériorité de la combinaison des détecteurs et descripteurs proposés.

5. CHAPITRE : RECONNAISSANCE

A l'issue des étapes de détection et de description, nous disposons d'une représentation des vues des modèles par un ensemble de primitives constituées des coordonnées 3D de points saillants et d'un histogramme de mesures des propriétés de la forme sur le voisinage de ces points.
Suite à l'étude comparative des détecteurs, la méthode de classification des types de surfaces qui combine les deux espaces SC et HK et l'approche de l'état de l'art basée sur l'indice de forme SI ont montré des performances comparables.
Concernant les descripteurs, trois d'entre eux ont montré des efficacités semblables. Le premier est le descripteur connu, spin image, qui utilise la distribution spatiale des points de la surface. Les deuxièmes et troisièmes descripteurs sont le SHOT et le nouveau descripteur proposé IndSHOT, qui mixent une information sur la variation locale de la surface (différentielle des normales pour le SHOT, et des indices de formes pour l'IndSHOT) avec une information de la distribution spatiale des points du support.
Dans une tâche de reconnaissance d'objets à partir de vues partielles, nous proposons de combiner ces différentes approches et de mettre en correspondance les PIs et les descripteurs résultants.
Dans la première partie de ce chapitre, nous présentons des techniques issues de la recherche sur l'apprentissage automatique (« Machine Learning ») en vue de reconnaitre ou classer une requête inconnue par rapport aux données d'apprentissage.
Dans la deuxième partie, nous évaluons différents systèmes de reconnaissance, basés sur les détecteurs et descripteurs proposés dans notre étude, pour conclure sur la (ou les) combinaison(s) la plus performante en termes de taux de reconnaissance. La validation de ces approches est faite sur quatre bases de données présentant des caractéristiques différentes.

5.1. Techniques d'apprentissage statistique supervisé pour la classification

Un algorithme d'apprentissage de classification est dit supervisé si l'étiquette (la classe) associée à chaque observation est fournie en entrée. Le but de cet algorithme est de minimiser l'erreur de séparation (classification) sur la base d'apprentissage pour minimiser l'erreur en généralisation (problème d'optimisation).
Dans ce qui suit, nous présentons quelques techniques d'apprentissage utilisés dans la classification d'objets, à partir de notre étude bibliographique.

5.1.1 Classificateur par le plus proche voisin

La comparaison entre une requête (objet inconnu) et les modèles de la base d'apprentissage correspond à une recherche dans l'espace des vecteurs de primitives. En d'autres termes, les vecteurs de primitives sont considérés comme des points dans l'espace de recherche et la meilleure correspondance sera donnée par le voisin le plus proche en termes de distance. Le calcul des distances utilise en général les normes L_1 ou L_2. Ce classificateur est plus connu en anglais sous le nom K-Nearest Neighbor (K-NN). Les vecteurs de primitives sont stockés en mémoire sans leur apporter de

modification. Pour prédire la classe d'une requête, l'algorithme cherche les K plus proches voisins de chaque vecteur de la requête et fournit comme classe estimée celle la plus fréquente parmi ces K plus proches voisins. Deux paramètres sont utilisés : le nombre K et la fonction de similarité pour comparer les vecteurs de la requête à ceux en apprentissage.

5.1.1.1 Métriques d'appariement

L'extraction des primitives des objets en apprentissage se fait au préalable, souvent en hors-ligne. En phase de reconnaissance, une mise en correspondance des objets un à un, compare les vecteurs primitives extraits de l'objet test avec tous ceux de la base d'apprentissage. Pour chaque vecteur (descripteur), l'objet le plus similaire reçoit un vote. L'objet avec le plus grand nombre de votes est choisi comme objet correspondant.

Pour l'appariement des descripteurs définis sur les points d'intérêt, une mesure de similarité est calculée. Dans le cas des histogrammes, les mesures de similarité sont souvent les distances usuelles (Euclidienne (Bay, et al., 2006), Mahalanobis, Minkowski, Manhattan,…) ou encore la distance de programmation dynamique (DPD)). Ces métriques d'appariement comparent des bins de mêmes indices et sont par conséquent très sensibles à la quantification. Inversement, la mesure « Earth Mover Distance » EMD (distance de transport), parfois appelée distance du cantonnier, s'affranchit de cette limitation et compare des bins d'indices différents. Si f et g sont deux histogrammes de N bins, la distance EMD entre leurs histogrammes est égale à la distance L_1 entre leurs histogrammes cumulés ($||F - G||_1$). Dans le cas circulaire, la distance EMD entre f et g est le minimum en k des distances L_1 entre les histogrammes F_k et G_k cumulés circulairement à partir de la k-ième cellule de quantification.

- Tangelder, et al. (Tangelder, et al., 2003), utilisent, après extraction de la signature des modèles polyèdres dans leur approche de points saillants pondérés, la mesure Proportionnal Transportation Distance (PTD) inspirée de la distance (EMD) et qui satisfait l'inégalité des triangles. Cette mesure d'appariement permet de garantir l'invariance géométrique des descripteurs mis en correspondance.

- Pour la mise en correspondance, le critère le plus intuitif est le seuillage de la mesure de similarité. Le choix d'un seuil global pour la distance du plus proche descripteur est problématique et ne peut pas être généralisé pour toutes les données et tous les descripteurs.
- D. Lowe (Lowe, 2004) propose un critère consistant à comparer la distance au plus proche voisin avec la distance au deuxième plus proche voisin. Le plus proche voisin est apparié si le rapport de ces distances est inférieur à un seuil de « validité » de l'appariement. Cette approche exige que le plus proche voisin soit significativement plus proche que le plus proche appariement erroné. Des bons résultats attestent de l'efficacité de ce critère. Cependant, cela restreint la classification à l'appariement avec les deux plus proches voisins outre l'handicap du choix manuel du seuil.
- Pour résoudre ces difficultés, Rabin et al. (Rabin, et al., 2007) proposent l'utilisation de la méthode a contrario dont l'idée principale est de valider des groupes en rejetant une hypothèse d'indépendance des structures à classer. En effet, cette méthode est adaptée pour réaliser une validation des mises en correspondances de descripteurs présentant des distances très petites sous l'hypothèse de l'indépendance des distances entre les composantes des deux descripteurs. Par la suite, une probabilité de similarité entre deux descripteurs bornée par un seuil de détection est définie.
- Biasotti et al. (Biasotti, et al., 2006) , dans une étude comparative pour des méthodes de classification 3D, en prenant 4 descripteurs de forme différents, comparent la performance de

classification de 5 mesures de similarité: classifieur de distance minimum (Minimum Distance), Maximum distance classifier, Average distance classifier, Centroid distance classifier et Atipicity distance classifier. La conclusion était que la mesure de similarité du plus proche voisin donnait les meilleurs résultats.

- Hetzel et al. (Hetzel, et al., 2001) mentionnent trois mesures de similarité pour comparer deux distributions de probabilité (histogrammes): une version modifiée de la méthode χ2-test, une mesure d'intersection et la divergence de Kullback-Leibler.

$$X^2(Q,V) = \sum_i \frac{(q_i - v_i)^2}{q_i + v_i} \; ; \qquad \cap\,(Q,V) = \sum_i \min\,(q_i, v_i)\;;$$

$$KL(Q||V) = \sum_i (q_i - v_i) ln \frac{q_i}{v_i}$$

Avec le théorème Bayesien et en supposant l'équiprobabilité, ils calculent également :

$$p(o_n|\Lambda m_k) = \frac{\Pi_k p(m_k|o_n)}{\sum_i \Pi_k p(m_k|o_i)}$$

Comme cette mesure utilise une probabilité à postériori, elle réalise des meilleurs résultats de reconnaissance et donne une estimation de la confiance de la décision.

5.1.1.2 Kd-Tree

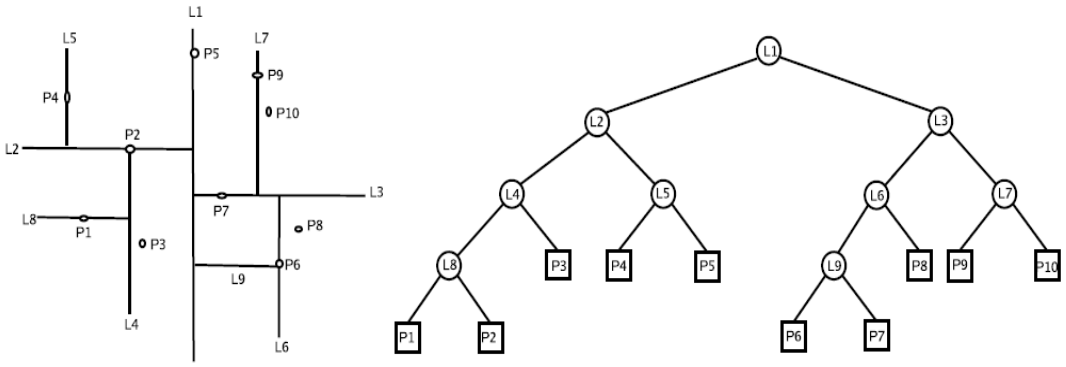
Dans le but d'organiser les données et d'optimiser le temps de recherche du plus proche voisin, on utilise souvent le Kd- tree (« K-dimensional tree »). Cette structure de données est un cas particulier des BSP trees (« Binary Space Partitionning trees », qui partitionnent l'espace des primitives sous forme d'arbre binaire. Chaque nœud de cet arbre contient un point de dimension K. Cette structure décompose l'espace en volumes englobants (ou voxels) et découpe chaque voxel (nœud non terminal) en deux sous-voxels (branches gauches et droites du nœud courant) grâce à un plan séparateur toujours perpendiculaire aux axes du repère de l'espace.

L'algorithme général de construction d'un Kd-Tree est donné par le Tableau 5-1. Une représentation de la division du plan pour la construction et la recherche d'une requête dans un Kd-Tree est donnée dans la
Figure 5-1 (Hemant, 2005). Si d est la dimension du descripteur et n le nombre d'entrée du KDtree, alors la complexité de construction de l'arbre est O(n log n) et celui de la recherche du plus proche voisin pour une requête est de O ($n^{1-1/d}$+k). L'étape de structuration des données est réalisée en hors ligne et est suivie de l'étape de recherche qui se fait en ligne.

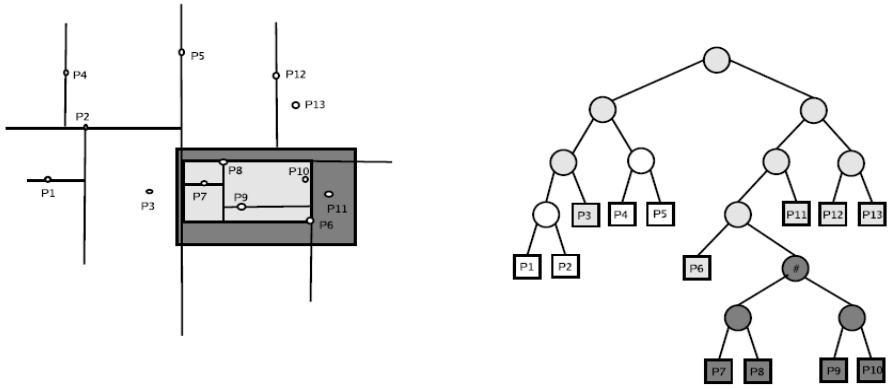
- Cette méthode d'appariement permet de trouver un compromis entre précision et rapidité.

Algorithme de construction d'un Kd-Tree
recConst (voxel V, objets O) 1- Tester s'il faut diviser le voxel courant V : - Si non : Rendre la Feuille(O) contenant la liste des objets O - Si oui : Continuer 2- Trouver le bon plan séparateur p 3- Couper le voxel V avec p pour obtenir VG et VD 4- Répartir les objets O dans VG et VD => OG et OD 5- Rendre le Nœud (p, recConst(VG ,OG), recConst(VD, OD))
Algorithme de parcours d'un Kd-Tree
recParcours(Voxel V , rayon r) 1 Tester si V est une feuille : Si oui : Rendre la plus proche intersection entre r et les objets de V Si non : Continuer 2 Trouver le fils « near » et le fils « far » 3 Tester quels sont les fils traversés par le rayon Cas le fils « near » ⇨ réitérer: recParcours(Vnear ,r) Cas le fils « far » ⇨ réitérer : recParcours(Vfar ,r) Cas les deux fils ⇨ réitérer: recParcours(Vnear ,r) Tester si une intersection a été trouvée dans le fils « near » Si oui Rendre cette intersection Si non réitérer : recParcours(Vfar ,r)

Tableau 5-1 Algorithme de construction et de parcours récursives d'un Kd-Tree (Fleury)



-a -



-b-

Figure 5-1 Représentation –a- de la division du plan pour la construction et –b- la recherche d'une requête dans un Kd-Tree (Hemant, 2005)

5.1.1.3 Hachage géométrique

Le hachage géométrique (Geometric hashing) permet de sélectionner un système de coordonnées orthogonales à partir des primitives, et exprimer le reste des primitives en fonction de la nouvelle base. Ces propriétés de position relative des primitives entre elles sont utilisées pour indexer une base de données et accélérer la phase de mise en correspondance. La description des relations géométriques entre les primitives extraites pallie aux problèmes d'occultation et d'invariance aux transformations. Le processus de reconnaissance est divisé en deux parties : prétraitement et reconnaissance. La complexité d'un algorithme sans indexation pour la mise en correspondance d'un nombre f de primitives d'un objet, est de O(log(f)). Pour comparer deux objets, la complexité est de O(f*log(f)). Si nous avons "m" objets en apprentissage et "n" objets requêtes, la complexité de tout le processus de mise en correspondance est O(m*n*f*log(f)), alors que pour un algorithme d'indexation, la complexité d'appariement est de O(m*n*f).

- **Repères de coordonnées centrées**

Dans la phase de prétraitement, Aykut Hozatlı (Hozatlı, 2009) et Ömer Eskizara (Eskizara, 2009) construisent, à partir des primitives, des repères de coordonnées centrés pour représenter l'objet par rapport à son axe intrinsèque, indépendamment de l'angle de vue. Ces repères peuvent être définis à partir de trois points non colinéaires. Si la base de données contient N points, 3 d'entre eux sont utilisées pour définir le repère, les N-3 points restants sont décrits par rapport à ce repère. Ce traitement est répété pour chaque triplet de points non colinéaires de la base. La nouvelle représentation des points est stockée dans une table de hachage en utilisant, par exemple, les coordonnées du point comme clé pour accéder au bin correspondant dans la table.

Dans la phase de reconnaissance, la table de hachage préalablement préparée est utilisée pour la recherche du correspondant du modèle en requête. Les mêmes étapes que le prétraitement sont appliquées à savoir l'extraction des points primitives et l'expression de ces points dans un repère de coordonnées. La nouvelle position des points est utilisée pour accéder au bin de la table de hachage et trouver le meilleur correspondant.

- **Filtrage et groupement des paires de PIs par contraintes géométriques**

Pour accélérer le processus de comparaison, Chen et al. (Chen, et al., 2007) créent une table de hachage, indexée par la paire (μ, σ) représentant la moyenne et la variance des valeurs de l'indice de forme sur le voisinage du PI. L'information (ID du modèle, histogramme, type de la surface, coordonnées 3D) de chaque PI est insérée dans le bin correspondant. La Figure 5-2 schématise ce processus où Si est la fonction de l'indice de forme au point i, L est le nombre de voisins de i, μ et σ sont données par:

$$\mu = \frac{1}{L}\sum_{l=1}^{L} S_i(p_l);\ \sigma^2 = \frac{1}{L-1}\sum_{l=1}^{L}(S_i(p_l) - \mu)^2$$

Pour chaque modèle, ce processus est répété pour construire la base des modèles. Pour un objet test, la similarité, entre son descripteur et ceux des modèles correspondants associés au bin de la table, est calculée à l'aide de la mesure de divergence $\chi 2$ entre les histogrammes.

Deux PIs sont mis en correspondance si la distance de leurs histogrammes est inférieure à un seuil et s'ils ont le même type de surface.

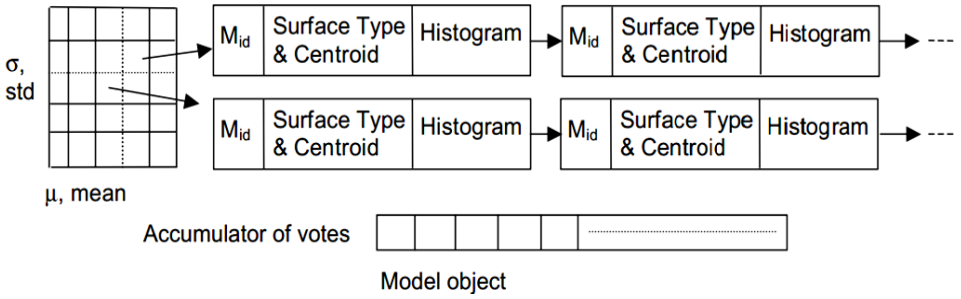


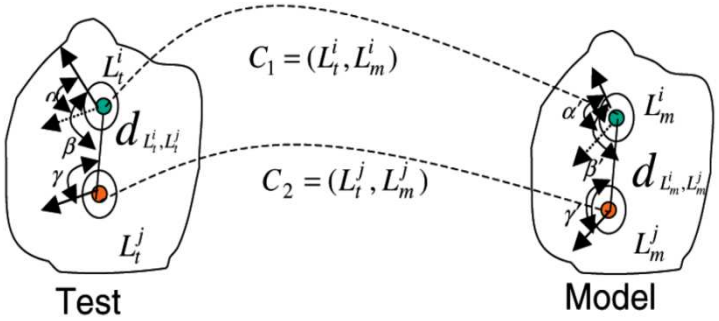
Figure 5-2 Représentation de la table de hachage (Chen, et al., 2007)

Après l'étape de mise en correspondance des PIs par le biais d'une table de hachage (Chen, et al., 2009), un ensemble de proches voisins est retourné. Les paires potentiellement mises en correspondance sont filtrées et regroupées en se basant sur les contraintes géométriques présentées dans les équations de la Figure 5-3 qui présente le cas de deux correspondances ; t désigne la zone de la surface test, et m celle du modèle. Le terme $d_{L_x^i,L_x^j}$ désigne la distance euclidienne entre deux PIs de la même vue d'un objet x. α est l'angle formé par les normales des PIs des deux patches L_t^i et L_t^j. β est l'angle entre la normale du PI de la zone L_t^i et la ligne formée par les centroïdes des deux zones L_t^i et L_t^j. γ est l'angle entre la normale du PI de la zone L_t^j et la ligne formée par les centroïdes des deux zones L_t^i et L_t^j. α', β' et γ' sont définis de la même façon. La valeur absolue $d_{C1,C2}$ de la différence entre les distances des PIs dans la vue test et les distances des PIs dans le modèle doit être inférieure à un seuil ϵ_1. Le maximum de ces distances doit être inférieur à un seuil ϵ_2. La valeur absolue de la différence entre les angles doit être inférieure à un seuil ϵ_3. Une paire de PIs mis en correspondance est dite consistante, si ces deux équations sont vérifiées.

Pour une liste des paires de PIs mis en correspondance, la procédure de groupage est la suivante :

- Initialiser chaque paire du groupe.
- Pour chaque groupe, ajouter d'autres paires si elles satisfont les équations des contraintes géométriques
- Répéter la procédure pour chaque groupe. Sélectionner le groupe ayant la plus grande taille.

Le modèle affecté à l'objet test est celui correspondant au groupe ayant le plus grand nombre de votes.



$$d_{C_1,C_2} = |d_{L_t^i,L_t^j} - d_{L_m^i,L_m^j}| < \epsilon_1 \quad max(d_{L_t^i,L_t^j}, d_{L_m^i,L_m^j}) > \epsilon_2$$
$$(|\alpha - \alpha'|, |\beta - \beta', |\gamma - \gamma'|) < \epsilon_3$$

Figure 5-3 Filtrage des paires de PIs selon des contraintes géométriques (Chen, et al., 2009)

5.1.2 Séparateur à vaste marge

Un séparateur à vaste marge (SVM: acronyme de Support Vector Machines) propose de trouver une frontière linéaire entre deux classes différentes. Le choix particulier du séparateur se base sur la condition de trouver la "marge" maximale. La marge est définie comme la distance minimale entre les points d'apprentissage et la frontière. Les vecteurs supports constituent les éléments délimitant cette frontière : plus la frontière est large, plus les risques d'erreurs de classification sont rares. La fonction de classification retourne alors une valeur positive sur un côté de la ligne de division et négative de l'autre (Cornuéjols, 2002).

Le principe d'un SVM est de projeter les descripteurs dans un nouvel espace appelé espace des caractéristiques (ou un espace de re-description). Par la suite, des fonctions noyau correspondant à un produit scalaire dans cet espace sont définies. Ces fonctions noyau fournissent une mesure de similarité sur les données. Elles encodent également le degré de covariance entre ces données (leur interdépendance). Cet aspect convient parfaitement aux attentes d'un système de reconnaissance quant aux invariances géométriques. En général, l'utilisateur doit tâtonner: essayer des fonctions noyau (linéaire, gaussien, polynomial, etc) et voir si elles permettent l'obtention de bonnes séparatrices. En revanche, le réglage du paramètre de la fonction noyau et du paramètre C (utilisé dans la fonction d'erreur à minimiser) est crucial. Néanmoins, ces choix se font normalement de façon automatisée par validation croisée.

Ce classifieur est utilisé dans (Li, et al., 2007) avec un noyau pyramidal. La technique proposée met en correspondance les descripteurs qui sont transformés en des matrices de similarité en appliquant un noyau pyramidal dont le principe est le suivant : si X_1 et X_2 représentent deux surfaces 3D avec une signature de d-dimension. L'espace de signature est subdivisé en histogrammes multi-résolution de L+1 nivaux avec b_l bins à un niveau l. Le nombre de vecteur matché entre les deux surfaces à un niveau l (s'ils appartiennent au même bin) est :

$$I_l\,(X_1, X_2) = \sum_{n=0}^{b_l} \min\,(H_l({X_1}^n), H_l({X_2}^n))$$

Le noyau qui mesure la similarité entre deux surfaces est donné par:

$$\kappa\,(X_1, X_2) = \sum_{l=0}^{L} \frac{1}{2^l}\,(I_l\,(X_1, X_2) - I_{l-1}\,(X_1, X_2))\ ;$$

Les matrices de similarités constituées par ce noyau sont introduites dans le classifieur SVM.

5.1.3 Adaboost (Adaptive Boosting)

Adaboost est une méthode de dopage (boosting) qui construit un classifieur « fort » avec une combinaison linéaire de classifieurs « faibles ». Par itérations successives, la connaissance d'un classifieur faible est ajoutée au classifieur final.

L'idée principale est de définir à chacune de ses étapes, une nouvelle distribution de probabilité a priori sur les exemples d'apprentissages en fonction des résultats de l'algorithme à l'étape précédente. Des poids sont introduits à chaque étape. Initialement, tous les exemples ont un poids identique, puis à chaque étape, les poids des exemples mal classés par le classifieur faible précédent sont augmentés. Ceci force l'apprenant à se concentrer sur les exemples difficiles de l'échantillon d'apprentissage. La performance de l'apprenant est mesurée par l'erreur pondérée, et idéalement on doit choisir comme classifieur faible celui dont l'erreur pondérée est la plus basse possible. A ce classifieur faible est de plus affecté un poids mesurant l'importance qui lui sera donnée dans la combinaison finale, qui consiste en un vote pondéré. Vers la fin de l'apprentissage, le poids des exemples difficiles à apprendre devient largement dominant et chaque hypothèse de classification est pondérée par un poids final. Le classifieur résultat est une combinaison de ces règles pondérées (Cornuéjols, et al., 2002). Une version améliorée d'AdaBoost, avec des prédictions classées par indice de confiance, est utilisée

par Jones et al. (Jones, et al., 2003). L'algorithme apprend une fonction, de similarité des visages, basée sur une combinaison linéaire de primitives "rectangle" appliquées à deux images.

.

5.1.4 Réseau de neurones

Quant on n'arrive pas à reconnaitre certaines choses dans la vie, l'expression « votre réseau de neurones ne s'est pas suffisamment activé ? » revient souvent. Mathématiquement, un réseau de neurones est constitué d'un grand nombre de cellules de base interconnectées. De nombreuses variantes sont définies selon le choix de la cellule élémentaire (contient des valeurs binaires ou réelles), de l'architecture du réseau (sans ou avec rétroaction) et de la dynamique du réseau (synchrone ou asynchrone). L'apprentissage est une phase du développement d'un réseau de neurones durant laquelle le comportement du réseau est modifié en ajustant ses paramètres (poids), jusqu'à l'obtention du comportement désiré. Un algorithme d'apprentissage permet de déterminer les valeurs des poids permettant à la sortie du réseau de neurones d'être aussi proche que possible de l'objectif fixé. Dans le cas d'un problème de classification supervisée, il s'agit de déterminer une surface de séparation. Cet apprentissage s'effectue grâce à la minimisation d'une fonction, appelée fonction de coût, calculée à partir des exemples de la base d'apprentissage et de la sortie du réseau de neurones (Stricker, 2000). Pour résumé, le principe de fonctionnement d'un réseau de neurones se présente comme suit :

- Apprentissage des poids avec un jeu d'exemples,
- Mesure de l'erreur sur des données conservées pour la validation,
- Arrêt de l'apprentissage quand cette erreur de validation stagne ou augmente

5.1.5 Algorithme de vote des PIs

Le système de vote peut être perçu comme complément à une première classification. En effet, il s'applique à la sortie des classifieurs. Le vote peut se faire selon différents critères (Plantié, 2008) :

– Vote à la majorité simple : choix des classes à la majorité des résultats des classificateurs.

– Vote par choix du maximum (respectivement, minimum) : choix de la classe attribuée par le classifieur qui a donné la probabilité la plus élevée (respectivement, faible) d'appartenance.

– Vote par somme pondérée : le choix de la classe attribuée est alors fondé sur la plus forte moyenne pour chaque classe la moyenne des probabilités de tous les classifieurs est calculée.

Un algorithme de vote par choix du maximum appliqué sur un classifieur calculant le plus proche voisin de PIs est proposé dans (Mayo, et al., 2009) et se présente dans le Tableau 5-2 comme suit :

Algorithme de vote des PIs
Entrées : un ensemble K de PIs extraits et étiquetés par les classes de la base d'apprentissage + un ensemble T de PIs non étiquetés extraits de l'image de test 1- Pour chaque t ∈T faire a. Trouver le plus proche k ∈ K en termes de distance euclidienne, dist(t, k) b. Attribuer l'étiquette de k à t c. Attribuer un poids de $\frac{1}{\text{dist}(t,k)}$ à t Fin pour 2- Chaque t ∈T vote pour la classe de son étiquette avec le poids attribué 3- Somme des votes pour chaque classe et normalisation du vecteur des poids des classes 4- La classification de l'image est la classe ayant **le plus grand** vote (plus grande probabilité) Sorties : la classification de l'image inconnue

Tableau 5-2. Algorithme de vote des PIs (Mayo, et al., 2009)

5.1.6 Conclusion

Dans notre cas, nous utilisons un ensemble de descripteurs des points d'intérêt pour la comparaison d'une requête inconnue aux données de la base d'apprentissage. Les PIs de la requête sont appariés un à un avec les PIs stockés dans la base. La méthode de mise en correspondance des primitives basée sur la recherche du plus proche voisin est la plus adoptée pour l'étape de classification. Nous avons testé, dans un premier temps, l'indexation des primitives PIs par leur valeur de la moyenne et la variance de l'indice de forme à la manière de Chen et al. (Chen, et al., 2007) dans une table de hachage, mais les résultats étaient médiocres. Pour accélérer la recherche du plus proche voisin, nous avons opté pour une structure de Kd-Tree. Par la suite, nous suggérons que l'approche la plus intuitive et la plus simple est un système de vote à la majorité simple suite aux résultats du KD-Tree. Nous considérons un tableau qui permet d'incrémenter les votes pour chaque classe et un tableau pour stocker la somme des distances de PIs. La classe ayant le plus grand vote est associée à la requête. En cas d'égalité entre deux classes, nous attribuons la classe avec le plus faible total de distance. Un filtrage des paires de PIs après leur mise en correspondance, basé sur les contraintes géométriques proposées dans le travail de Chen et al. (Chen, et al., 2009), a été également testé.

5.2. Evaluations des méthodes

Dans ce qui suit, nous allons évaluer la performance des différentes combinaisons des méthodes proposées (détection et description) en termes de « taux de bonne reconnaissance» sur les bases de données présentées dans le chapitre détection.
L'évaluation de l'invariance à l'angle et à l'échelle et de la robustesse au bruit des méthodes sera présentée uniquement sur les données de la base Minolta qui est la seule base intégrant ces variations. Pour le reste des bases des données, nous exposerons les résultats obtenus uniquement pour un scénario de variation d'angles de vues.

5.2.1 Paramètres

Dans cette section, nous exposons le choix des valeurs des paramètres pour notre système de reconnaissance. Quelques paramètres sont identiques entre les bases de données et d'autres changent selon les propriétés des ces bases.
Le nombre et la position des PIs sélectionnés par nos détecteurs combinés, se basent sur le type de leur surface locale. Nous avons menés des expérimentations en variant la combinaison des types des surfaces sélectionnées pour voir l'impact de ce choix sur le résultat de reconnaissance. Les types (dôme (peak), ornière (ridge), selle ornière, non-défini, plan, hyperbolique, cuvette (pit), crête (valley), selle crête) sont notés dans l'ordre de 1 à 9. Cinq façons de combiner ces types sont envisageables. Les tests sont faits avec le détecteur SC_HK_confi et les descripteurs IndSHOT et IndThrift, sur 10 vues de 9 objets de la base Minolta. Les résultats sont donnés dans le Tableau 5-3. Nous constatons que la sélection la combinaison des types {1, 3, 5, 6, 7, 9} donne le plus haut taux de reconnaissance pour les deux descripteurs. En effet, les types peak, saddle ridge, plane, hyperbolique, pit et saddle valley correspondent à des types surlesquels se base, en général, la perception humaine pour différencier les objets. En revanche, le nombre de PIs est supérieur avec les représentations sélectionnant plus de type (par exemple 320 PIs pour la sélection{1, 3, 5, 6, 7, 9}, contre 290 PIs pour la sélection{1, 3, 5, 7, 9}). Il est à noter que les formes des objets de cette base présentent peu d'arêtes de très faibles courbures. Ce type de formes est rencontré surtout sur des objets géométriques, pour lesquelles la sélection des types peut changer par rapport à celle sélectionner pour cette base.

Ces tests ont été réalisés vers la fin de notre travail et nous n'avons pas pu testé la performance des algorithmes avec la meilleure sélection des types sur chaque base. Parmi nos points références dans notre démarche, nous nous sommes basés sur les résultats présentés par Ömer Eskizara (Eskizara, 2009) et qui suggèrent la sélection des types {1, 3, 5, 7, 9}. Nous avons utilisé cette dernière sélection pour toutes nos expérimentations. D'après le Tableau 5-3, la différence, entre les deux sélections, étant de 2% en taux de reconnaissance, nous estimons que les résultats présentés dans ce chapitre auront une légère amélioration avec un détecteur combiné basé sur la sélection des types{1, 3, 5, 6, 7, 9}.

Types testés	{1, 3, 7, 9}	{1, 3, 5, 7, 9}	{1, 2, 3, 5, 7, 9}	{1, 3, 5, 6, 7, 9}	{1, 3, 5, 7, 8, 9}
% Reconnaissance avec IndSHOT	95.5	95.5	92.2	**97.7**	95.5
%Reconnaissance avec IndThrift	91.1	91.1	91.1	**93.3**	92.2

Tableau 5-3 Résultats de reconnaissance en variant les types des PIs

Dans ce qui suit, nous présentons le choix des valeurs du reste des paramètres.
Les paramètres des algorithmes de détection, description et classification, qui ne varient pas sur les différentes bases sont :

- les types de surface sélectionnés sont en forme de dôme, cuvette, plan, selle ornière et selle crête (peak, saddle ridge, plane, pit et saddle valley). Les seuils de classification des types des surfaces sur H, K, SI et C sont fixés.
- le nombre minimal de voisins égal à 18 (correspond à 2 rings de voisinage sur le maillage).
- la distance minimale tolérée des sommets par rapport au bord correspond à celle du rayon du voisinage.
- le nombre de bins pour les histogrammes est déduit de l'évaluation des paramètres dans le chapitre « Descripteurs ». Ces valeurs sont: nBin_CosNormals (12), nBin_Index (10) et nBin_Spin (10). Ce qui donne au final, les dimensions : 416, 768, 120, 120, 100 et 12, pour respectivement les descripteurs SHOT, IndSHOT, IndThrift, LSP, Spin et Thrift.

Les paramètres dont la valeur change, selon la base testée, sont : le rayon (Rayon) utilisé dans le calcul du voisinage (détection et descripteur) et qui correspond à une proportion de la diagonale (Diag), et le paramètre (ParSelect_PIs) utilisé pour filtrer les PIs selon la distance spatiale entre eux, après une première extraction par les algorithmes combinés de détection. Selon la nature de la base de données utilisée, ces valeurs sont comme suit:

- Minolta : La valeur du Rayon est 0.08 x Diag; et celle du ParSelect_PIs est égale à 0.07 x Diag. Ces paramètres, qui décident du nombre et positions des PIs, sont déterminés empiriquement par rapport aux tests effectués dans les chapitres Détecteurs et Descripteurs. L'optimisation à la fois du temps de calcul et du pouvoir descriptif passe par le choix de ces paramètres. Les paramètres des autres bases sont choisis relativement à ceux de cette base.
- Stuttgart : La valeur Rayon est égale à 0.04 x Diag. Le rayon est plus faible puisque la densité de points est plus grande et la résolution est plus faible (distance des points est plus petite que celle de Minolta). La valeur choisie de ParSelect_PIs (0.04 x Diag) permet de sélectionner un nombre de PIs proportionnel à la densité des données.
- Pour les bases Carotte et RGBD, les données acquises par la Kinect ont une faible densité de points et les objets ont des petites tailles. Nous avons utilisé les valeurs: 0.15 x Diag pour le

Rayon et 0.09 x Diag pour le ParSelect_PIs. Le nombre de PIs est relativement plus faible que celui des autres bases.

En dehors de la base des 9 objets Minolta, les valeurs de ces paramètres sur les autres bases sont déterminées d'une façon approximative. Le nombre de primitives extraites sur les différentes bases varie de 4 à 400 keypoints par vue, selon l'angle de vue et la complexité de la forme de l'objet.

Concernant le calcul des quantités différentielles (normales et courbures), nous avons testé deux algorithmes: le premier utilise le système de Monge Form et le deuxième calcule la matrice de covariance comme introduite par (Tombari, et al., 2010). Pour les résultats présentés dans ce chapitre, nous avons utilisé le système de Monge Form qui donne des meilleurs résultats d'estimation des quantités mais qui est plus lent en temps de calcul. Par ailleurs, pour le calcul de la RF utilisée dans les descripteurs SHOT et IndSHOT nous avons gardé l'algorithme d'origine (celui du SHOT).

Dans la suite de ce chapitre, nous complétons notre évaluation qualitative des méthodes proposées par une comparaison de la performance de reconnaissance avec des approches de la littérature, sous différents scénarios. Sachant que, dans les chapitres précédents, les évaluations et le paramétrage de nos algorithmes ont été faits sur la base des 9 objets Minolta (avec les 18 vues de 20° d'écart d'angle), nous nous attendons à ce que les résultats de reconnaissance soient plus performants sur cette base que sur les autres bases à tester.

5.2.2 Evaluation de l'invariance et de la robustesse sur la base Minolta

Rappelons que la base Minolta est une collection d'images réelles de profondeur, présentant un jeu de données avec des variations d'angles et d'échelles. Chen et al. (Chen, et al., 2007) ont sélectionné 20 objets de la base Minolta pour tester la performance de leurs descripteurs LSP. Toutes les paires modèle-test ont un écart d'angle de 20° sauf les paires des objets 3, 14 et 19 qui ont un écart d'angle de 36° (illustration dans la partie haute de la Figure 3-19 du chapitre 3). Leur résultat de reconnaissance sur cette base est de 85%. Le scénario de reconnaissance n'étant pas clairement décrit dans leur travail, nous ne pouvons pas nous comparer à leur résultat. Par contre, nous avons ajouté à ces 20 objets 5 autres objets dont deux de forme géométrique plane et cône (représentés dans la Figure 5-4) et nous avons menées nos expérimentations sur ces 25 objets. Notre version du descripteur LSP est alors testée sur cette base. Une étude comparative des différentes combinaisons des méthodes proposées est donnée pour une évaluation de l'invariance à l'angle de vue sur cette base.

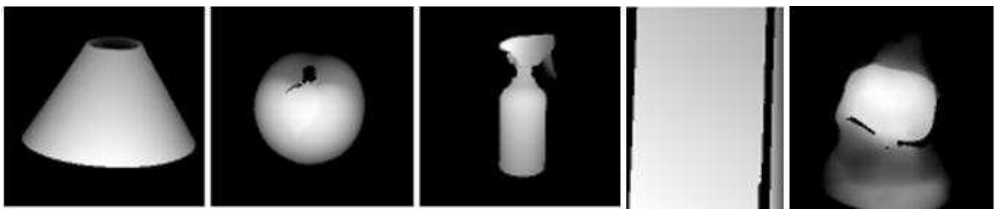


Figure 5-4 Les 5 objets Minolta ajoutés au 20 objets utilisés pour évaluer la robustesse à la rotation

D'autre part, nous considérons les objets de la base Minolta présentant une variation d'échelles et que nous avons introduits dans le chapitre détection (dans la Figure 3-20). Une évaluation de l'invariance des méthodes à un changement d'échelle est proposée pour cette base. La question à ce niveau est de savoir si le faible nombre d'objets de cette base peut donner une réelle évaluation de l'invariance à l'échelle. Bien que l'étude soit faite sur seulement 7 objets, la forme de ces objets regroupe des objets à aspect naturel (visages, brain) et des objets à aspect industriel (gc_bottle, valve et yellowhorn). D'autre part, quelques objets présentent des similitudes entre eux comme reddino et bluedino. En

outre, les 11 vues prises par objet permettent d'augmenter la variation dans le jeu de données et d'avoir un résultat moyennant plusieurs tests. Finalement, avec cette base, nous constituons, une représentation miniature des contraintes qui peuvent exister dans une plus grande base, et une possibilité de se positionner dans un cas de scénario réelle de reconnaissance.

5.2.2.1 Invariance à la rotation

❖ **Description**

Parmi les défis majeurs dans notre conception d'un système de reconnaissance, l'invariance aux transformations géométriques a le plus grande poids. Pour mesurer la performance des différents systèmes proposés à une variation d'angles de vues, une évaluation est faite selon le scénario décrit dans la Figure 5-5. Ce protocole prend, à chaque expérimentation, une vue test d'un objet et apprend sur le reste des vues de cet objet et sur toutes les vues des autres objets. Nous l'appelons scénario de rotation.

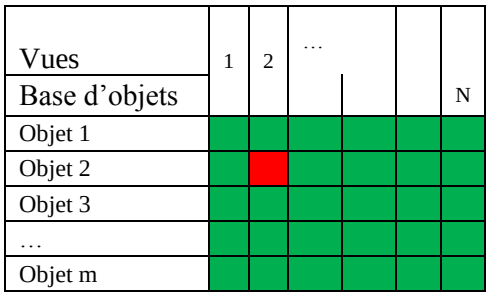


Figure 5-5 Scénario de reconnaissance pour la rotation: la case rouge représente la vue n°2 prise en test pour l'objet 2 et les cases vertes correspondent aux vues d'apprentissage

Nous commençons par une évaluation sur les 9 objets de la base Minolta avec les 18 vues de 20° d'écart d'angle. En requête, nous donnons une vue d'un objet et apprenons sur les 161 vues restantes. Nous répétons cette opération pour toutes les vues de chaque objet. Les taux de reconnaissance, présentés dans le Tableau 5-4, sont moyennés sur le total de 162 (18x9) tests de tous les objets et toutes les vues de la base. En dehors du SURF, les systèmes de reconnaissance testés combinent les détecteurs : SC_HK_FQ, SC_HK_confi, SC_HK_C, SI, HK, SC, Harris_fract et Harris_clust, avec les descripteurs: IndSHOT, SHOT, IndThrift, LSP, Spin et Thrift. Dans le Tableau 5-5, nous donnons le résultat de reconnaissance sur la base constituée des 25 objets Minolta.

❖ **Résultats**

	IndSHOT	SHOT	IndThrift	LSP	Spin	Thrift	SURF
SC_HK_FQ	**92.0**	85.2	92.0	86.5	71.1	17.8	
SC_HK_confi	**95.7**	84.6	93.2	88.3	73.0	19.6	
SC_HK_C	**93.8**	87.7	92.6	87.7	72.4	22.7	
SI	**92.6**	81.6	84.6	81.6	70.5	15.9	
HK	88.9	80.3	**92.0**	85.8	73.0	22.0	
SC	**93.8**	85.2	91.4	85.8	75.4	27.6	
Harris_fract	85.8	84.6	**88.9**	81.6	65.0	30.6	
Harris_clust	**74.8**	66.2	70.5	58.9	50.3	14.7	
SURF							**92.6**

Tableau 5-4 Résultats de reconnaissance pour les 9 objets Minolta- les taux les plus élevés, pour chaque détecteur, sont mis en gras et le meilleur taux du tableau est mis en rouge

- **Résultats sur les 9 objets**

➢ Dans l'ensemble, les détecteurs proposés SC_HK_FQ, SC_HK_confi et SC_HK_C ont le même ordre de performance. Une légère supériorité du détecteur SC_HK_confi composé avec le descripteur IndSHOT, avec un taux de 95.7% de bonne reconnaissance, est notée. Les meilleurs taux réalisés par les approches de détection testées de la littérature, sont 92.6% pour le SURF, 92.6% pour SI- IndSHOT, 93.8% pour SC- IndSHOT, 92.0% pour HK- IndThrift et 88.9% pour Harris_fract-IndSHOT. La comparaison entre le Harris_fract et le Harris_clust suggère qu'une sélection des PIs, basée sur l'extraction d'une proportion des PIs les plus saillants (avec la plus grande réponse de Harris) donne des meilleurs résultats qu'une sélection par clustering contrainte à la distance géométrique entre les PIs.

➢ Indépendamment du détecteur choisi, la comparaison entre les descripteurs permet de confirmer la nette supériorité des descripteurs combinés IndSHOT et IndThrift. L'écart réalisé avec le Thrift est remarquable par rapport au reste des descripteurs. Les performances médiocres de ce dernier sont probablement dues à la faible dimension choisie pour ce descripteur qui est égale à 12 par rapport aux dimensions des autres descripteurs de taille supérieure à 100; de plus, le formalisme du Thrift est le seul descripteur (à part le SHOT) à ne pas combiner plusieurs propriétés. Par ailleurs, nous ne sommes pas intéressés à optimiser le fonctionnement de ce descripteur, à partir du moment où notre nouvelle combinaison de ce descripteur avec l'information de l'indice de forme (IndThrift) a donnée une nette meilleure performance pour un même paramétrage.

➢ Le succès de la reconnaissance sur la base des 9 objets Minolta, s'explique par le degré et la clarté des détails présents sur ces objets. En effet, le processus d'extraction et de description parvient à formuler des primitives descriptives et distinctes. A l'exception du cas où les zones planes sont dominantes sur la surface du modèle, notre détection des PIs favorise essentiellement les zones de forte variation locale de la forme, par un tri selon une mesure de courbure (C, FQ ou Conf). En outre, la qualité des nuages de points de cette base (densité de points et bruit) permettent une bonne estimation des éléments fondamentaux de nos algorithmes, à savoir les courbures et les normales. Cependant, le processus d'extraction, sur ce type de données, reste couteux en temps de calcul.

- **Résultats sur les 25 objets**

Ces résultats sont consolidés par les tests effectués sur les 25 objets Minolta et qui sont donnés dans le Tableau 5-5. Le descripteur Thrift n'est pas représenté dans ce tableau compte tenu de sa très faible performance en comparaison avec les autres descripteurs.

➢ D'après ce deuxième tableau, le taux de reconnaissance des différentes méthodes chute par rapport à celui des 9 objets, ce qui était prévisible et attendu sachant que le nombre des faux appariements augmentent quand le nombre d'objets de la base est plus grand.

➢ Pareillement, pour les résultats de reconnaissance des 25 objets, les mêmes explications, données plus haut sur les 9 objets, sont valables. En utilisant le descripteur IndSHOT, le détecteur SC_HK_C donne un meilleur taux de reconnaissance (84.5%) que celui donné par HK (83.7%) et par SC (82.9%), ce qui prouve l'amélioration apportée par notre contribution de coupler les deux espaces de classifications. D'autre part, la combinaison du détecteur SI avec le descripteur IndSHOT excède les autres méthodes avec un taux de 86.6%. En prenant le même détecteur SI, le taux réalisé avec le SHOT est de 75.3%, avec l'IndThrift est de 81.4%, et le LSP est de 74.7%. Le formalisme de ce nouveau descripteur permet une nette hausse de la performance de reconnaissance. La robustesse de ce descripteur pour un changement de l'angle de vue est confirmée. Concernant le temps de calcul, les descripteurs ont le même ordre de grandeur d'exécution sauf pour l'IndThrift qui est 7 fois plus couteux que les autres.

- En comparaison avec l'état de l'art, l'approche du SURF (détecteur et descripteur) réalise un taux de 82.8% de bonne reconnaissance sur la base des 25 objets. Nous concluons que les combinaisons proposées utilisant le descripteur IndSHOT avec les détecteurs SC_HK_confi, SC_HK_C et SI apportent une amélioration en reconnaissance invariante à l'angle de vue, par rapport à l'état de l'art.

	IndSHOT	SHOT	IndThrift	LSP	Spin	SURF
SC_HK_FQ	**83.1**	76.4	79.3	69.9	55.9	
SC_HK_confi	**84.1**	74.7	80.1	74.1	57.6	
SC_HK_C	**84.5**	83.7	80.1	67.8	58.2	
SI	**86.6**	75.3	81.4	74.7	58.8	
HK	**83.7**	72.0	81.4	73.7	58.6	
SC	**82.9**	74.5	80.1	71.1	56.3	
Harris_fract	**83.5**	83.3	79.5	73.2	64.5	
Harris_clust	**79.1**	73.9	67.2	60.9	54.9	
SURF						**82.8**

Tableau 5-5 Résultats de reconnaissance pour les 25 objets Minolta

- **Taux de reconnaissance par objet** :

Un autre aspect de comparaison des taux de reconnaissance consiste à évaluer ces taux pour les objets séparément. Si nous calculons ce taux par objet, nous remarquons que certains objets sont plus reconnaissables que d'autres selon la complexité de la forme et la densité des points dans le modèle. Cette comparaison est donnée dans le Tableau 5-6 qui montre le résultat du détecteur SC_HK_confi combiné avec les deux descripteurs IndSHOT et IndThrift. Nous pouvons noter que l'objet buddha a le plus faible taux de reconnaissance, ce qui est probablement du à la densité élevée des points dans ses vues en comparaison avec les vues des autres objets. Par exemple, la vue 0 contient 14672 points et 28279 facettes, par rapport à la vue 0 de bunny qui comporte 6787 points et 12668 facettes. De plus, pour un objet dont la forme présente beaucoup de détails, comme le cas de l'objet buddha et frog, le nombre de PIs extraits croit et peut induire à plus de confusion dans le processus d'appariement si ces mêmes primitives apparaissent sur d'autres objets. Ce nombre important peut être, par contre, avantageux si la forme décrite se distingue des autres objets (comme le cas de l'objet Duck). D'autre part, un nombre faible de PIs par objet peut ne pas être suffisant pour discriminer l'objet et l'appariement risque de disperser le vote sur plusieurs objets, tel est le cas pour l'objet Bunny.

Objets testés		IndThrift	IndSHOT
Angel	ID du modèle	1	1
~100 PIs	Taux de reconnaissance	**94 .4%**	**94.4%**
Pooh	ID du modèle	2	2
~115 PIs	Taux de reconnaissance	**100%**	**100%**
Bird	ID du modèle	3	3
~120 PIs	Taux de reconnaissance	**100%**	**100%**
Buddha	ID du modèle	4	4
~130 PIs	Taux de reconnaissance	77.8%	**88.9%**
Dough_boy	ID du modèle	5	5
~45 PIs	Taux de reconnaissance	**100%**	**100%**
Duck	ID du modèle	6	6
~140 PIs	Taux de reconnaissance	94.4%	**100%**
Frog	ID du modèle	7	7
~135 PIs	Taux de reconnaissance	**88.9%**	83.3%
Lobster	ID du modèle	8	8
~115 PIs	Taux de reconnaissance	**100%**	**100%**
Bunny	ID du modèle	9	9
~80 PIs	Taux de reconnaissance	83 ,3%	**94.4%**

Tableau 5-6 Taux de reconnaissance par objet pour le détecteur SC_HK_confi

- **Impact du choix des vues pour le protocole** : Par ailleurs, le choix du protocole de reconnaissance est crucial et intervient sur le résultat final. En effet, selon la position (angle) de la vue test choisie parmi les vues apprises, le taux de reconnaissance varie. Ce taux est, sans surprise, plus élevé pour des vues tests dont l'angle est situé entre deux angles de vues apprises comme le montre le Tableau 5-7 (angle 140° vs angle 0°). La raison derrière ce résultat est que la majorité des PIs de la vue intermédiaire sont appariés avec la vue qui précède et la vue qui suit la vue en question, faisant augmenter, ainsi, le pourcentage des bonnes correspondances.

	Frame 0°	Frame 140°	Frame 280°
Taux de reconnaissance IndThrift	88.9%	**100%**	88.9%
Taux de reconnaissance **IndSHOT**	100%	**100%**	88.9%

Tableau 5-7 Taux moyen de reconnaissance selon la vue, pour le détecteur SC_HK_confi sur les 9 objets Minolta

❖ **Conclusion**

A l'issu des résultats obtenus, nous confirmons que les combinaisons détecteurs – descripteurs proposées ont une performance satisfaisante concernant l'invariance à un changement d'angles de

vues. Sur l'ensemble des objets Minolta, le descripteur IndSHOT a une assez nette supériorité par rapport au descripteur IndThrift et la performance des détecteurs est complétée par celle du descripteur.

En conclusion de cette partie, un système de reconnaissance d'objets est sensible à la nature du couple détecteur-descripteur choisi, à la forme des objets de la base et au protocole de reconnaissance utilisé (les vues choisies pour l'apprentissage et pour le test). En général, selon l'application finale, un ajustement de ces éléments est à prendre en compte pour maximiser la performance de reconnaissance.

Dans la suite de notre évaluation de l'invariance aux transformations géométriques, nous abordons l'impact d'un changement d'échelles. En se basant sur les résultats de cette partie, nous allons restreindre notre évaluation des systèmes de reconnaissance aux deux meilleurs descripteurs IndSHOT et IndThrift.

5.2.2.2 Variation d'échelles

❖ **<u>Scénarios</u> :**

Dans cette partie, nous proposons 3 scénarios de reconnaissance pour évaluer l'invariance du système à un changement d'échelle.

Scénario1 : Ce scénario prend une vue par objet en test et apprend sur le reste des vues et échelles. On répète cette requête pour toute les vues de chaque objet (Figure 5-6). Nous pouvons dire que ce scénario combine la variance de l'échelle et de l'angle de vue. Ce protocole suppose que l'apprentissage soit riche et fait sur un grand nombre de vues et échelles possibles.

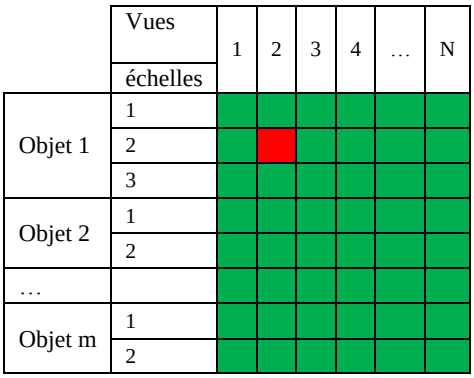


Figure 5-6 Premier scénario de reconnaissance pour la variation d'échelle

Scénario 2 : Ce scénario prend, pour un objet, une vue à une échelle donnée en test et apprend sur les autres vues des échelles différentes de la vue et échelle prises en test (Figure 5-7). Dans ce scénario, la vue test est complètement différente des vues apprises ce qui parait être un cas assez proche des applications réelles. En effet, généralement, la vue requête n'est pas forcement capturée à la même échelle ou au même angle de vue que les vues contenues dans la base d'apprentissage. Ce scénario d'évaluation de l'invariance à l'échelle est, de ce fait, le plus critique pour un système de reconnaissance.

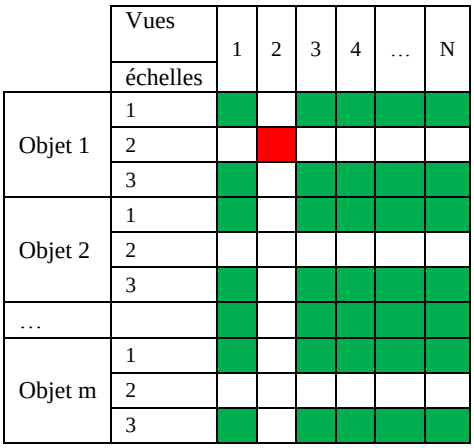


Figure 5-7 Deuxième scénario de reconnaissance pour la variation d'échelle

Scénario 3 : Ce scénario prend en test une vue à une échelle par objet et apprend sur la même prise de vue dans les autres échelles (Figure 5-8). Ce cas de figure correspond par exemple à une application où la caméra est fixe (l'angle de vue ne change pas) et que plusieurs captures de l'objet soient prises à différentes distances (objet se déplaçant dans un couloir étroit). Ce protocole permet d'évaluer l'impact d'une variation d'échelle à un angle fixe pour notre système de reconnaissance. En outre, ce scénario permet de mesurer la performance d'un système qui apprend sur un faible nombre de vues par objet.

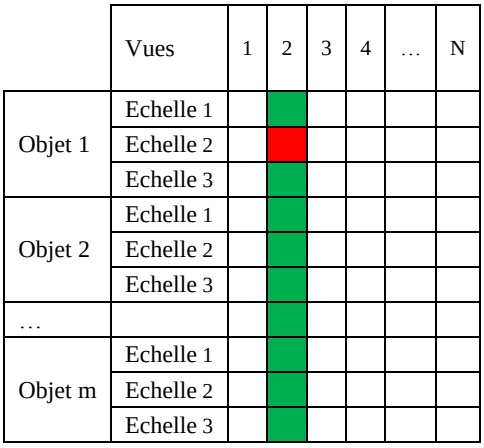


Figure 5-8 Troisième scénario de reconnaissance pour la variation d'échelle

Pour les tests qui suivent, nous présentons le taux de reconnaissance comme un pourcentage des vues reconnues par rapport à l'ensemble des vues testées de la base.

<u>Base à variation *artificielle* d'échelles</u>

Dans un premier temps de nos expérimentations, nous proposons d'utiliser la base à variation artificielle d'échelles introduite dans le chapitre 3 pour l'évaluation des détecteurs. Cette base est constituée de trois variations d'échelles.

Les résultats du scénario 1 et 3 sur cette base sont donnés dans le Tableau 5-8 et le Tableau 5-9. Ces résultats permettent d'avoir une première idée sur le comportement de nos algorithmes pour un changement d'échelles, simulé à une variation dans la densité des données. Cependant, nos

conclusions finales seront essentiellement basées sur les évaluations de la base de variation réelle d'échelles et nous verrons si ses résultats consolident ceux obtenus sur cette première base artificielle. L'analyse des résultats de la première base sera donnée conjointement avec l'analyse des résultats de la deuxième base.

Base à variation *réelle* d'échelles

Dans un deuxième temps de notre évaluation, les données des 7 objets de la base Minolta, présentant une variation de vues et d'échelles (11 vues sur 3 à 4 échelles), sont utilisées dans les trois scénarios décrits plus haut. Pour le scénario 1, toutes les vues et échelles sont considérées et pour le scénario 3, les 11 vues sur 3 échelles de chaque objet sont utilisées. Pour les tests du scénario 2, nous avons sélectionné pour chaque objet 5 vues de chacune des 3 échelles. Les résultats sont présentés dans les tableaux: Tableau 5-10, Tableau 5-11 et Tableau 5-12. Le taux de reconnaissance correspond au pourcentage des vues correctement reconnues par rapport aux vues testées, pour 7 détecteurs et trois descripteurs.

❖ **Résultats et analyses**

	IndSHOT	IndThrift	SURF
SC_HK_confi	97.1	**97.9**	
SC_HK_FQ	**97.7**	97.1	
SC_HK_C	**99.0**	97.7	
SC	**99.0**	97.1	
HK	**98.5**	97.9	
SI	**98.5**	96.7	
Harris_fract	**100**	**100**	
Harris_cluster	99.0	**99.8**	
SURF			**99.8**

Tableau 5-8 Résultats du scénario 1 sur les données d'échelle artificielle des 9 objets Minolta

	IndSHOT	**IndThrift**	SURF
SC_HK_confi	90.3	**98.3**	
SC_HK_FQ	92.0	**98.7**	
SC_HK_C	94.2	**98.7**	
SC	92.0	**98.1**	
HK	91.7	**98.1**	
SI	98.7	**98.3**	
Harris_fract	**100**	**100**	
Harris_cluster	**100**	**100**	
SURF			**100**

Tableau 5-9 Résultats scénario 3 sur les données d'échelle artificielle des 9 objets Minolta

	IndSHOT	IndThrift	SURF
SC_HK_confi	**97.2**	**97.2**	
SC_HK_FQ	97.2	**98.0**	
SC_HK_C	**98.0**	96.8	
SI	**97.2**	95.6	
Harris_fract	**97.2**	92.5	
Harris_cluster	**92.9**	88.1	
SURF			**97.2**

Tableau 5-10 Résultats du scénario 1 sur les données d'échelle réelle des 7 objets Minolta

	IndSHOT	IndThrift	SURF
SC_HK_confi %Reco /echelle	86.3 **84.3 - 91.4 - 84.3**	**86.6**	
SC_HK_C %Reco /echelle	**88.6** **87.1 – 90.0 – 88.6**	86.6	
SC_HK_FQ %Reco /echelle	86.2 **85.7 87.1 85.7**	**88.6**	
SI %Reco /echelle	79.0 **80 78.6 78.6**	**80.9**	
Harris_cluster %Reco /echelle	67.6 **65.7 68.6 68.6**	**68.6**	
Harris_fract %Reco /echelle	79.3 **80.9 82.8 74.3**	**82.8**	
SURF			**71.9**

Tableau 5-11 Résultats du scénario 2 sur les données d'échelle réelle des 7 objets Minolta

	IndSHOT	IndThrift	SURF
SC_HK_confi	92.2	**94.4**	
SC_HK_C	93.1	**93.5**	
SC_HK_FQ	92.2	**93.5**	
SI	**90.9**	89.6	
Harris_fract	**94.8**	86.6	
Harris_cluster	**82.6**	79.7	
SURF			**93.5**

Tableau 5-12 Résultats du scénario 3 sur les données d'échelle réelle des 7 objets

- Pour le **scénario 1**, sur la **première base**, le Harris_fract et le Harris_cluster combinés respectivement avec l'IndSHOT et l'IndThrift donnent 100% de taux de reconnaissance, suivi par le SURF avec 99.8%. Alors que sur la deuxième base qui nous donne l'estimation réelle de l'invariance, c'est le couple SC_HK_C- IndSHOT qui l'emporte avec 98.0%. Concernant les descripteurs, les résultats sont de même ordre avec les deux descripteurs IndSHOT et IndThrift avec une légère supériorité du IndSHOT. Sur le premier scénario, notre conclusion n'est pas tranchante sur le couple (détecteur, descripteur) réalisant la meilleure performance

pour une variation d'échelles sachant surtout que les resultants, dans ce cas de figure, peuvent être influencé par la variation d'angles de vue.

- Pour le **scénario 3**, la combinaison du Harris_fract avec l'IndSHOT donne le meilleur résultat sur les deux bases d'échelles. Le couple formé par SC_HK_confi et IndThrift donne quasiment le même résultat que ce dernier avec un taux prometteur de 94.4% sur la deuxième base. Nous pouvons également confirmer la supériorité des résultats obtenus avec le descripteur IndThrift par rapport à ceux de l'IndSHOT, pour les trois nouveaux détecteur proposés. Par ailleurs, ces résultats appuient la pertinence de nos algorithmes pour un scénario qui répond au défi de reconnaitre les objets à partir de peu de vues en apprentissage (ici, seulement 2 vues par objet en apprentissage).
- Dans le tableau du **scénario 2**, la colonne du descripteur IndSHOT représente le résultat global moyenné sur les trois échelles en première ligne, et en deuxième ligne, le résultat individuel correspondant à la moyenne des taux de reconnaissance des vues tests prises à chacune des trois échelles. D'après ce tableau, les couples (SC_HK_FQ, IndThrift) et (SC_HK_C, IndSHOT) réalisent le plus haut taux avec 88.6% par rapport à 71.9% donné par le SURF. Sachant que ce scénario est le plus déterminant de la performance d'un système de reconnaissance dans des conditions réelles, nous pouvons conclure de la légère amélioration apportée par les nouveaux détecteurs – descripteurs.
- Pour une **analyse plus détaillée**, le taux de reconnaissance est meilleur quand les vues de test se situent sur une échelle intermédiaire entre deux échelles ce qui correspond au deuxième taux sur les 3 taux de la ligne (exemple 91.4% par rapport à 84.3% sur les deux autres échelles). Ce résultat est complètement prévisible. En effet, plus l'échelle est grande (on se rapproche de la caméra) plus les détails de la forme sont visibles, ce qui en résulte un nombre de primitives plus important avec une description plus fine de la forme. Si l'apprentissage se fait sur les données d'une échelle plus proche de la caméra que celle de la vue en requête, le processus de mise en correspondance réussit à mieux apparier les PIs entre eux. A l'inverse, si l'apprentissage se fait sur les vues les plus éloignées, les PIs d'une vue test à une échelle plus proche ne seront pas tous correctement appariés car le degré de détails est différent. Pour la situation où la vue test se positionne à une échelle intermédiaire, la mise en correspondance se fait avec les vues des deux échelles qui la bornent, ce qui réduit le nombre des faux appariements par rapport aux deux autres cas. D'autre part, dans cette base, les vues prises à une échelle proche du capteur, présentent souvent des parties manquantes de l'objet. Nous pouvons dire que cette évaluation permet aussi d'estimer la robustesse du système de reconnaissance à des scénarios d'occultation.

- Sur les trois scénarios, la complémentarité du détecteur SC_HK_C avec le descripteur IndSHOT donne relativement une meilleure stabilité pour une variation d'échelle sur des données réelles. Cependant, la remarque générale est que les résultats varient selon le choix du couple détecteur et descripteur. En outre, les caractéristiques liées à la <u>densité</u> des points, <u>l'angle de la prise de vue</u> et la <u>nature</u> des <u>détails</u> de la forme, impactent sur la performance de la reconnaissance en variation d'échelle.

5.2.2.3 Robustesse au bruit

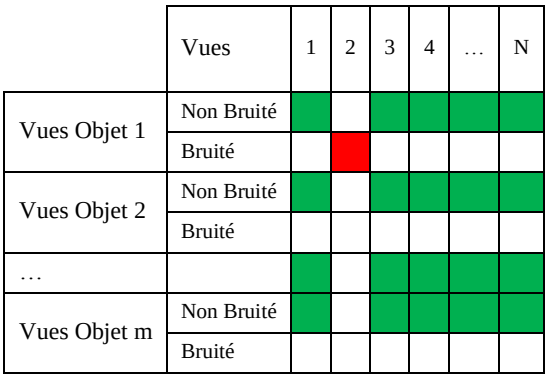
❖ **<u>Scénarios</u>**

Un autre point important pour la validation de notre système de reconnaissance est de tester son comportement pour des données bruitées. Nous suggérons trois façons d'étudier cette problématique. La première suppose que les conditions d'acquisition liées au capteur ou à l'objet, au moment de la

reconnaissance, soient différentes de celles de l'apprentissage, ce qui donne, par exemple, un scénario d'un apprentissage sur des données non bruitées et un test sur des données bruitées. Le deuxième cas suggère que les vues, en apprentissage comme en test, soient acquises à plusieurs moments avec différentes conditions, ce qui en résulte un apprentissage et un test sur des données mélangées bruitées et non bruitées. Et le troisième scénario suppose que les données soient de nature bruitées aux moments de l'apprentissage et du test. Ce dernier cas correspond à une évaluation de la reconnaissance directement sur les données des deux bases Kinect (Carotte et RGBD) qui présentent cette propriété. Les résultats de cette évaluation seront données dans les parties 5.2.4 et 5.2.5. L'étude de la robustesse au bruit pour le premier et deuxième scénario est faite sur les 9 objets Minolta.
Le premier protocole proposé teste sur une vue bruitée et apprend sur les vues non bruitées dont l'angle est différent de celui de la vue en test. Un cas d'application possible pour ce protocole peut être l'utilisation d'un capteur en apprentissage et d'un autre différent en test. La Figure 5-9 illustre ce scénario qu'on appelle scénario B1. Nous disposons pour chaque objet de 5 vues initiales non bruitées et de 5 vues bruitées. L'ensemble des vues bruitées de chaque objet est pris en test. Trois expérimentations sont faites en ajoutant aux données, un bruit gaussien de variance égale à 0.1 MR, 0.2 MR et 0.5 MR.

Les résultats du scénario B1 sont donnés dans le Tableau 5-13 et Tableau 5-14 en termes de taux de reconnaissance. Le Tableau 5-13 présente une évaluation de la stabilité à un bruit de variance 0.2 MR pour les 7 détecteurs et trois descripteurs. Quant aux résultats de l'ajout d'un bruit de variance 0.1 MR et 0.5 MR, ils sont données pour les 4 détecteurs SC_HK_confi, SC_HK_C, SC_HK_FQ et SI et les deux descripteurs IndSHOT et IndThrift. Nous traçons, dans la Figure 5-10, les courbes illustrant l'évolution du taux de reconnaissance pour ces 4 détecteurs en fonction des quatre degrés du bruit (0 MR, 0.1 MR, 0.2 MR et 0.5 MR).

Le deuxième protocole prend un mélange de vues bruitées et non bruitées en apprentissage et essaie de reconnaitre une vue (bruitée ou pas) différente des vues apprises. Un tel scénario peut correspondre à un cas d'application réelle utilisant différents capteurs en apprentissage et en test. La Figure 5-9 illustre ce scénario qu'on appelle scénario B2. Dans nos tests, nous sélectionnons, pour chaque objet, 9 vues des vues paires non bruitées et 9 vues des vues impaires de l'ensemble des vues bruitées. Les résultats sur l'ensemble des 18 vues pour un ajout de bruit de variance 0.2 MR sont donnés dans le Tableau 5-15.



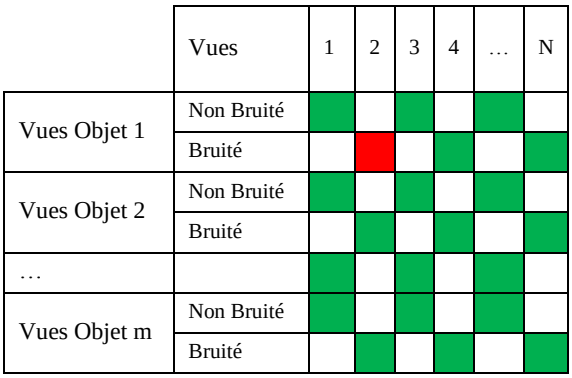


Figure 5-9 Les deux scénarios de reconnaissance pour l'évaluation de la robustesse au bruit- en haut: scénario B1 et en bas: scénario B2

	Sans Bruit			Bruit 0.2 MR		
	IndSHOT	IndThrift	SURF	IndSHOT	IndThrift	SURF
SC_HK_confi	**86.6**	80.0		68.9	**75.5**	
SC_HK_C	**91.1**	86.6		77.8	**77.8**	
SC_HK_FQ	**88.9**	88.9		75.5	**80.0**	
SI	**84.4**	77.8		**80.0**	60.0	
Harris_fract	80.0	**82.2**		**86.6**	84.4	
Harris_cluster	**80.0**	66.6		71.1	73.3	
SURF			**71.1**			**71.1**

Tableau 5-13 Résultats du scénario B1 pour un ajout de bruit de variance 0.2 MR vs les résultats sans bruit, sur 9 objets Minolta

	Sans Bruit		0.1MR		0.5MR	
	IndSHOT	IndThrift	IndSHOT	IndThrift	IndSHOT	IndThrift
SC_HK_confi	**86.6**	80.0	**84.4**	82.2	**13.3**	11.1
SC_HK_C	**91.1**	86.6	86.6	**88.9**	**11.1**	8.9
SC_HK_FQ	**88.9**	88.9	**84.4**	82.2	**15.5**	8.9
SI	**84.4**	77.8	**80.0**	75.5	**11.1**	11.1

Tableau 5-14 Résultats du scénario B1 pour un ajout de bruit de variance 0.1 MR et 0.5 MR sur les 9 objets Minolta

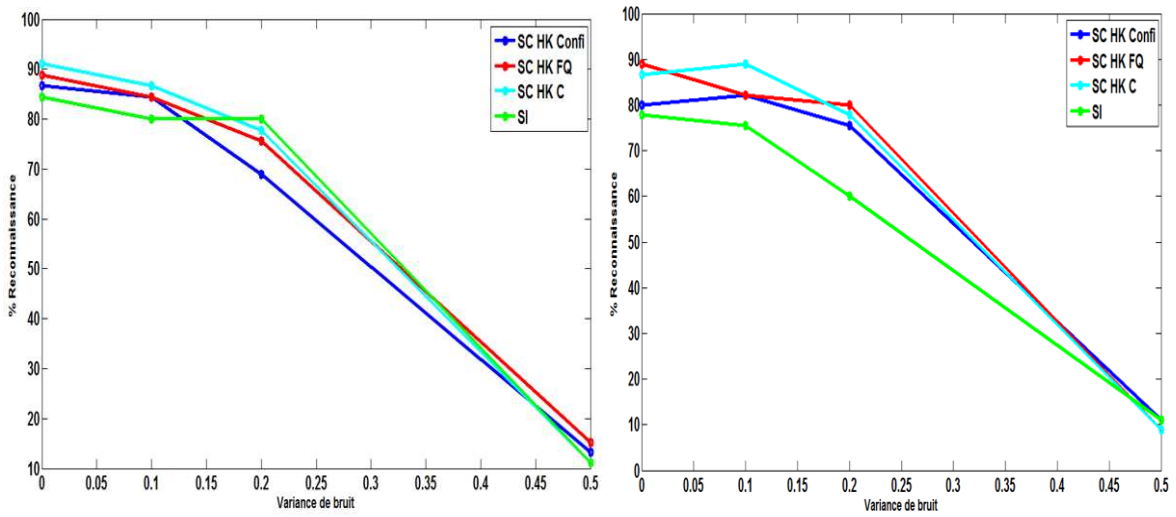


Figure 5-10 Courbes d'évolution du taux de reconnaissance pour les 4 détecteurs SC_HK_confi, SC_HK_C, SC_HK_FQ et SI en fonction de 4 degrés de bruit, pour le descripteur IndSHOT (à gauche) et le descripteur IndThrift (à droite).

❖ **Résultats et analyses**

- Dans une observation globale des résultats, il n'est pas évident de décider de la combinaison détecteur- descripteur la plus performante compte tenu à la fois du taux de reconnaissance et de la stabilité au bruit. Avec l'ajout d'un bruit de variance 0.2MR, la performance du SURF reste stable, mais avec un taux de reconnaissance plus faible que celui de nos systèmes. Nous remarquons que le taux obtenu avec le détecteur Harris_fract augmente en présence de bruit. Le même comportement est observé avec les couples SC_HK_C- IndThrift et SC_HK_confi – IndThrift pour le degré 0.1MR. Nous expliquons ce phénomène par le fait que le nombre des PIs augmente en présence de bruit pour ces détecteurs. Pour ces nouveaux points, bien qu'ils peuvent correspondre à du bruit, leur voisinage est décrit d'une façon robuste au bruit. Le processus de mise en correspondance réussit à apparier les nouvelles paires de PIs.

- Pour nos détecteurs combinés, en augmentant le niveau du bruit, les courbes du descripteur IndThrift chutent moins rapidement que celles du IndSHOT. Par exemple, sur la courbe du détecteur SC_HK_C, nous passons de 86.6% à 88.9% puis à 77.8% alors qu'avec la courbe du IndSHOT, nous passons de 91.1% à 86.6% puis à 77.8%. Il est clair que le descripteur IndSHOT est plus performant pour des données non bruitées (91.1% vs 86.6%) comme nous l'avons démontré plus haut dans la section 5.2.2.1. Pour des plus hauts degrés de bruit (ex.0.5MR), la

performance de l'ensemble des systèmes chute considérablement. Cette limite ne fait pas partie des objectifs de notre travail actuel.

Selon le **scénario B1**, le descripteur IndThrift est plus robuste au bruit.

	IndSHOT	IndThrift	SURF
SC_HK_confi	77.0	**88.2**	
SC_HK_C	80.1	**83.2**	
SC_HK_FQ	73.3	**87.6**	
SI	**80.7**	75.8	
Harris_fract	83.9	**85.8**	
Harris_cluster	69.7	69.7	
SURF			**84.5**

Tableau 5-15 Résultats du scénario B2 pour un ajout de bruit de variance 0.2 MR sur 9 objets Minolta

- Dans les résultats du **scénario B2**, le descripteur IndThrift est plus performant que l'IndSHOT et sa combinaison avec le détecteur SC_HK_confi réalise un taux de 88.2%, ce qui suggère la robustesse de ce système à un faible bruit. Ce résultat est en accord avec le celui obtenu dans le scénario B1. Les taux sur ce scénario sont plus élevés que ceux du B1 puisque, de un, le protocole apprend sur plus de vues, et de deux, l'apprentissage et le test sont fait sur le même type de données (bruitées et non bruitées). En effet, si nous supposons que la vue en requête est bruitée alors ses PIs vont être apparié majoritairement avec les PIs des vues de même nature (bruitée) et inversement.
- Rappelons que le descripteur IndThrift, croise une information sur les angles des normales calculées sur deux rayons, avec une mesure de l'indice de forme, aux points du voisinage des PIs. Le calcul de ces normales est fait d'une façon robuste au bruit puisque la normale d'une zone de voisinage est résultante de la moyenne des normales des points de ce voisinage. D'autre part, le choix d'un rayon petit et d'un autre plus grand permet de décrire la variation locale d'une façon robuste à une petite perturbation des données. L'impact d'un faible bruit sera négligeable sur le cumul final des valeurs croisées. A ce stade, nous mentionnons que parmi les paramètres déterminant de la robustesse de notre système au bruit, le rayon choisi pour la détection et la description des PIs. Plus ce rayon est grand, plus l'impact du bruit est moins visible. Nous avons évalué l'influence du choix de ce rayon, seulement, sur la performance des descripteurs. Il serait judicieux, dans une étude plus approfondie, de voir l'effet de la variation du rayon sur le taux de reconnaissance. Pour une approche multi-échelle de notre système, la problématique du choix du rayon en description ne se pose pas vu qu'il est déterminé automatiquement suite au processus d'extraction des PIs. Pour cette raison, nous pensons que pour une approche multi échelle la robustesse au bruit augmentera.

➢ Ce résultat prometteur rejoint les résultats déjà obtenus dans notre évaluation des détecteurs et descripteurs. En effet, la robustesse des détecteurs combinés SC_HK_confi, SC_HK_C et SC_HK_FQ a été prouvé ainsi que la stabilité des descripteurs IndSHOT et IndThrift. A l'issu de ces résultats, nous déclarons que pour nos détecteurs composés, le descripteur IndThrift est plus robuste au bruit que l'IndSHOT.

5.2.3 Base Stuttgart

❖ **Scénario**

Dans cette partie, nous proposons de faire une comparaison avec les résultats de la littérature sur une base de données publique. En particulier, nous nous sommes référés au travail d'Aykut Hozatlı (Hozatlı, 2009) et d'Ömer Eskizara (Eskizara, 2009) qui utilisent pour leur détection de primitives, respectivement, un espace de classifications SI et un espace d'échelles à base de HK. Les descripteurs correspondent, pour le premier système (Hozatlı, 2009), à des splashs des angles entre les normales, et à des vecteurs formés par une information sur l'angle et la longueur géométrique entre les primitives pour le deuxième système (Eskizara, 2009). Dans ces deux systèmes, un hachage géométrique est appliqué aux primitives en phase de pré- classification. De plus, une série de test a été menée dans le but de déterminer les valeurs des seuils des paramètres de leur algorithme et conclure sur le paramétrage qui donne le plus haut taux de reconnaissance sur les objets de la base Stuttgart. Une évaluation de la reconnaissance est faite sur trois groupes d'objet de taille 5, 25 puis 42. Pour Ömer Eskizara, les 5 objets sont agfa, machine, igea, vette et pitbull. Nous notons cet ensemble E1. Aykut Hozatlı a choisi les objets : agfa, bunny, auto, balljoint et banane (Figure 5-11). Nous appelons cet ensemble E2.

Les 25 objets sélectionnés pour les tests sont dans l'ordre: 1.agfa, 2.balljoint, 3.bunny, 4.carrier, 5.chiken, 6.copter, 7.deoround, 8.dino, 9.ente, 10.fighter, 11.hip, 12.igea, 13.knot, 14.liberty, 15.pedaltopf, 16.pitt_plane, 17.rocker, 18.screwdriver, 19.scull, 20.seahorse, 21.Sfb, 22.teeth, 23.turbine, 24.vette, et 25.watch.

Nous rappelons que, dans notre travail, nous n'avons pas vraiment mené les tests nécessaires dans le but d'optimiser les paramètres de nos algorithmes et les adapter aux différentes bases testées. Notre objectif est d'avoir un système le plus générique possible et non d'ajuster nos paramètres selon une base donnée pour un meilleur fonctionnement, tel est le cas pour ces deux derniers travaux.

Le scénario de reconnaissance utilisé sur cette base est différent des protocoles utilisés dans les parties précédentes. Nous appliquons le même schéma que les auteurs de (Eskizara, 2009), en prenant en apprentissage 66 vues par objet et testant sur 258 vues dont les 66 prises en apprentissage. La comparaison est faite sur les résultats des bases avec les trois groupes d'objets.

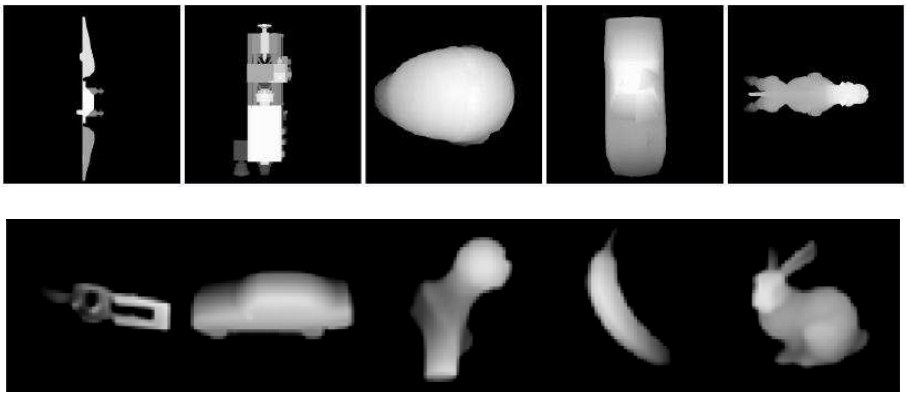


Figure 5-11 Les 5 objets sélectionnés de la base Stuttgart pour l'évaluation des primitives respectivement, de haut vers le bas, dans (Hozatlı, 2009) et (Eskizara, 2009)

Les résultats sont affichés sous forme de diagramme de Hinton qui est une représentation qualitative de la matrice de confusion. A chaque valeur d'un taux de reconnaissance est associé un carré de

couleur en niveau de gris avec un degré proportionnel à cette valeur (plus le carré est claire plus le taux est faible).

❖ **Résultats et analyses**

- Le couple SC_HK_C - IndSHOT donne le meilleur résultat parmi les 6 combinaisons détecteur-descripteur (Tableau 5-16). Ce résultat est en accord avec celui obtenu sur les 25 objets de la base Minolta.

% Reconnaissance	IndSHOT	IndThrift
SC_HK_confi	95.03	92.63
SC_HK_C	**95.35**	92.86
SC_HK_FQ	95.26	93.17

Tableau 5-16 Résultats de reconnaissance sur les 5 objets Stuttgart de l'ensemble E1

- D'après les diagrammes de la Figure 5-12, sur les cinq premiers objets, notre résultat est de 95.35% contre les taux 94,57% et 99.45% (Eskizara, 2009), et de 99.61% par rapport à 98,99% (Hozatlı, 2009), sur le deuxième groupe d'objets. Nous avons une légère amélioration par rapport à l'approche d'Aykut Hozatlı et nous nous situons entre les deux évaluations, primaires et améliorée, réalisées par Ömer Eskizara. Un exemple de mise en correspondances des primitives du modèle pitbull obtenu avec le système SC_HK_C – IndSHOT est illustré dans la Figure 5-13. Cette composition du détecteur-descripteur permet de bien apparier les paires des PIs grâce à leur pouvoir descriptif et leur robustesse à une variation d'angle.

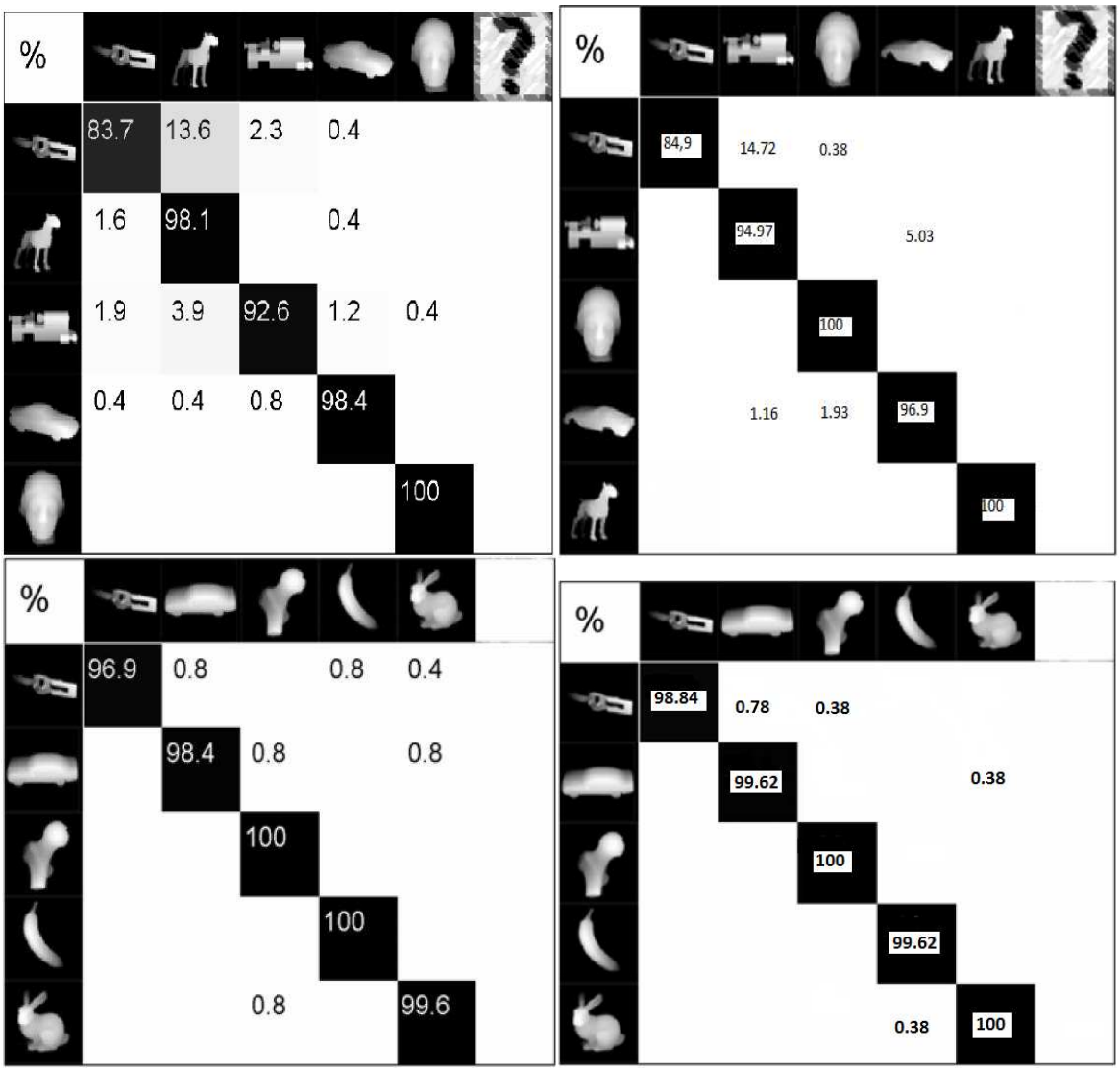


Figure 5-12 Diagrammes de Hinton pour les deux ensembles E1 et E2 de la base Stuttgart. Première ligne : à gauche l'approche HK (Eskizara, 2009) avec un taux primaire de 94,57% (taux amélioré pour atteindre 99.45%), et à droite, notre résultat avec 95.35%. Deuxième ligne, à gauche, l'approche SI (Hozatlı, 2009) avec un taux de 98,99%, et à droite, notre résultat avec 99.61%.

Nous illustrons, dans la Figure 5-14 et la Figure 5-15, les diagrammes de Hinton pour respectivement les 25 et 42 objets. En définitif, en augmentant le nombre d'objets de la base, le taux de reconnaissance diminue légèrement entre la taille 5 et 25 objets et reste quasiment inchangé pour les tailles 25 et 42.

- Sur le résultat des 25 objets, notre taux 94.93% est supérieur au taux 93.04% réalisé par Ömer Eskizara. D'après le diagramme de Hinton, tout nos taux par objet sont supérieurs à 81% et 9 objets ont un taux égal à 100%, alors que d'après leur diagramme (Eskizara, 2009), les trois objets : copter, pitt_plane et turbine ont les taux 66.3%, 78,7% et 80,6% et seulement 2 objets ont 100% de taux de bonne reconnaissance. La combinaison SC_HK_C – IndThrift, sur ces 25 objets, donne un taux de 91.15%. En plus de ce taux plus bas, nous dénonçons le cout du temps de calcul pour nos systèmes utilisant l'IndThrift.

- Il est à noter que les taux de reconnaissances des objets de types mécaniques, comme agfa (1), carrier (4), copter (6), scull (19) et vette (24), sont les plus faibles la base. Nous remarquons que

pour ces objets, les surfaces planes sont dominantes et les arêtes de leurs surfaces sont faiblement courbées. Or dans nos primitives, les descripteurs IndSHOT des PIs extraits sur des zones planes sont identiques partout et sont, pour cette raison, faussement appariés avec les PIs de même type sur d'autres objets (ex. l'objet agfa est apparié avec l'objet Sfb de 3.49%). Un processus de mise en correspondance plus contrôlée a été proposé englobant une information sur la distance spatiale entre les PIs. Nous n'avons pas présenté les résultats qui ont été médiocres. Nous pensons que cette piste est à creuser d'avantage en ajoutant plus de contraintes géométriques sur les PIs, et une information sur la texture au niveau du descripteur. Ces modifications amélioreront la performance de l'appariement pour les objets ayant des structures planes.

- En contre partie, notre système a plus de faciliter à reconnaitre les objets de type naturel dont le degré de courbures est plus prononcé. Il est clair que des surfaces sous forme de cuvette, dôme et selles sont plus détectables sur des objets comme le balljoint (2), bunny(3), ente (9) et igea (12) dont le taux est de 100% de bonne reconnaissance. La diversité de la nature des détails de la forme au sein d'un même objet crée une unicité dans la représentation donnée par nos primitives. En effet, d'une part, le nombre de PIs extraits est plus important et le descripteur IndSHOT réussit à caractériser d'une façon distincte la surface locale sur ces PIs. Le processus de mise en correspondance apparie correctement les PIs de ces objets.

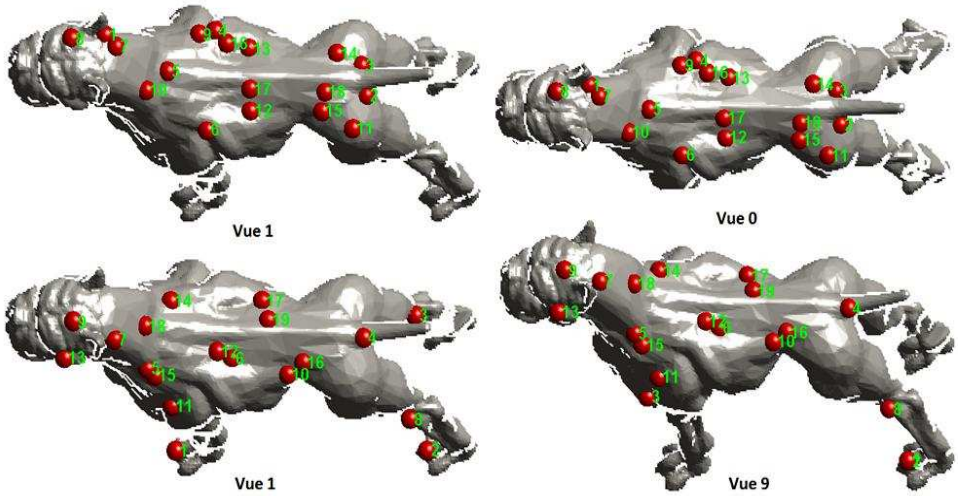


Figure 5-13 Exemples de mise en correspondance des PIs de l'objet pitbull entre la vue test n°1 (à gauche) et les deux vues d'apprentissage n°0 et n°9 (à droite)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	89.53	0.39					0.39		0.39	0.39				0.39	0.39			0.39			3.49		2.39	0.39	1.55
2		100																							
3			100																						
4	1.16		0.77	82.2								0.39			7.75		0.39		0.77		4.26	0.77	0.39	0.77	0.39
5					95.74				2.32				1.94												
6		1.55	2.32			87.6	1.55		3.87		1.93						0.77					0.39			
7							100																		
8		0.39						99.22														0.39			
9									100																
10		0.39								93.41		0.77			0.39						3.49			1.55	
11											100														
12												100													
13													100												
14														95.35	1.93						2.33			0.39	
15															96.12						3.39			0.39	
16															1.93	97.28							0.39		
17																	100								
18									1.55									98.45							
19															3.88				82.56		12.8			0.77	
20			0.77						1.94							0.39			0.39	96.52					
21															7.75						91.09			1.16	
22			0.77						1.16				0.39		0.39	0.39					1.94	94.58		0.39	
23																							100		
24												1.94			9.3						6.98			81.78	
25															5.43						2.32			0.39	91.86

Figure 5-14 Diagramme de Hinton pour les 25 modèles de la base Stuttgart (94.93%)

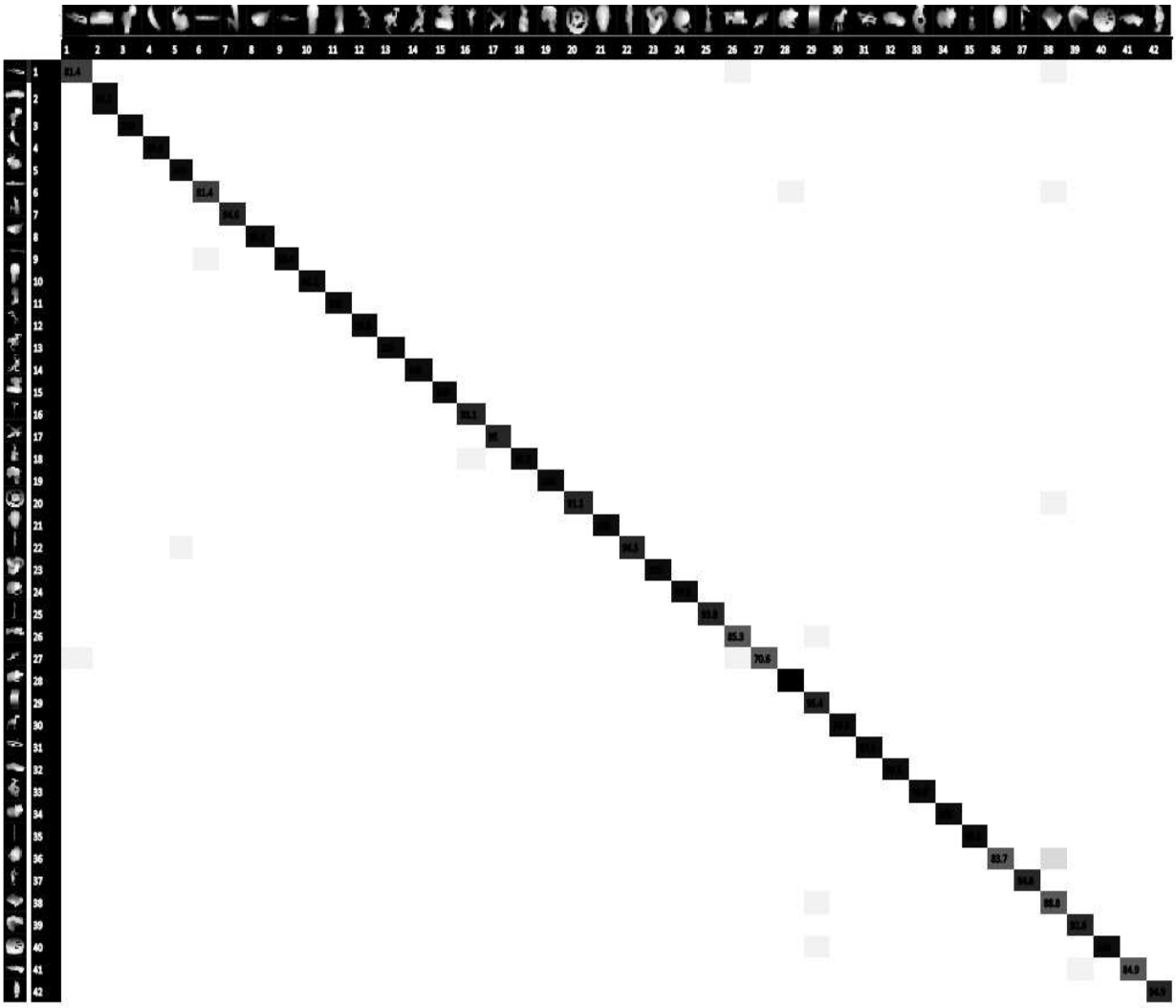


Figure 5-15 Diagramme de Hinton pour les 42 modèles de la base Stuttgart (94.90%)

- Le taux obtenu de **94.90**% sur les 42 objets est encourageant et améliore les taux de la littérature qui sont de 90,97% (Eskizara, 2009) et de 91.03% (Hozatlı, 2009) sur cette même base.

- Nous observons sur ce diagramme que 36 objets sont reconnu avec un taux supérieur à 90% et que 12 objets ont un taux égal à 100% (ex. balljoint, dinopet, mole, et turbine). L'extraction de primitives distinctes, saillantes et décrivant la forme, justifie ces taux élevés. Cependant, quand ces éléments saillants sont redondants sur plusieurs objets, le nombre de faux appariements entre les PIs augmente. Les modèles ayant les plus faibles taux de reconnaissance sont : agfa (81.4%), copter (83.34%), scull (83.73%), carrier (81.4%), manta (70.55%), machine (85,3%), Sfb (88.75%) et vette (84.89%). La caractéristique commune entre ces modèles est la forte présence de zones planes et de peu de variation locale de la forme. Comme expliqué plus haut pour la base des 25 objets, pour ces modèles, les primitives locales extraites ne sont pas suffisamment représentatives de l'unicité de leur forme dans la base. Notre approche étant locale, et ne tenant pas compte des relations relatives entre les PIs, engendre des confusions entre les zones locales de faible degré de courbure (zones planes ou quasi planes). Par exemple, les objets mécaniques agfa, machine et Sfb sont souvent appariés entre eux.

- D'un autre coté, la forte similarité entre quelques modèles de la base, tel le cas des modèles (auto, porsche et vette) fait hausser le nombre des fausses correspondances. Cette limite est légitime dans les cas où à l'œil nu, nous n'arrivons pas à différencier quelques objets entre eux comme l'exemple de l'objet deoflach et l'objet deoround.

Nous résumons les résultats de notre comparaison dans le tableau suivant :

Nombre d'objets	5		25	42
Méthode	E1	E2		
HK (Eskizara, 2009)	**99.45**		**93.04**	90,97
SI (Hozatlı, 2009)		98,99		91.03
Notre méthode SC_HK_C- IndSHOT	95.35	**99.61**	**94.93**	**94.90**

Tableau 5-17 Comparaison des taux de reconnaissance de notre méthode avec deux méthodes de la littérature

- En conclusion, avec le nouveau combiné **SC_HK_C- IndSHOT**, nous avons réalisé une performance qui excède celle de l'état de l'art sur une base de 42 objets. La particularité de cette base par rapport à la base Minolta, outre sa taille, consiste d'un coté en la diversité des prises de vues des objets autour de plusieurs axes de rotations dans l'espace, et d'un autre coté, en un apprentissage fait sur un nombre de vues égale à environ le tiers du nombre de vues prises en test. Ces composantes présentent un réel défi pour un système de reconnaissance.

5.2.4 Base Carotte

❖ **Scénario**

Rappelons que la base Carotte est constituée de 20 objets et acquise avec la caméra Kinect dans le cadre du projet Carotte. Les défis qui se présentent dans cette base concernent la faible résolution des données, le bruit présent dans les données et le faible nombre d'angles de vues par objet (3 à 15 vues). Une autre nouveauté de cette base est que ses objets ont des formes différentes de celles présentes dans les bases de données précédentes. De point de vue applicatif, ses modèles correspondent à des objets utilisés dans notre quotidien (chaise, bottes, bouteille, balles, …). De ce fait, il est intéressant d'évaluer la performance de notre système sur ce type d'objets pour une tache de reconnaissance d'objets domestiques.

Dans nos expérimentations, nous appliquons un filtrage bilatéral aux données de la kinect pour lisser les nuages de points bruités tout en conservant la forme de l'objet. Il en résulte que le degré de bruit est réduit sans être complètement éliminé. Le scénario de reconnaissance utilisé pour cette base est le scénario de rotation avec un nombre fixe égale à 6 de vues requêtes pour tous les objets sauf pour les objets 4, 13 et 17 (car leur nombre de vues est inférieur à 6).
Le descripteur utilisé est IndThrift, qui d'après notre évaluation 5.2.2.3, a montré une meilleur stabilité au bruit. Parmi nos trois détecteurs combinés, nous avons sélectionné SC_HK_FQ qui donne la meilleure performance sur ces données (nous ne présentons pas la comparaison des performances des trois détecteurs). Les résultats sont présentés dans le diagramme de la Figure 5-16 qui indique également le nombre de vue utilisées en apprentissage et en test pour chaque objet. La description du protocole de test donnée dans la colonne gauche de la figure montre que, par exemple pour l'objet 1, le taux de reconnaissance (100%) donné dans la case (1, 1) est la moyenne des taux de 6 tests, où à

chaque fois, le système apprend sur 7 vues de l'objet 1 et toutes les vues des autres objets et teste sur une vue de l'objet 1 différente de celles apprises.

❖ **Résultats et analyses**

Le taux de reconnaissance total obtenu sur les 20 objets est de 80%, qui est un résultat prometteur par rapport à la nature de la base. En analysant ce diagramme, quelques objets sont bien reconnus (100% de taux) comme la balle(1), le ventilateur(19) et le robot(20). L'explication est que ces objets se distinguent par une forme assez différente de celles des autres objets. Par ailleurs, le processus d'appariement confond les objets ayant une forte similarité de forme, par exemple les deux objets bottes (2 et 3) et les objets parallélépipédiques (boite à outils (5) - parallélépipède mauve (14) - livre (12)). Les objets avec des surfaces planes comme le cylindre vert, parallélépipède mauve (14) et livre (12) sont appariés, faussement, avec d'autres objets de même type. Il est clair que la similarité entre les modèles de cette base rend difficile la tache de reconnaissance. Compte tenu du formalisme de notre système, notamment ce qui concerne la mise en correspondance de primitives indépendamment de leurs positions relatives, ces confusions sont justifiables. A ce niveau, nous pouvons prédire que pour un exercice de catégorisation, notre système réussira à identifier la catégorie associée aux objets (ex. catégorie Botte, catégorie objets géométriques…).

Néanmoins, ces résultats restent encourageants quant au fait de reconnaitre des objets, à partir de peu de vues en apprentissage, sur une base de nature bruitée et ayant des formes similaires. Nous rappelons que le bruit de ces données est assez élevé, ce qui baisse la performance de notre détecteur et descripteur dont la sensibilité à des hauts degrés de bruit a été démontrée dans les chapitres détecteurs et descripteurs. Une amélioration de la qualité des données, par un prétraitement de filtrage plus poussé, peut hausser le taux de reconnaissance.

	Apprentissage	test
1	7	1X6
2	10	1X6
3	14	1X6
4	3	1X4
5	8	1X6
6	14	1X6
7	6	1X6
8	11	1X6
9	8	1X6
10	14	1X6
11	13	1X6
12	13	1X6
13	2	1X1
14	11	1X6
15	14	1X6
16	14	1X6
17	2	1X2
18	9	1X6
19	14	1X6
20	14	1X6

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	100																			
2		67	33																	
3		33	67																	
4	50			0				50												
5					50									50						
6						83	17													
7							100													
8								100												
9									33			67								
10										100										
11											100									
12								17				83								
13													100							
14												33		67						
15															100					
16																100				
17																	100			
18													33		17			50		
19																			100	
20																				100

Figure 5-16 Diagramme de Hinton pour les 20 objets de la base Carotte (80%) en utilisant SC_HK_FQ et IndThrift (à gauche- la description du protocole de test et à droite le diagramme de Hinton)

5.2.5 Base RGBD

❖ **Scénario**

De la base RGB-D, nous a avons sélectionné 37 objets de différentes formes, notés de 1 à 37 dans l'ordre suivant : 1.apple_1, 2.ball_1, 3.banana_1, 4.bell_peper_1, 5.binder_1, 6.calculator_1, 7.camera_1, 8.cap_1, 9.cell_phone_1, 10.cereal_box_3, 11.coffee_mug_1, 12.flashlight_1, 13.food_bag_1, 14.food_box_1, 15.food_can_1, 16.food_cup_1, 17.greens_1, 18.hand_towel_1, 19.instant_noodles_1, 20.Kleenex_1, 21.lemon_1, 22.lime_1, 23.notebook_1, 24.onion_1, 25.orange_1, 26.peach_1, 27.pear_1, 28.pitcher_1, 29.plate_1, 30.scissors_1, 31.shampoo_1, 32.soda_can_1, 33.sponge_1, 34.stapler_1, 35.tomato_1, 36.toothpaste_1 et 37.watter_bottle_1.

Un nouveau challenge s'ajoute à ceux présents dans la base Carotte, concernant les angles des prises de vues suite à des rotations autour de plusieurs axes de rotations dans l'espace 3D. Nous sélectionnons 30 vues par objet. Le scénario de reconnaissance est comme suit : pour un objet x, une vue est gardée en requête et ses 29 vues restantes avec toutes les vues des autres objets sont prises en apprentissage. Ce scénario est répété 10 fois pour calculer la moyenne des taux de reconnaissance par objet. Pareillement qu'avec la base Carotte, la combinaison SC_HK_FQ – IndThrift est utilisée pour une évaluation sur cette base. Egalement, un filtrage bilatéral est appliqué en phase de prétraitement sur les données.

❖ **Résultats et analyses**

Le diagramme de Hinton obtenu pour cette base est illustré dans la Figure 5-17.

Avec un taux de reconnaissance totale est de **78.10%** sur les 37 objets, ce résultat reste acceptable compte tenu des contraintes posées dans cette base.

Les mêmes explications que celles énumérées pour la base Carotte sont valables pour cette base. La similarité des formes entre les modèles de cette base est aussi un handicap pour notre système de reconnaissance, comme le cas de confusion entre l'objet ball_1 (2) et l'objet lemon_1 (21).

Il est à noter que dans cette base, certains objets sont de petite taille et ont pour ce fait une faible densité de points par rapport aux autres objets. Des exemples d'objets de petite taille sont le calculator_1 (6), le peach_1 (26) et le tomato_1 (35) dont les taux sont respectivement, 40%, 10% et 40%. Un nombre faible de points sur la surface d'un objet, n'est pas pris en considération dans le filtrage des points de nos algorithmes de traitement. En effet, nous prenons en compte seulement les points, ayant un nombre de voisinage supérieur à un seuil fixe sur toutes les base, et situés à une distance du bord supérieure au rayon du support. Avec toutes ces conditions, le détecteur sort peu de PIs pour ces objets et le taux d'erreur de reconnaissance est plus important comme le montre le diagramme de Hinton.

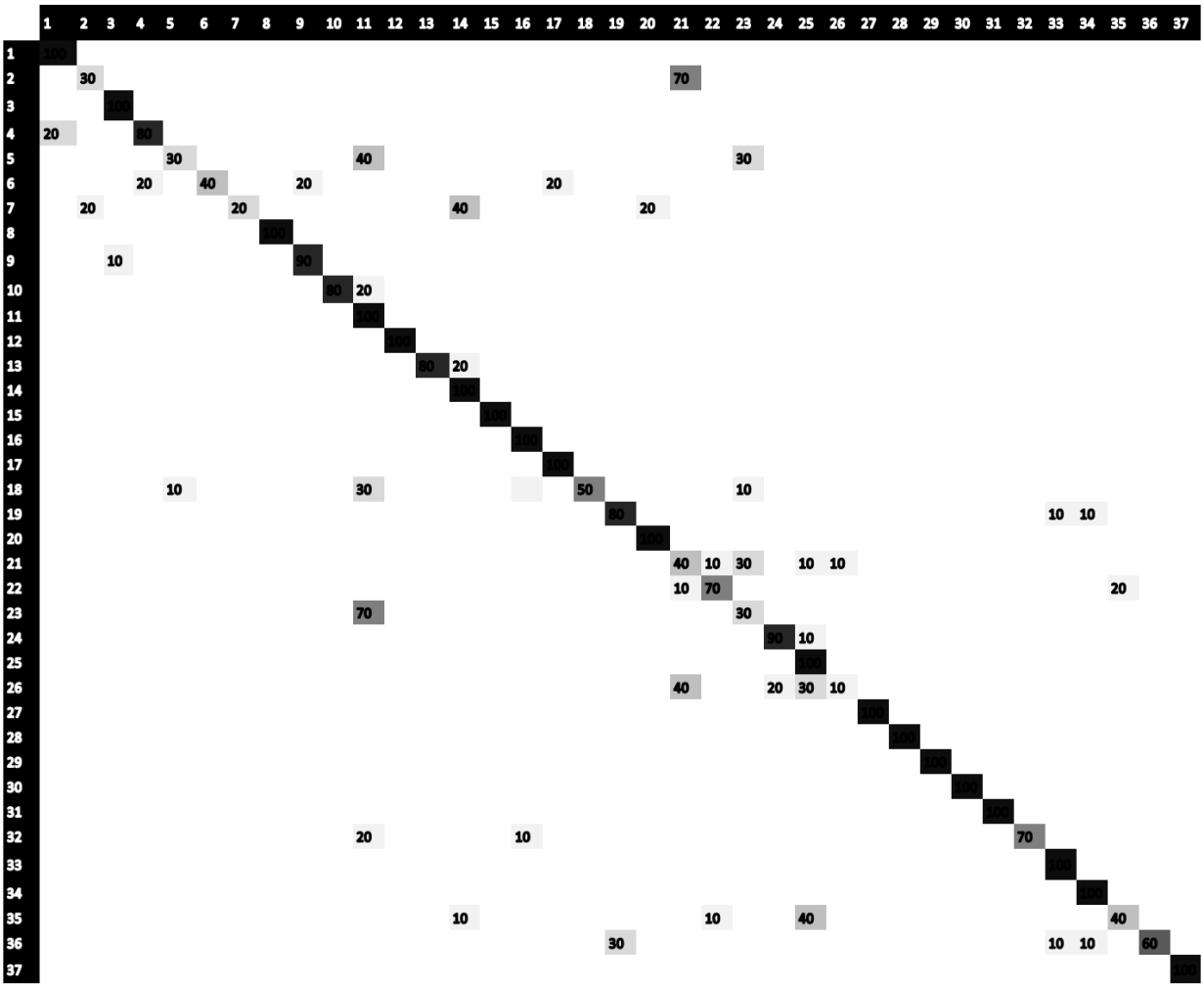


Figure 5-17 Diagramme de Hinton pour les 37 modèles de la base RGB-D (78.10%)

5.2.6 Conclusions

- Avec les nouvelles combinaisons (détecteur- descripteur) présentées, nous obtenons des meilleurs taux de reconnaissance que ceux de la littérature sur la base Stuttgart de 42 objets. Pour notre évaluation de la robustesse au bruit et de l'invariance aux changements d'échelles et d'angles de vues, ces combinaisons ont une performance de même ordre que celles de la littérature. Les expérimentations menées sur plusieurs bases de données (nature des objets, densité et qualité des données, type de capteur) ont montré l'efficacité et la faisabilité de nos algorithmes sur différent types de données.
- Dans notre analyse des différentes bases de données, les objets mécaniques avec des surfaces planes ont un faible taux de reconnaissance. L'explication donnée est que notre extraction, description et appariement des primitives sont plus adaptés à des objets « naturels » dont la surface se caractérise plus communément par des formes de cuvette, dôme et selles. Le processus de mise en correspondance est à enrichir par l'ajout de contraintes géométriques de positions relatives entre les PIs.
- En revanche, un aspect qui reste à optimiser pour notre approche est le temps de calcul de l'étape de détection. Le nombre de primitives sélectionnées est crucial pour la performance de la reconnaissance. Le temps de mise en correspondance est de l'ordre de 0.06s pour 150 PIs avec le

descripteur IndSHOT. Ce temps de calcul est influencé par le nombre des PIs extrait et reste insensible à la dimension des descripteurs utilisés, comme le montre le Tableau 5-18. Un autre facteur, intervenant légèrement sur le coût des calculs, est le nombre d'objet dans la base d'apprentissage (ex. le passage de 5 à 25 objets de la base Stuttgart avec 258 vues par objet, augmente le temps d'environ 0.05s par vue). Dans le Tableau 5-19, nous récapitulons le cout du processus total des phases de détection, description et mise en correspondance pour les couples (SC_HK_C, IndSHOT) et (SC_HK_C, IndThrift) et pour l'approche du SURF. Ces résultats montrent que l'exécution avec le couple SC_HK_C + IndSHOT reste comparable en temps de calcul avec celle du SURF.

- Il est à noter que ce temps est variable selon la nature des données de la base testée. En effet, par exemple, les données de la Kinect sont nettement moins denses que ceux de la base Stuttgart, ce qui diminue le temps de calcul des 3 phases de notre système. Pour une application contrainte à un temps d'exécution en temps quasi réel, nos approches restent faisables.

Dimension		**IndSHOT 768**	**SHOT 416**	**IndThrift 120**	**Spin 100**	**LSP 120**
Nombre de PIs	Nombre d'objets					
75	**5**	0.05	0.03	0.03	0.03	0.03
	25	0.09	0.08	0.08	0.08	0.08
250	**5**	0.18	0.11	0.12	0.11	0.11
	25	0.30	0 .29	0 .29	0 .29	0.30

Tableau 5-18 Temps de calcul (s) de la phase de mise en correspondance

Base : densité de points	Détection+description+appariement		
	Surf	SC_HK_C + IndSHOT	SC_HK_C + IndThrift
Stuttgart : 10948 points	**15.20**	**18.64**	**20.98**
Carotte : 4685 points	**10.13**	**9 .62**	**11.33**

Tableau 5-19 Temps de calcul total des phases de détection, description et mise en correspondance pour les couples (SC_HK_C, IndSHOT) et (SC_HK_C, IndThrift) et pour l'approche du SURF

Nous montrons, dans le Tableau 5-20, un résumé de la performance des systèmes combinés proposés. D'après ce récapitulatif, le couple SC_HK_C- IndSHOT est sélectionné pour répondre aux objectifs posés dans notre travail.

	Ordre de performance des Détecteur-descripteurs	Ordre de performance des Détecteurs	Ordre de performance des Descripteurs
Invariance à la rotation	1 SI-IndSHOT 2 SC_HK_C-IndSHOT	**1- SC_HK_C** 2- SC_HK_confi	**IndSHOT** IndSHOT
Invariance à l'échelle	1SC_HK_FQ-IndThrift 1 SC_HK_C-IndSHOT	**1- SC_HK_C** 1- SC_HK_FQ	**IndSHOT** IndThrift
Robustesse au bruit	1 SC_HK_C-IndThrift 2 Harris_fract-IndSHOT	**1- SC_HK_C** 1- SC_HK_FQ	**IndThrift** IndThrift
Temps de calcul	1 Harris_fract-IndSHOT 2 SC_HK_C-IndSHOT	**1- SC_HK_C** 2- SC_HK_confi 2- SC_HK_FQ	**IndSHOT** IndThrift

Tableau 5-20 Résumé sur la performance des systèmes proposés

- Dans notre étude, un compromis existe entre les trois composantes: nombre de points, complexité et représentation compact. Un nombre de PIs optimum (pas très faible et pas très grand) améliorera nettement la performance de notre système de reconnaissance.
- En outre, nous pensons que la version multi-échelle de notre détecteur apportera des améliorations sur la performance de notre système de reconnaissance. En effet, l'efficacité de l'attribution d'une information sur une échelle et un rayon adéquat à chaque PI, a déjà été démontrée par rapport à une approche mono-échelle, dans maints travaux de la littérature.

Conclusions et perspectives

En conclusion de ce manuscrit, nous avons pu accomplir les réalisations et tâches suivantes :

- Etat de l'art
  - Enumération (non exhaustive) des principales techniques 3D de la littérature touchant aux trois sujets sous-jacents (détecteurs, descripteurs et classification)
  - Brève description et analyse des atouts et limites de ces approches

- Détection
  - Privilégier une solution qui combine les critères et proposition d'une détection de points d'intérêt 3D, basée sur deux espaces de classification de la forme locale des surfaces (espace SC-HK)
  - Implémentation de quatre nouveaux détecteurs, SC_HK_C, SC_HK_ FQ, SC_HK_Conf et SC_HK_Con, basés sur des mesures de la courbure
  - Evaluation quantitative de la répétabilité et de la robustesse de ces détecteurs
  - Performance de SC_HK_C et SC_HK_ FQ, supérieure en changement d'angles de vue et comparable en variation d'échelles et en présence de bruit, en comparaison avec les détecteurs existants : SI, Harris 3D et SURF 3D
  - Proposition d'une extension de la méthode proposée pour une version multi-échelle

- Description :
  - Ajout du critère de l'indice de forme aux deux descripteurs existants SHOT et Thrift
  - Comparaison avec des descripteurs existants comme le Spin, LSP, SHOT et Thrift
  - Stabilité des nouveaux descripteurs IndSHOT et IndThrift pour une variation d'angles de vue et d'échelles
  - Pouvoir descriptif élevé, et robustesse pour des perturbations des données par du bruit ou un lissage, de ces descripteurs

- Reconnaissance :
  - Efficacité des nouvelles combinaisons (détecteur- descripteur) proposées pour une tâche de reconnaissance d'objet avec des meilleurs taux en comparaison avec des systèmes de la littérature
  - Faisabilité de nos algorithmes sur différent types de données d'images de profondeur
  - Meilleur taux de reconnaissance que ceux de la littérature sur la base Stuttgart de 42 objets

Du point de vue applicatif, notre technique a été testée dans le contexte d'une tâche de reconnaissance d'objet avec une Kinect par un robot mobile (projet CAROTTE). D'autres cadres applicatifs sont envisageables, puisque nous avons mis en place un système complet de reconnaissance d'objets 3D, comportant les étapes de détection de points d'intérêt, leur description et leur mise en correspondance sur des vues 3D de l'objet. Nous pensons que le champ d'applications, pour notre système, reste vaste.

A l'issue de ce travail, plusieurs perspectives s'ouvrent sur les divers sujets traités. Nous proposons d'énumérer les plus principales :

- Approfondir l'analyse et le développement de l'approche multi-échelle proposée
- Proposer une mesure qui permet de distinguer les positions des PIs sur des zones planes
- Optimiser le temps de calcul de l'étape calcul des quantités différentielles
- Raffiner l'étape de prétraitement pour des données bruitées
- Formuler un descripteur tridimensionnel en rajoutant une information sur la texture
- Tester l'approche Sac de mots pour la classification des PIs
- Utiliser les positions relatives des PIs dans le processus de mise en correspondance
- Rajouter un degré de confiance (utilisé dans notre approche mais pas avec la même signification) au résultat de reconnaissance, qui peut être calculé à partir du résultat du vote et de la distance des classes.

- Image illustrant quelques perspectives
  - Evaluation du système pour un scénario d'occultation
  - Evaluation du système pour un scénario de catégorisation
  - Intégration et validation des composantes implémentées dans un système d'application réelle (robot)

Et pour finir, nous suggérons une ouverture sur la dualité 2D/3D. Dans les applications où l'on souhaite reconnaitre des objets avec le maximum de précision, la dualité de la 3D et la 2D est un domaine à explorer pour pouvoir profiter des avantages complémentaires des deux représentations.

Liste des publications

Conférences internationales avec comité de relecture

- "3D Object Recognition and Person Facial Identification Using Time-averaged Single-views from Time-of-flight 3D Depth-Camera", Fabien Moutarde, Hui Ding and Ayet Shaiek, Eurographics Workshop on 3D Object Retrieval (**3DOR 2010**), Norrköping, Suède, 2 Mai 2010.

- "3D keypoint detectors and descriptors for 3D objects recognition with TOF camera", Ayet Shaiek and Fabien Moutarde, 3D Image Processing (**3DIP**'11) and Applications III (Conférence SPIE), San Francisco, USA, 26 Janvier 2011.

- "3D Keypoints Detection for Objects Recognition", Ayet Shaiek and Fabien Moutarde, International Conference on Image Processing, Computer Vision, and Pattern Recognition (**IPCV'12**), Las Vegas, USA, 19 Juillet 2012.

- "Fast 3D keypoints detector and descriptor for view-based 3D Objects Recognition", Ayet Shaiek and Fabien Moutarde, International Workshop on Depth Image Analysis (**WDIA 2012**) of 21st International Conference On Pattern Recognition (ICPR'2012), Tsukuba, Japan, 11 Novembre 2012.

Conférences nationales, et autres communications

- "*Détecteurs de points d'intérêt 3D basés sur la courbure*", Ayet Shaiek et Fabien Moutarde, COmpression et REprésentation des Signaux Audiovisuels (**CORESA 2012**), Lille, 24 et 25 Mai 2012

- "Détection et caractérisation en temps quasi-*réel de points d'intérêt 3D pour la* reconnaissance d'objets", Ayet Shaiek et Fabien Moutarde, présentation à la journée « Approches du suivi des caractéristiques dans des images et leurs applications à la navigation » du GDR ISIS, Paris, 13 Janvier 2012

Bibliographie

[En ligne] // CGAL. - 2010. - http://www.cgal.org/.

[En ligne] // VTK. - 2010. - http://www.vtk.org/.

[En ligne] // Stuttgart. - Janvier 2010. - http://range.informatik.uni-stuttgart.de.

[En ligne] // Minolta. - 2011. - http://cheepnis.cse.nd.edu/~flynn/3DDB/3DDB/RID/index.htm

[En ligne] // RGBD. - 2012. - http://www.cs.washington.edu/rgbd-dataset/.

Akagündüz E. et Ulusoy I. Scale and Orientation Invariant 3D Interest Point Extraction Using HK Curvatures [Article] // Computer Vision Workshops (ICCV Workshops). - 2009. - pp. 697 - 702.

Akagündüz E. 3D Object recognition using scale space of curvatures [Rapport]. - Middle East : Graduate School Of Natural Ana Applied Sciences Of Middle East Technical University, 2011.

Al-Osaimi FR, Bennamoun M. et Mian A. Interest-point Based Face Recognition from Range Images [Article] // British Machine Vision Conference (BMVC). - 2007.

Ashbrook A.P., Fisher R. B., Robertson C. et Werghi N. Finding surface correspondence for object recognition and registration using pairwise geometric histograms [Article] // European Conference on Computer Vision (ECCV'98). - 1998. - p. 674.

Atmosukarto I. et Shapiro Linda G. A Learning Approach to 3D Object Representation for Classification [Article] // Proceedings of the 2008 Joint IAPR International Workshop on Structural, Syntactic, and Statistical Pattern Recognition (SSPR & SPR '08). - 2008. - pp. 267-276.

Bay H., Tuytelaars T. et Gool L. Van Surf: Speeded up robust features [Article] // European Conference on Computer Vision (ECCV'06). - 2006. - pp. 404-417.

Bayramoglu Neslihan. et Aydın Alatan A. Shape Index SIFT: Range Image Recognition Using Local Features [Article] // International Conference on Pattern Recognition (ICPR 2010). - 2010. - pp. 352-355.

Besl P. J. et Jain R. C. Segmentation Through Variable-Order Surface Fitting [Revue] // IEEE Transactions on Pattern Analysis and Machine Intelligence. - 1988. - 2 : Vol. 10. - pp. 167-192.

Biasotti S., Giorgi D., Marini S., Spagnuolo M. et Falcidieno B., A comparison framework for 3D object classification methods [Article] // Multimedia Content Representation, Classification and Security (MRCS). - 2006. - pp. 314-321.

Biederman I. Recognition-by-components: A theory of human image understanding [Revue] // Psychological review. - 1987. - Vol. 94. - pp. 115-147.

Biosca J.M. et Lerma, J.L. Unsupervised robust planar segmentation of terrestrial laser scanner point clouds based on fuzzy clustering methods [Revue] // ISPRS Journal of Photogrammetry and Remote Sensing. - 2008. - Vol. 63. - 1.

Bowyer K.W., Chang K. et Flynn P. A survey of 3D and multi-modal 3D+ 2D face recognition [Rapport] : Technical Report / Notre Dame Department of Computer Science and Engineering . - 2004.

Bozkurt N., Halıcı U., Ulusoy İ. et Akagündüz E. 3D Data Processing for Enhancement of Face Scanner Data [Article] // Signal Processing and Application Conference SIU. - 2009.

Campbell R.J. et Flynn P.J. A survey of free-form object representation and recognition techniques [Article] // Computer Vision and Image Understanding. - 2001. - Vol. 81. - pp. 166-210.

Cantzler H. et Fisher R. Comparison of HK and SC Curvature Description Methods [Article] // Third International Conference on 3-D Digital Imaging and Modeling (3DIM '01). - 2001. - p. 285.

Castellani U., Fusiello A. et Murino V Sparse points matching by combining 3D mesh saliency with statistical descriptors [Article] // Computer Graphics Forum. - 2008. - 2 : Vol. 27.

Chang K., Bowyer K. et Flynn P. Face recognition using 2D and 3D facial data [Article] // ACM Workshop on Multimodal User Authentication. - 2003. - pp. 25-32.

Chaouch M. et Verroust-Blondet A. A new descriptor for 2D depth image indexing and 3D model retrieval [Article] // Proceedings of the International Conference on Image Processing (ICIP 07). - 2007. - pp. 373-376.

Chen H. et Bhanu B. 3D free-form object recognition in range images using local surface patches [Revue] // Pattern Recognition Letters. - 2007. - Vol. 28. - pp. 1252-1262.

Chen H. et Bhanu B. 3D free-form object recognition in range images using local surface patches [Article] // International Conference on pattern Recognition (ICPR). - Cambridge : [s.n.], August 2004.

Chen H. et Bhanu B. Efficient Recognition of Highly Similar 3D Objects in Range Images [Revue] // IEEE Trans. Pattern Anal. Mach. Intell. - 2009. - 1 : Vol. 31. - pp. 172-179.

Choksuriwong A., Laurent H. et Emile B. A Comparative Study of Objects Invariant Descriptor [Article] // ORASIS 2005 , journées francophones des jeunes chercheurs en vision par ordinateur. - 2005.

Cornuéjols A. Une nouvelle méthode d'apprentissage :Les SVM. Séparateurs à vaste marge. [Article] // Bulletin de l'AFIA. - juin 2002. - 14--23 : Vol. 51.

Diego I.M., Serrano Á., Conde C. et Cabello E. Face verification with a kernel fusion method [Revue] // Pattern Recognition Letters. - 2010.

Dorai C. et Jain A.K. Cosmos - a representation scheme for 3d free-form objects [Revue] // IEEE Trans. Pat.Anal.and Mach. Intel.. - 1997. - 10 : Vol. 19. - pp. 1115–1130.

Dufournaud Y., Schmid C. et Horaud R. Matching images with different resolutions [Article] // Computer Vision and Pattern Recognition (CVPR00). - 2000. - pp. 612-618.

Eskizara Ö. 3D Geometric Hashing Using Transform Invariant Features [Rapport]. - Middle East : Electrical and Electronics Engineering Department, Middle East Technical University, 2009.

Cornuéjols A. et Miclet L. Apprentissage par combinaison de décisions :le dopage (boosting).. - 2002.

Tombari F., Salti S. et Di Stefano L., A combined texture-shape descriptor for en-hanced 3D feature matching [Article] // International Conference on Image Processing (ICIP'11). - Brussels, Belgium : [s.n.], 11-14 September 2011.

Tombari F., Salti S. et Di Stefano L., Unique Signatures of Histograms for Local Surface Description [Article] // European Conference on Computer Vision (ECCV'10). - 2010.

Fischler M.A. et Bolles, R.C. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography [Revue] // Communications of the ACM. - 1981. - Vol. 24. - pp. 381--395.

Fleishman S., Drori I. et Cohen-Or D. Bilateral mesh denoising [Revue] // ACM Transactions on Graphics (TOG). - July 2003. - 3 : Vol. 22. - pp. 950-953 .

Fleury C. Présentation de M2RI - Module CTR " Le kd-Tree : une méthode de subdivision spatiale ".

Flint A., Dick A. et van den Hengel A. Thrift: Local 3D Structure Recognition [Revue] // Proc. Digital Image Computing: Techniques and Applications (DICTA'07). - 2007. - pp. 182–188.

Frome A., Huber D., Kolluri R., Biiulow T. et Malik J. Recognizing objects in range data using regional point descriptors [Article] // European Conference on Computer Vision(ECCV'04). - 2004. - pp. 224–237.

Harris C. et Stephens M. A combined corner and edge detector [Article] // In Alvey vision conference. - 1988. - Vol. 15. - pp. 147–152.

Hebert M., Ikeuchi K. et Delingette H. A spherical representation for recognition of free-form surfaces [Revue] // IEEE Transactions on Pattern Analysis and Machine Intelligence. - 1995. - Vol. 17. - pp. 681-690.

Hemant M. Kakde. Range Searching using Kd Tree. - 2005.

Hetzel G., Leibe B., Levi P. et Schiele B. 3D Object Recognition from Range Images using Local Feature Histograms [Conférence] // Proceedings of Computer Vision and Pattern Recognition (CVPR'01). - 2001. - pp. 394-399.

Hilaga M., Shinagawa Y., Kohmura T. et Kunii T. Topology matching for fully automatic similarity estimation of 3D shapes [Article] // Proceedings of the 28th annual conference on Computer graphics and interactive techniques. - 2001. - pp. 203-212.

Ho H.T. et Gibbins D. A Curvature-based Approach for Multi-scale Feature Extraction from 3D Meshes and Unstructured Point Clouds [Revue] // Iet Computer Vision. - 2009. - 4 : Vol. 3.

Horn K. P. Extended gaussian images [Revue] // Proceedings of the IEEE . - 1984. - 12 : Vol. 72. - pp. 1671-1686.

Hozatlı A. 3D Object Recognition by Geometric Hashing for Robotics Applications [Rapport]. - Middle East : Electrical and Electronics Engineering Department, Middle East Technical University, 2009.

Johnson A. et Hebert M. Surface Matching for Object Recognition in Complex Three-Dimensional Scenes [Revue] // Image and Vision Computing. - 1998. - Vol. 16. - pp. 635-651.

Johnson AE et Hebert M. Using spin images for efficient object recognition in cluttered 3Dscenes [Revue] // IEEE Transactions on Pattern Analysis and Machine Intelligence. - 1999. - Vol. 21. - pp. 433-449.

Jones M.J. et Viola P. Face recognition using boosted local features [Article] // Proceedings of International Conference on Computer Vision (ICCV'03). - 2003.

Jones S. et Van Rijsbergen C.J. Information retrieval test collections [Revue] // Journal of Documentation 32. - 1976. - pp. 59–75.

Kadir T., Zisserman A. et Brady M. An affine invariant salient region detector [Article] // European Conference on Computer Vision (ECCV'04). - 2004. - pp. 228-241.

Kalogerakis E., Nowrouzezahrai D., Simari P. et Singh K. Extracting lines of curvature from noisy point clouds [Article] // Computer-Aided Design. - 2009. - Vol. 41. - pp. 282-292.

Kalogerakis E., Simari P., Nowrouzezahrai D. et Singh K. Robust statistical estimation of curvature on discretized surfaces [Article] // Proceedings of the fifth Eurographics Symposium on Geometry Processing Eurographics Association. - 2007. - p. 22.

Kazhdan M. et Funkhouser T. Harmonic 3d shape matching [Article] // ACM SIGGRAPH 2002 Technical Sketch. - 2002. - p. 191.

Ke Y. et Sukthankar R. PCA-SIFT: A More Distinctive Representation for Local Image Descriptors [Article] // Computer Vision and Pattern Recognition (CVPR). - 2004. - Vol. 2. - pp. 506-513.

Knopp J., Prasad M., Willems G., Timofte R. et Van Gool L. Hough transform and 3D SURF for robust three dimensional classification [Article] // European Conference on Computer Vision (ECCV'10). - 2010. - pp. 589-602.

Koenderink J. et Doorn A. J. Surface shape and curvature scale [Revue] // Image Vis. Comput.. - 1992. - 8 : Vol. 10. - pp. 557–565.

Li X. et Guskov I. 3D object recognition from range images using pyramid matching [Article] // IEEE 11th International Conference on Computer Vision. - 2007. - pp. 1-6.

Li X. et Guskov I. Multi-scale features for approximate alignment of point-based surfaces [Article] // Proceedings of the third Eurographics symposium on Geometry processing. - 2005. - p. 217.

Li Y. et Hancock E. R. Face Recognition using Shading-Based Curvature Attributes [Article] // Proceedings of the 17th International Conference on Pattern Recognition (ICPR 2004). - 2004. - Vol. 3. - pp. 538 - 541.

Lo T.-W. R. et Siebert J. P. Local feature extraction and matching on range images: 2.5D SIFT [Revue] // Computer Vision and Image Understanding. - 2009. - Vol. 113. - pp. 1235-1250.

Lo T.W.R. et Siebert J.P. SIFT keypoint descriptors for range image analysis [Revue] // Annals of the BMVA. - 2008. - 3 : Vol. 2008. - pp. 1–18.

Loménie N., Gallo L., Cambou N. et Stamon G. Structuration plane d'un nuage de points 3D non structuré et détection des zones d'obstacle [Article] // Vision Interface. - 1999. - pp. 164-171.

Lowe D.G. Distinctive image features from scale-invariant keypoints [Revue] // International journal of computer vision. - 2004. - Vol. 60. - pp. 91-110.

Maes C., Fabry T., Keustermans J., Smeets D., Suetens P. et Vandermeulen D. Feature detection on 3D face surfaces for pose normalisation and recognition [Article] // IEEE International Conference on Biometrics : Theory Applications and Systems (BTAS). - 2010.

Mayo M. et Zhang E. 3D Face Recognition Using Multiview Keypoint Matching [Article] // Proceedings of the Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS '09). - 2009. - pp. 290-295.

Medioni G.G. et François A.R.J. 3-D structures for generic object recognition [Article] // Computer Vision and Image Analysis. - 2000. - Vol. 1. - pp. 30-37.

Mian A., Bennamoun M. et Owens R. Face recognition using 2D and 3D multimodal local features [Article] // Advances in Visual Computing. - 2006. - pp. 860-870.

Mian A., Bennamoun M. et Owens R. On the Repeatability and Quality of Keypoints for Local Feature-based 3D Object Retrieval from Cluttered Scenes [Revue] // International Journal of Computer Vision. - 2009. - pp. 1-14.

Mian A.S., Bennamoun M. et Owens R. Keypoint detection and local feature matching for textured 3D face recognition [Revue] // International Journal of Computer Vision (IJCV). - 2008. - Vol. 79. - pp. 1-12.

Mikolajczyk K. et Schmid C. A performance evaluation of local descriptors [Revue] // IEEE Transactions on Pattern Analysis and Machine Intelligence. - 2005. - 10 : Vol. 27. - pp. 1615–1630.

Mikolajczyk K. et Schmid C. Scale & affine invariant interest point detectors [Revue] // International Journal of Computer Vision (IJCV). - 2004. - Vol. 60. - pp. 63-86.

Moravec H.P. Towards automatic visual obstacle avoidance [Article] // Proceedings of the 5th International Joint Conference on Artificial Intelligence. - 1977. - Vol. 99. - p. 584.

Murase H. et Nayar S.K. Visual learning and recognition of 3-D objects from appearance [Revue] // International Journal of Computer Vision (IJCV). - 1995. - Vol. 14. - pp. 5-24.

Osada R., Funkhouser T., Chazelle B. et Dobkin D. Matching 3D Models with Shape Distributions [Article] // Proceedings of the International Conference on Shape Modeling and Applications (ICSMA). - 2001. - p. 154.

Plantié M., Roche M. et Dray, G. Un système de vote pour la classification de textes d'opinion [Article] // Proceedings of the 8th journées francophones d'Extraction et gestion des Connaissances. - Spohia -Antipolis : France : [s.n.], 2008.

Rabin J., Delon J. et Gousseau Y. Mise en correspondance de descripteurs géométriques locaux par méthode a contrario [Article] // 21° Colloque GRETSI sur le traitement du signal et des images. - 2007. - pp. 11-14.

Ramaswamy S., Rastogi R. et Shim K. Efficient algorithms for mining outliers from large data set [Article] // International Conference on Management of Data (SIGMOD). - 2000. - pp. 427–438.

Roth P.M. et Winter M. Survey of appearance-based methods for object recognition [Rapport] : Technical Rapport / Institute for Computer Graphics and Vision, Graz University of Technology. - Austria : [s.n.], 2008.

Rudinac M., Lenseigne B. et Jonker P. Keypoint extraction and selection for object recognition [Article] // IAPR Conference on Machine Vision Applications (MVA2009). - 2009.

Samir C., Daoudi M. et Srivastava A. Reconnaissance de Visages 3D Utilisant l'Analyse de Formes des Courbes Faciales [Article] // 10èmes Journées CORESA (COmpression et REprésentation des Signaux Audiovisuels). - 2006. - pp. 9-10.

Schmid C., Mohr R. et Bauckhage C. Evaluation of Interest Point Detectors [Revue] // International Journal of Computer Vision (IJCV'00). - 2000. - 2 : Vol. 37. - pp. 151–172.

Scovanner P., Ali S. et Shah M. A 3-dimensional sift descriptor and its application to action recognition [Article] // Proceedings of the 15th international conference on Multimedia. - 2007. - p. 360.

Sipiran I. et BUSTOS B. A robust 3D interest points detector based on Harris operator [Article] // Eurographics Workshop on 3D Object Retrieval. - 2010. - pp. 7–14.

Song W., Liu L. et Yang X. Fast and Robust Ridge-Ravine Detection on Triangular Meshes by Smooth Parametric Functions [Article] // Proc. Pacific Graphics. - 2005. - pp. 109-111.

Stricker M. Réseaux de neurones pour le traitement automatique du langage : conception et réalisatiion de filtres d'informations [Rapport]. - Paris : Université Pierre et Marie Curie - Paris VI, 2000.

Sun J., Ovsjanikov M. et Guibas L. A concise and provably informative multi-scale signature based on heat diffusion [Article] // Eurographics Symp. Geom. Processing. - 2009. - pp. 197-206.

Swadzba A. et Wachsmuth S. Categorizing Perceptions of Indoor Rooms Using 3D Features [Article] // Proceedings of the 2008 Joint IAPR International Workshop on Structural, Syntactic, and Statistical Pattern Recognition. - 2008. - pp. 734-744.

Tangelder JWH. et Veltkamp RC. Polyhedral model retrieval using weighted point sets [Article] // Shape Modeling International (SMI'03). - 2003. - pp. 119-129.

Ho H. T. et Gibbins D. A Curvature-based Approach for Multi-scale Feature Extraction from 3D Meshes and Unstructured Point Clouds [Revue] // Computer vision, IET. - 2009. - 4 : Vol. 3. - pp. 201-212.

Tsalakanidou F., Tzovaras D. et Strintzis M.G. Use of depth and colour eigenfaces for face recognition [Revue] // Pattern Recognition Letters (PRL). - 2003. - 9-10 : Vol. 24. - pp. 1427-1435.

Unnikrishnan R. et Hebert M. Multi-scale interest regions from unorganized point clouds [Article] // Computer Vision and Pattern Recognition Workshop on S3D (Search in3D). - 2008.

Vranic D.V. et Saupe D. 3D model retrieval [Article] // Proc. Spring Conference on Computer Graphics and its Applications (SCCG2000). - 2000. - pp. 3-6.

Vranic D.V. et Saupe D. 3D shape descriptor based on 3D Fourier transform [Conférence] // Proceedings of the EURASIP Conference on Digital Signal Processing for Multimedia Communications and Services (ECMCS 2001). - 2001. - pp. 271-274.

Willems G., Tuytelaars T. et Van Gool L. An efficient dense and scale-invariant spatio-temporal interest point detector [Article] // European Conference on Computer Vision (ECCV'08). - 2008. - pp. 650-663.

Zaharescu A., Boyer E. et Varanasi K. Surface feature detection and description with applications to mesh matching [Article] // Computer Vision and Pattern Recognition (CVPR'09). - 2009. - pp. 373-380.

Zaharia T. et Prêteux F. Indexation de maillages 3D par descripteurs de forme [Article] // Actes 13eme Congres Francophone AFRIF-AFIA Reconnaissance des Formes et Intelligence Artificielle (RFIA'02). - 2002. - pp. 48-57.

Zhao W., Chellappa R., Rosenfeld A. et Phillips P.J. Face recognition: A literature survey [Revue] // ACM Computing Surveys (CSUR). - 2003. - 4 : Vol. 35. - pp. 399-458.

Zhong Y. Intrinsic shape signatures: A shape descriptor for 3d object recognition [Article] // in International Conference on Computer Vision Workshop on 3D Representation for Recognition(3DRR'09). - 2009.

Reconnaissance d'objets 3D par points d'intérêt

RESUME : Soutenue par les progrès récents et rapides des techniques d'acquisition 3D, la reconnaissance d'objets 3D a suscité de nombreux efforts de recherche durant ces dernières années. Cependant, il reste à résoudre dans ce domaine plusieurs problématiques liées à la grande quantité d'information, à l'invariance à l'échelle et à l'angle de vue, aux occlusions et à la robustesse au bruit.

Dans ce contexte, notre objectif est de reconnaitre un objet 3D isolé donné dans une vue requête, à partir d'une base d'apprentissage contenant quelques vues de cet objet. Notre idée est de formuler une méthodologie locale qui combine des aspects d'approches existantes et apporte une amélioration sur la performance de la reconnaissance.

Nous avons opté pour une méthode par points d'intérêt (PIs) fondée sur des mesures de la variation locale de la forme. Notre sélection de points saillants est basée sur la combinaison de deux espaces de classification de surfaces : l'espace SC (indice de forme- intensité de courbure), et l'espace HK (courbure moyenne-courbure gaussienne).

Dans la phase de description de l'ensemble des points extraits, nous proposons une signature d'histogrammes, qui joint une information sur la relation entre la normale du point référence et les normales des points voisins, avec une information sur les valeurs de l'indice de forme de ce voisinage. Les expérimentations menées ont permis d'évaluer quantitativement la stabilité et la robustesse de ces nouveaux détecteurs et descripteurs.

Finalement nous évaluons, sur plusieurs bases publiques d'objets 3D, le taux de reconnaissance atteint par notre méthode, qui montre des performances supérieures aux techniques existantes.

Mots clés : Reconnaissance 3D, Points d'intérêt, Images de Profondeurs, Nuages de Points 3D, Détecteurs, Descripteurs, Objet, Courbures, Normales.

3D object recognition with points of interest

ABSTRACT : There has been strong research interest in 3D object recognition over the last decade, due to the promising reliability of the 3D acquisition techniques. 3D recognition, however, conveys several issues related to the amount of information, to scales and viewpoints variation, to occlusions and to noise.

In this context, our objective is to recognize an isolated object given in a request view, from a training database containing some views of this object. Our idea is to propose a local method that combines some existent approaches in order to improve recognition performance.

We opted for an interest points (IPs) method based on local shape variation measures. Our selection of salient points is done by the combination of two surface classification spaces: the SC space (Shape Index-Curvedness), and the HK space (Mean curvature- Gaussian curvature).

In description phase of the extracted set of points, we propose a histogram based signature, in which we join information about the relationship between the reference point normal and normals of its neighbors, with information about the shape index values of this neighborhood. Performed experiments allowed us to evaluate quantitatively the stability and the robustness of the new proposed detectors and descriptors.

Finally we evaluate, on several public 3D objects databases, the recognition rate attained by our method, which outperforms existing techniques on same databases.

Keywords : 3D Recognition, Interest Points, Range Images, 3D Point Clouds, Detectors, Descriptors, Object, Curvatures, Normals.

MINES ParisTech

ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY