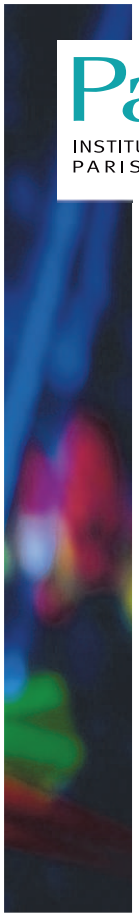
Fine-grained object categorization : plant species identification

Asma Rejeb Sfar



ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUT OF TECHNOLOGY

TELECOM ParisTech

2014-ENST-0046

EDITE - ED 130

Doctorat ParisTech

T H È S E

pour obtenir le grade de docteur délivré par

TELECOM ParisTech

Spécialité Informatique et Réseaux

présentée et soutenue publiquement par

Asma REJEB SFAR

10/07/2014

Fine-Grained Object Categorization: Plant Species Identification

Directeurs de thèse: **Nozha BOUJEMAA** et **Donald GEMAN**

Jury

M. Alain TROUVÉ, Professeur, ENS Cachan	Rapporteur
M. François FLEURET, Chargé de recherche (HDR), Idiap Research Institute	Rapporteur
Mme Françoise PRÉTEUX, Adjointe au directeur de la recherche, Mines ParisTech	Examinateur
M. Benoît POCHON, Responsable de l'équipe R&D en multimédia, Parrot	Examinateur
M. Hichem SAHBI, Chargé de recherche (HDR), LTCI Lab Télécom ParisTech	Examinateur
M. Donald GEMAN, Professeur, Johns Hopkins University	Examinateur
Mme Nozha BOUJEMAA, Directrice de INRIA Saclay Ile de France	Examinateur

TELECOM ParisTech
école de l'Institut Mines-Télécom - membre de ParisTech
46 rue Barrault 75013 Paris - (+33) 1 45 81 77 77 - www.telecom-paristech.fr

THÈSE

Acknowledgements

First I would like to thank Dr. Nozha Boujemaa who gave me the opportunity to join the iMedia team under her guidance and especially to work with Professor Donald Geman.

I am really grateful to Professor Geman for guiding me deeply during my thesis, for his consistent support and continuous encouragement. This thesis would not have been possible without him. Exchanges of ideas and fruitful discussions with Don were the key success factors of this work. For me, he was the perfect advisor both humanly and scientifically.

I am also grateful to the members of my oral defense committee : Alain Trouvé, François Fleuret, Françoise Préteux, Benoît Pochon and Hichem Sahbi, for their time and insightful questions.

I would like to thank my parents, my sister and my brother for their love, emotional support and eternal encouragements. To them, I dedicate this thesis. Also, words fail me to express my gratitude to my husband Mohamed Riadh TRAD whose patience, guidance and persistent confidence in me, have taken the load off my shoulder.

Finally, thanks are due to all my friends and everyone who was a member of the IMedia Team.

ABSTRACT

Fine-Grained Object Categorization: Plant Species Identification

We introduce models for fine-grained categorization, focusing on determining botanical species from leaf images. Images with both uniform and cluttered background are considered and several identification scenarios are presented, including different levels of human participation. Both feature extraction and classification algorithms are investigated.

We first leverage domain knowledge from botany to build a hierarchical representation of leaves based on *IdKeys*, which encode invariable characteristics, and refer to geometric properties (i.e., landmarks) and groups of species (e.g., taxonomic categories). The main idea is to sequentially refine the object description and thus narrow down the set of candidates during the identification task. We also introduce *vantage feature frames* as a more generic object representation and a mechanism for focusing attention around several vantage points (*where to look*) and learning dedicated features (*what to compute*).

Based on an underlying coarse-to-fine hierarchy, categorization then proceeds from coarse-grained to fine-grained using local classifiers which are based on likelihood ratios. Motivated by applications, we also introduce on a new approach and performance criterion: report a subset of species whose expected size is minimized subject to containing the true species with high probability. The approach is model-based and outputs a *confidence set* in analogy with confidence intervals in classical statistics. All methods are illustrated on multiple leaf datasets with comparisons to existing methods.

Résumé

1 Introduction

La vision par ordinateur est la science permettant de doter les machines de la capacité à "comprendre" et à interpréter le contenu d'une image au niveau sémantique. La vision par ordinateur couvre les technologies de base de l'analyse automatique d'images qui est utilisé dans de nombreux domaines. Les applications vont du traitement d'images de bas niveau à l'annotation automatique d'images, l'interprétation de scènes et la reconnaissance d'objets.

Dans le cadre de cette thèse, nous nous intéressons à la reconnaissance automatique d'objets, plus précisément aux objets appartenant à un même concept de base telle que la reconnaissance des différentes espèces de plantes via des images de feuilles. Distinguer les sous-concepts d'objets, peut s'avérer extrêmement difficile pour un "simple" utilisateur et nécessite souvent l'avis d'un expert, contrairement à la distinction de concepts différents (voiture, piéton, animal). Cette problématique de reconnaissance qui aborde l'espace sémantique dans une résolution très fine est un domaine d'étude de plus en plus actif, guidé par les applications ainsi que le défi intellectuel. Nous nous intéressons, en particulier, au domaine de la botanique. Nous étudions toute la chaine d'identification, notamment la représentation et l'extraction des caractéristiques discriminatives de l'objet, l'apprentissage et la recherche. Nous introduisons de nouveaux mécanismes de représentation de l'objet ainsi que de nouvelles méthodes de classification. Plusieurs scénarios semi-automatiques sont proposés à l'utilisateur pour l'identification d'espèces de plantes via différents types d'images de feuilles.

2 Motivations

L'un des défis majeurs dans le domaine de l'intelligence artificielle et la vision par ordinateur est de pouvoir doter les systèmes informatiques de la capacité humaine à discerner des objets visuellement similaires. Cependant, contrairement à la reconnaissance de concepts assez différents (tels que la reconnaissance des voitures ou des personnes parmi d'autres objets), même les humains pourraient avoir des difficultés dans la reconnaissance de (sous)catégories très similaires appartenant à un même concept de base (telles que les espèces de plantes ou les races de chiens). En effet, de telles catégories sont généralement reconnues par des spécialistes. Dans le cas des espèces botaniques, la reconnaissance est d'autant plus difficile que les botanistes se servent souvent de guides d'identification contenant différentes illustrations et propriétés végétales afin de faciliter la tâche de l'identification.

Souvent, les taxonomistes présentent des descriptions écrites et organisées des espèces similaires de sorte que d'autres biologistes peuvent identifier les spécimens inconnus. En botanique, cette tâche implique des comparaisons entre les caractéristiques déjà enregistrées et celles observées, puis l'attribution d'une plante à un rang taxonomique particulier (ex. famille, genre), pour finalement arriver au niveau des espèces. Toutefois , en raison de la grande variation des caractéristiques et le très grand nombre de catégories de plantes actuellement connues dans le monde (plus de 200, 000 espèces), l'identification des espèces botaniques peut être une tâche ardue et de longue haleine, même pour les experts. Trouver une espèce assez rapidement nécessite souvent de connaître à l'avance le nom de la famille ou du genre concerné. Cependant, un spécialiste dans un genre (ou une famille) particulier(ère) peut ne pas être familier avec un(e) autre. La difficulté est de plus en plus importante avec la pénurie actuelle de taxonomistes qualifiés.

Un système d'identification automatique (ou semi-automatique) pourrait donc être d'une utilité particulière dans de nombreuses applications. Les outils automatisés peuvent accélérer la reconnaissance et ainsi aider les spécialistes à reconnaître ou découvrir de nouvelles espèces. Dans le cas de la botanique, accélérer la collecte et la classification des observations est une étape cruciale vers le développement durable de l'agriculture et la conservation de la biodiversité.

3 Challenges

Dans ce qui suit, nous nous intéressons à des objets appartenant à un même concept, notamment les feuilles de plantes. Bien que de tels objets, pour la plus part, sont visuellement très similaires (même couleur, forme générale, système de nervures..), ils peuvent néanmoins largement différer au sein d'une même catégorie (espèce). Pis encore, il peut même y avoir moins de variation d'aspect entre deux images de deux catégories différentes qu'au sein d'une même catégorie. Et bien que la forme globale s'avère suffisamment discriminante entre certaines catégories, d'autres catégories affichent des variations fines uniquement.

En particulier, les feuilles peuvent présenter des apparences totalement différentes dues au contexte local, tels que l'emplacement et les conditions climatiques. Par exemple, les feuilles ont souvent des couleurs et des formes plus ou moins différentes tout au long des saisons. Une feuille morte a une texture très différente d'une en vie. Les mêmes espèces peuvent présenter des caractéristiques assez différentes dans deux régions différentes. Aussi, les feuilles d'une même tige, peuvent changer graduellement de forme avec l'âge; un phénomène appelé *développement hétéroblastique*. Ainsi, les feuilles dites *juvéniles* issues d'une jeune plante, sont presque toujours différentes des feuilles dites *adultes*.

Le fond des images peut également représenter un problème clé pour un système de reconnaissance automatique. Les images de feuilles sur fond naturel constituent, par exemple, un véritable défi pour la reconnaissance automatique et des algorithmes de segmentation. Le fait que lobjet d'intérêt soit situé au sein d'un environnement composite, dans lequel on risque de retrouver des objets de couleurs très proches (feuillage, autres plantes), rend par exemple totalement inopérantes les méthodes globales comme le seuillage pour séparer l'objet d'intérêt du fond. Dans l'ensemble, les conditions d'acquisitions sont peu maîtrisées, ce qui peut produire des photographies difficiles à traiter.

Dans la littérature, la plus part des travaux sur l'identification ou la classification automatique des plantes, ont eu recours à des images avec fond uniforme. D'ailleurs, plusieurs bases publiques sont constituées de scans de feuilles sur fond blanc. Cependant, des bases plus récentes contiennent de plus en plus d' images de feuilles avec fond naturel.

4 Positionnement par rapport à l'état-de-l'art

En raison des besoins et des défis mentionnés ci-dessus, de plus en plus de travaux en vision se consacrent à la catégorisation dite "fine", notamment la classification de (sous)catégories. Plusieurs méthodes ont été adaptées ou introduites pour discriminer des objets similaires et plusieurs approches ont eu recours à l'intervention humaine [13, 109, 33] pour améliorer les performances.

Généralement, le scénario standard était de fournir à l'utilisateur une estimation unique de la catégorie recherchée [35, 45, 119, 68, 117, 78, 13, 2]. Plusieurs autres travaux ont choisi de retourner les k premières catégories les plus similaires [7, 67, 75], où k, généralement, varie de dix à vingt afin de garantir un niveau de précision élevé. Bien sûr, on peut toujours retenir la bonne catégorie en retournant relativement un large ensemble d'estimations; ce qui a peu de valeur dans les applications réelles. En outre, se concentrer uniquement sur la liste des toutes premières estimations pourrait être complètement inutile à l'utilisateur, surtout si on ne garantit pas que la bonne réponse y figure.

Ici, nous nous concentrons sur (1) des stratégies hiérarchiques qui réduisent considérablement l'espace des candidats, (2) une description locale discriminante des différentes (sous)catégories, (3) un nouveau critère de performance: retenir le plus petit sous-ensemble possible d'espèces qui contiendrait le bonne estimation avec une probabilité élevée. En particulier, nous avons pu surpasser les performances d'autres méthodes de l'état-de-l'art et atteindre des taux de reconnaissance très élevés sur différents ensembles de feuilles avec un fond uniforme.

En outre, la plupart des méthodes de l'état-de-l'art, dans le domaine de l'identification des plantes, sont basées sur l'analyse des images de feuilles. Plus précisément, ces approches sont généralement basées sur des images d'une seule feuille sur un fond uniforme. Seuls quelques travaux ont abordé le problème d'identification de feuille sur un fond naturel (complexe), ce qui est plus susceptible d'être le scénario réel. Les plus efficaces, dans ce cas, étaient ceux qui ont procèdé à un processus de segmentation, que ce soit manuel ou interactif [19, 20, 4]. Dans cette thèse, nous abordons, en plus des images de feuilles avec fond uniforme, le cas des images avec fond non uniforme (naturel) dans le but de proposer les scénarios les plus utiles et les plus informatifs à l'utilisateur.

Néanmoins, aucun processus de segmentation n'est utilisé dans ce cas.

5 Contributions

Dans le cadre de cette thèse, nous étudions dans quelle mesure l'automatisation du processus d'identification des plantes peut-elle minimiser l'intervention humaine, tout en assurant des taux de reconnaissance élevés. Pour cela, nous nous intéressons à différents scénarios automatiques et semi-automatiques, pour atteindre le bon compromis entre une identification erronée mais entièrement automatisée et une identification très précise mais entièrement manuelle.

Nous utilisons des stratégies hiérarchiques, basées sur des caractéristiques discriminatives de l'objet; le but étant de réduire considérablement l'espace des candidats. Le résultat peut être une estimation unique ou un ensemble d'estimations. Le degré de l'intervention de l'utilisateur peut alors varier d'inexistant à significatif dans des situations ambiguës. Il dépend aussi des données. Par exemple, pour des photos naturelles de feuilles, le degré d'interaction entre l'utilisateur et le système peut être plus important que pour les images sur fond uniforme.

Plus formellement, soit $\mathcal{Y}$ l'ensemble des catégories et $Y(I)$ la bonne catégorie de l'image I. Le but ultime est de prédire Y. Pour cela, nous étudions deux stratégies: (i) proposer une liste ordonnée d'estimations. Seule la première estimation dans la liste est considérée dans le cas de référence, le cas d'un scénario sans aucune intervention humaine, (ii) proposer un *ensemble de confiance* (CS) qui dépend de I et tel que $Y \in CS$ avec une probabilité très élevée: $P(Y \in CS) \geq 1 - \epsilon$. Si la performance de la première stratégie peut être mesurée par la fréquence à laquelle la bonne catégorie figure parmi les k meilleurs estimations, la performance de la deuxième, est essentiellement mesurée par la taille du CS retourné.

Ces deux stratégies sont basées sur une représentation hiérarchique de $\mathcal{Y}$. La hiérarchie $\mathcal{T}$ permet de définir des vecteurs descripteurs ainsi que des scores (discriminants) locaux X_t , $t \in \mathcal{T}$. Nous étudions des approches utilisant de la connaissance botanique ainsi que des approches plus génériques. Nous nous intéressons en particulier à (1) de nouvelles représentations d'objet ainsi que (2) de nouveaux algorithmes de classification.

5.1 Nouvelles représentations de l'objet

Une classification efficace est fortement liée à une description discriminative de l'objet. Dans cette thèse, nous explorons la description et la représentation d'objets, plus particulièrement d'objets botaniques.

Nous introduisons, tout d'abord, une représentation hiérarchique de clés d'identification botaniques *IdKeys* dans le but d'imiter le processus d'identification décrit par les botanistes eux-mêmes. Une telle représentation est appliquée sur des images de feuilles sur fond uniforme, où une feuille est représentée par un ensemble ordonné d'attributs correspondants à des *IdKeys*. En particulier, ces clés codent des caractéristiques invariables - indépendantes des espèces végétales ainsi que du contexte, comme l'endroit, les conditions climatiques ou la saison. Ils peuvent représenter des propriétés géométriques qui ne sont pas directement observables, telles que des points de repère, ou des groupes d'espèces prédéterminés.

Ensuite, nous proposons une représentation plus générique basée sur ce nous avons appelé des *fenêtres de description*, dont les caractéristiques clés sont l'emplacement, l'orientation et la description locale. Alors que l'emplacement de ses fenêtres est le même pour toutes les espèces, ce sont les caractéristiques locales qui permettent de lever l'ambiguité entre les estimations candidates. Deux aspects importants sont étudiés dans le cadre de cette description (1) où précisément chercher? (2) que devrions nous calculer?

5.2 Classification

L'idée principale dans les scénarios d'identification proposés, est de profiter d'une représentation hiérarchique $\mathcal{T}$ des données.

Nous construisons, d'abord, des classifieurs locaux, en utilisant le rapport de vraisemblance et des fonctions discriminantes locales. La hiérarchie est ensuite parcourue en largeur allant des classifieurs les plus "grossiers" vers les plus "fins": à chaque niveau, tous les enfants d' un noeud t *positif* sont retenus et testés au niveau suivant. Alors que les faux positifs peuvent être réduits successivement, si la bonne hypothèse est rejetée au niveau d'un certain noeud, alors elle ne peut plus être récupérée. Les espèces *positives* (contenues dans les noeuds terminaux) sont finalement classées en fonction de

leurs rapports de vraisemblance .

Puis, par analogie avec les intervalles de confiance dans les statistiques classiques, nous construisons un modèle probabiliste qui permet de sélectionner un *ensemble de la confiance* (CS). La taille de cet ensemble joue le rôle de la largeur de l'intervalle de confiance dans les statistiques standards et la probabilité à posteriori que la bonne catégorie appartienne à l'ensemble de la confiance, joue le rôle du degré de confiance. L'idée est de restreindre les ensembles candidats aux sous-ensembles $\{C_t, t \in \mathcal{T}\}$ et d'intégrer tous les éléments de preuve à partir des scores des différents noeuds pour calculer la probabilité à posteriori $P(Y \in C_t|\mathbf{X} = x)$, où $\mathbf{X} = \{X_t, t \in \mathcal{T}\}$. Pour cela, nous utilisons la densité conditionnelle jointe $p(\mathbf{x}|c)$ des scores $\mathbf{x} = (x_t, t \in \mathcal{T})$, $c \in \mathcal{Y}$. (Nous supposons une distribution à priori $p(c)$ uniforme des espèces $c \in \mathcal{Y}$.) Notre modèle est basé un réseau bayésien gaussien.

6 Résultats

Nous avons procédé à différents tests de performance afin d'évaluer les approches proposées tout au long de cette thèse. Plusieurs bases de feuilles ont été utilisées. Les feuilles considérées appartenaient à différentes régions dans le monde (exemples: Suisse [104], France [50], Etats-Unis) et ont été numérisées de différentes manières. Des images avec fond uniforme ainsi que des images avec fond naturel ont été considérées.

Nous avons, tout d'abord, comparer la première stratégie de classification, en utilisant les deux descriptions proposées, avec différentes méthodes de l'état de l'art sous les mêmes protocoles d'évaluation. Des taux de reconnaissance de plus de 95% ont été atteints dès la première estimation pour la plus part des images de feuilles avec fond uniforme. Plusieurs méthodes de l'état de l'art ont été surpassées.

Aussi, les *ensembles de confiance* retournés à l'utilisateur contenaient en général moins de 3 espèces (avec un taux de reconnaissance très élevé) ce qui peut être d'une énorme aide aux spécialistes lors de la classifications de nouvelles collections de feuilles. Issus d'un clustering agglomératif, ces ensembles contiennent généralement des espèces visuellement très similaires, ce qui peut ouvrir d'autres perspectives sur l'étude des relations entre espèces.

Même si les taux de reconnaissance sont loin d'être parfaits pour les images avec

fond naturel, notre approche était meilleure que plusieurs méthodes, y compris celles qui ont eu recours à la segmentation. Cependant, des pistes plus spécifiques à ce genre de problématique sont encore à creuser. Le manque de protocole lors de l'acquisition de la photo peut affecter l'apprentissage et l'efficacité des cassifieurs. Par exemple, des photos de feuillage touffu ou de tout l'arbre ne peuvent guère aider pour identifier une photo d'une seule feuille.

7 Conclusion et perspectives

Dans le cadre de cette thèse, nous avons étudié la problématique de classification dite "fine" en se concentrant sur la détermination des espèces botaniques à partir d'images de feuilles. Le point fort de ce travail est le fait d'aborder l'ensemble de la chaîne d'identification. Nous nous sommes intéressés aussi bien à la description et la représentation de l'objet qu' aux algorithmes de classification et des scénarios d'identification utiles à l'utilisateur.

Nous nous sommes inspirés du processus manuel des botanistes pour introduire une nouvelle représentation hiérarchique des feuilles. Nous avons aussi proposé un nouveau mécanisme permettant d'attirer l'attention au tour de certains points caractéristiques de l'objet et d'apprendre des signatures spécifiques à chaque catégorie.

Nous avons adopté une stratégie de classification hiérarchique utilisant une série de classifieurs locaux allant des plus grossiers vers les plus fins; la classification locale étant basée sur des rapports de vraisemblance. L'algorithme fournit une liste d'estimations ordonnées selon leurs rapports de vraisemblance. Motivés par les applications, nous avons introduit un autre scénario proposant à l'utilisateur un ensemble de confiance contenant la bonne espèce avec une probabilité très élevée. Un nouveau critère de performance a été donc considéré, la taille de l'ensemble retourné. Nous avons proposé un modèle probabiliste permettant de produire de tels ensembles de confiance et démontré l'efficacité de cette stratégie sur plusieurs bases de feuilles.

Différents scénarios semi-automatiques ont été proposés à l'utilisateur pour l'aide à la décision et à l'identification d'espèces de plantes via différents types d'images de feuilles. Des performances très élevées ont été atteintes pour la reconnaissance des feuilles sur fond uniforme et des améliorations sont encore à faire pour celles avec fond

naturel. Notons que notre méthode basée sur les ensembles de confiance et l'intervention de l'utilisateur a réalisé le meilleur score sur les photos de ImageClef2011 en surpassant même celles qui ont eu recours à la segmentation.

Une interaction encore plus importante avec l'utilisateur peut améliorer les performances des systèmes de reconnaissance des photos de feuilles, en particulier celles prises sans aucun protocole pré-défini. Aussi, traiter chaque type de photos (photos d'une feuille, d'une branche, d'un feuillage, ou de tout l'arbre) différemment peut être une piste intéressante.

Etendre notre travail en utilisant les différentes stratégies proposées dans le cadre de la classification des requêtes définies par de multiple images d'organes botaniques tels que fleurs et fruits peut améliorer le processus d'identification et de classification des collections de plantes. Aussi, nous serons intéressés de tester le mécanisme basé sur les ensembles de confiance afin d'organiser la recherche d'autres objets non-botaniques.

8 Organisation de la thèse

Cette thèse est organisée comme suit:

- **Chapitre 2** dresse l'état-de-l'art en relation avec les problématiques abordées dans cette thèse.

- **Chapitre 3** présente deux nouvelles représentations d'objets, notamment la représentation hiérarchique de *clés d'identification*, basée sur des caractéristiques botaniques et les fenêtres de description permettant de discriminer des classes très simaires.

- **Chapitre 4** introduit deux autres contributions qui concerne le domaine de la classification, la première étant basée sur une première catégorisation grossière puis un raffinement séquentiel qui réduit à chaque fois l'espace des estimations potentielles. La deuxième étant un modèle probabiliste basé sur une analogie avec les intervalles de confiance en statistique.

- **Chapitre 5** détaille les différents scénarios d'identification que nous proposons, notamment l'identification entièrement automatique et semi-automatique en util-

isant une seule image ou des images multiples.

- **Chapitre 6** présente les bases de données et les expériences utilisées pour évaluer les techniques et les scénarios proposés. Des comparaisons et des analyses critiques sont fournis.

- **Chapitre 7** présente le bilan des contributions ainsi que les différentes perspectives.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Computer vision is the science of endowing machines with vision, or the ability to "see", to interpret and represent image content on semantic level. Computer vision covers the core technology of automated image analysis which is used in many fields. Applications range from low-level image processing to automated image annotation, scene interpretation and object recognition.

More specifically, automated object recognition, the focus of this dissertation, aims to address the issue of how to replicate or emulate human ability of recognizing and discriminating a multitude of object categories in images. Research in automated object recognition is currently very active, driven by applications as well as the intellectual challenge, and there have been notable recent advances using both discriminative learning and object modeling for detecting and localizing instances of generic object classes such as cars, cats and people appearing in digital images [10, 23, 28, 44, 49, 108, 120].

Motivated by applications in areas such as botany, agriculture, medicine and forestry, there has also recently been considerable interest in more fine-grained discrimination, such as species or breed recognition; see Figure 1.1. Unlike the socalled basic-level recognition, which refers to recognizing and distinguishing between generic object classes, fine-grained recognition works within a single base-level category and aims to distinguish among its (sub)categories. In other words, fine-grained recognition deals with the semantic space in a much *higher resolution*, that is, containing a large number of closely-related categories, and distinguishes itself from traditional object recognition

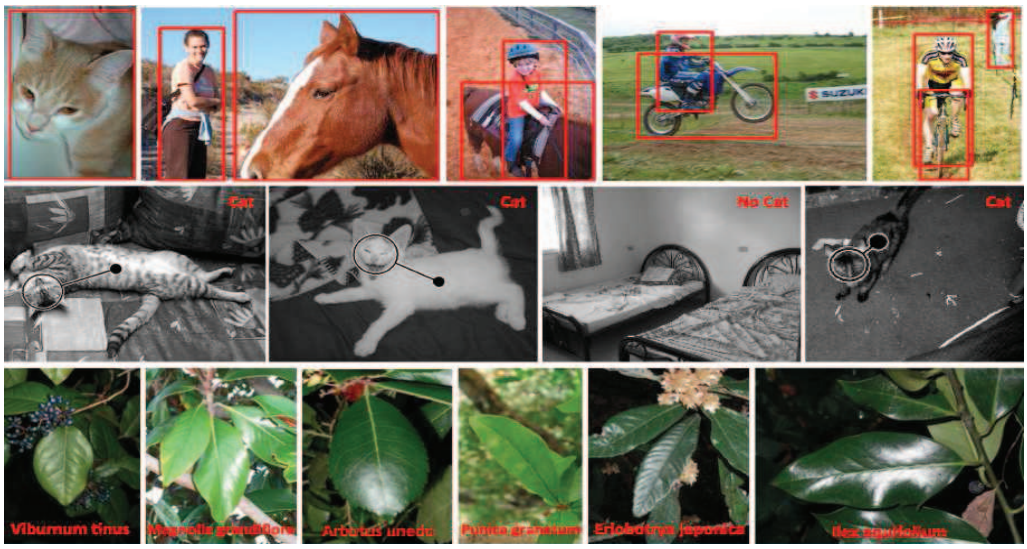


Figure 1.1: Examples of different recognition tasks. (1) Predict bounding boxes of instances of objects (top row). (2) Predict the presence/absence of a specific class with providing a detailed description of the pose of each detected instance (middle row). (3) Predict fine-grained categories of the same basic class (bottom row).

that studies a much *coarser* sampling of the semantic space.

Fine-grained recognition is generally more difficult than the basic-level recognition for both humans and computers. Differences between (sub)categories are often very fine and not noticeable to the common eye. This thesis deals with the problem of fine-grained recognition for determining botanical species from leaf images, but using a generic framework.

1 Motivations

One major challenge in artificial intelligence and computer vision research is endowing machines with human ability to discriminate among similar objects. However, unlike basic-level recognition, even humans might have difficulty with fine-grained categorization. (Sub)categories (e.g., species of plants, breeds of dogs) are usually recognized by experts, while one can recognize immediately basic categories. In the case of botanical species recognition, usually only a trained human expert (e.g., a botanist) can do the task, and usually not without following a complex identification procedure, using

Figure 1.2: Illustrations from some filed guides.

further filed guides [40]; see Figure 1.2.

Taxonomists often present organized written descriptions of the characteristics of similar species so that other biologists can identify unknown specimens. In botany, this task implies comparisons among stored and observed characteristics and then assigning a particular plant to a known taxonomic group, ultimately arriving at a species. However, due to the large variation of patterns among fundamental features and the very large number of biologically relevant plant categories (more than 200,000 [99]), identifying botanical species can be an onerous and time-consuming task even for experts. Generally, the situation is the same in other domains of fine-grained categorization. Finding a species quickly often requires knowing in advance the name of the family or the genus involved, and an expert on one genus or family may be unfamiliar with another. The difficulty is further increased by the ongoing shortage of skilled taxonomists (known as the *taxonomic impediment* [18]).

Since very few people can successfully distinguish among a large number of species, an automated (or a semi-automated) visual system for this task could be valuable in many applications for both experts and non-experts. Automated tools can accelerate recognition and thus help botanists in recognizing plant species or conjecturing new ones. Speeding up the collection and classification of botanical observation data is a crucial step towards a sustainable development of agriculture and the conservation of biodiversity. Other possible uses of such automated systems include developing educational tools, combating the illegal trade in endangered species and building accurate knowledge of, for example, poisonous plants.

2 Challenges

In this section, we discuss the challenges of fine-grained categorization, especially in the botanical field. In particular, we focus on the key ingredients: data, intra-species variability versus inter-species similarity, and the problem of cluttered images.

2.1 Lack of data

Unlike traditional datasets devoted to basic-level classification tasks and which contain different concepts such as animals, plants, and cars, a fine-grained image collection typically contains hundreds of categories sharing the same basic concept. For example, the Stanford Dogs dataset [66] contains 120 kinds of dogs and there are more than 200 plant species in the Smithsonian leaf dataset [7]. To find the small-variations between categories that share similar semantics, extra annotation is usually needed, such as taxonomic labels (e.g., family, genus, species), bounding boxes and part locations. Annotated data are generally a critical component for machine learning problems and especially for fine-grained problems.

Another crucial issue in machine learning and pattern recognition is the class imbalance problem [62], when the class distribution is highly imbalanced due to the lack of data. In particular, in datasets which deal with taxonomic categories (e.g., species), the number of samples per category often depends on the rarity and the interest of the species. Collecting and integrating botanical observation data could be very challenging since it needs specialized social network and expert botanist validation [65]. However, the size of the data set obviously has an important role in building a good classifier. For unbalanced data sets, the decision boundary established by standard machine learning algorithms tends to be biased towards the majority class; therefore, minority class instances are more likely to be misclassified. Indeed, lack of examples makes it difficult to uncover regularities within small classes.

2.2 Diversity of morphological characteristics

Fine-grained categories are by definition very similar since they belong to the same basic concept. Another problematic issue is then the large intra-class variability in

Figure 1.3: Intra-class similarity and inter-class variation for leaves. Displayed are two samples of *Quercus ilex* species on the top row and of *Ilex aquifolium* species on the bottom row.

addition to the inter-class similarity. Taxonomic categories (e.g., genus or species) are often determined by subtle differences in shape and texture. In fact, there can even be less variation in appearance between two images from two different (sub)categories than within a single one, as illustrated in Figure 1.3. And whereas the overall shapes may be sufficiently different to distinguish between some species, other species may display only subtle differences.

In particular, plants may exhibit different shape characteristics due to local context, such as location, climatic conditions and age. For example, leaves often have different colors and shapes throughout the seasons. A dead leaf have a very different texture than a living one. The same species could exhibit very different characteristics in two different regions. Figures 1.4 to 1.10 illustrate the intra-class variability over several criteria including global shape, margin, texture, color, number of lobes as well as number, shape and relative positions of leaflets.

Leaves may also vary continuously or discretely even along a single stem as they develop (known as *leaf heteroblasty*); for example, leaves in Figure 1.11 come from the same plant.

Figure 1.4: Leaf global shape variation of *Acer saccharinum* species.

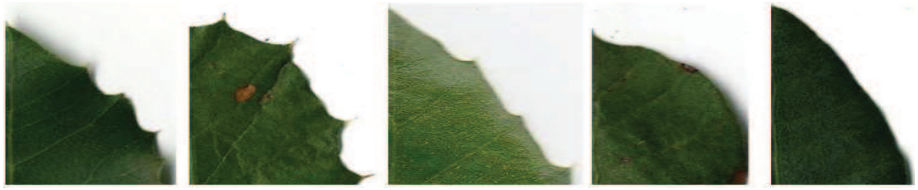


Figure 1.5: Leaf margin variation of *Quercus ilex* species.

2.3 Image background complexity

The background can also represent a key issue specific to an automatic recognition system. Cluttered background can serve as a challenging distractor, especially when the system assumes it is part of the object.

Most of the publicly available datasets used for plant species identification consist of leaf images on a uniform (white) background; examples include the Swedish [104], the Flavia [119] and the Smithsonian [7] datasets. However, the identification problem is far more challenging using cluttered images, which is the case of unconstrained photographs (e.g., branch and foliage photos, other plants in the background, or other objects such as fingers, yardsticks, grass, etc.). Such images were recently published in ImageClef benchmarks [1][2] and are accumulating at a staggering rate due to mobile devices and applications[3]; see Figure 1.12.

[1] http://www.imageclef.org/2011/Plants
[2] http://www.imageclef.org/2012/Plants
[3] https://itunes.apple.com/en/app/plantnet/id600547573

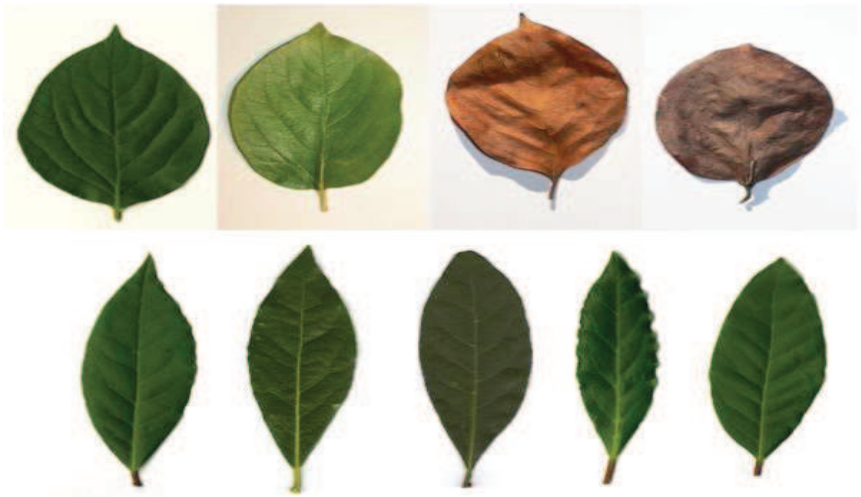


Figure 1.6: Leaf texture variation of *Diospyros kaki* (top row) and *Laurus nobilis* (bottom row) species.

3 Positioning in the literature

Due to the aforementioned needs and challenges, there is now a body of literature in computational vision devoted to fine-grained recognition, including identifying botanical species, and designed for both experts and non-experts. A variety of methods were adapted or introduced to discriminate similar objects and several approaches have made use of human input to improve accuracy. Most of this work is summarized in Chapter 2.

Generally, the *baseline* scenario was to provide the user with a single estimate; examples include [35, 45, 119, 68, 117, 78, 13, 2]. Other researchers chose to report the k most similar classes [7, 67, 75], where k usually ranges from ten to twenty in order to make it likely the true species is among the k reported ones. Many of the most efficient approaches for fine-grained recognition achieved about 70% accuracy on the first estimate and about 90% on the top-10 estimates while considering challenging and relatively large datasets, i.e., containing more than one hundred species with high inter-class similarity and intra-class variability; examples include [74, 7, 67, 75]. Of course, retaining the true species while returning a relatively large set of estimates has limited value in real-world applications. Also, focusing only on the few first estimates without achieving near-perfect (human-level) performances could be essentially useless

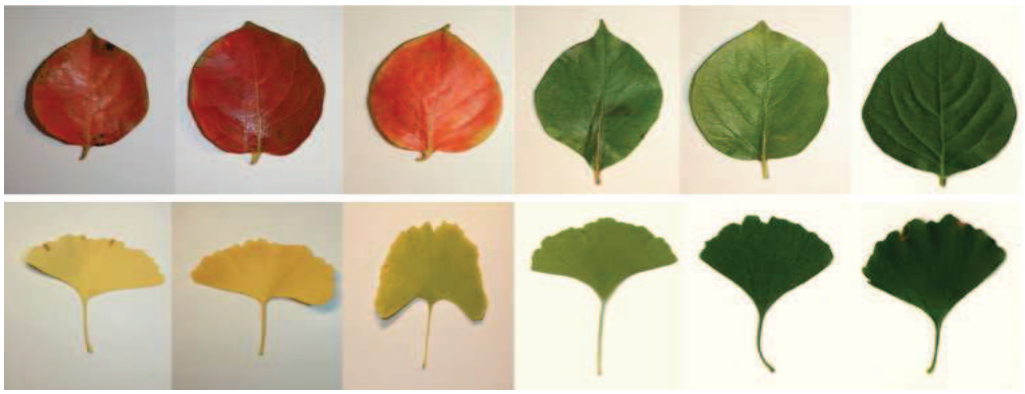


Figure 1.7: Leaf color variation of *Diospyros kaki* (top row) and *Ginkgo biloba* (bottom row) species.

Figure 1.8: Number of lobes variation of *Ficus carica* species.

and uninformative for the user. Here, we focus on (1) hierarchical strategies, based on discriminative features, in order to drastically reduce the space of candidate sets of estimates and (2) a new performance criterion: report a subset of species whose expected size is minimized subject to containing the true species with high probability. In particular, we outperform previous methods and achieve near-perfect recognition rates on several leaf images with uniform background; see Chapter 6.

Most state-of-the-art methods in plant identification are obviously based on leaf image analysis and in few cases on flowers [90, 29]. More specifically, such approaches are generally based on images of a single leaf with uniform background [26, 114, 112, 104, 36, 34, 79, 45, 74, 117, 67, 85, 81, 83]. Only a few papers addressed the problem of identifying leaf images on cluttered backgrounds [113, 19, 20, 4] which is the more likely real-world scenario. In this case, the most efficient approaches were those based

Figure 1.9: Leaflet shape and number variation of *Sambucus canadensis* species.

Figure 1.10: Leaflet relative position variation of *Parthenocissus quinquefolia* species.

on a segmentation process, either manual or interactive; more details can be found in §2.3. In this work, we also investigate leaf images with natural cluttered background (in addition to those with uniform background) with the aim of proposing the most useful and informative identification scenario to the user. However, no segmentation process is required.

4 Contributions

We investigate to what extent automated plant identification can minimize human intervention while ensuring high recognition rates. To this end, we focus on different scenarios, with a human in the loop, to achieve something sensible between the two extremes of an inaccurate but fully automated identification and a very accurate but fully manual identification.

We use hierarchical strategies, based on discriminative features, in order to drastically reduce the space of candidates. The output can be either a single estimate or a set of estimates. The degree of the user intervention can then range from non-existent to significant in ambiguous situations. It is also data-dependent. For example for natural photos of leaves, the degree of interaction between the user and the system may be more

Figure 1.11: Illustration of leaf heteroblasty for *Gleditsia triacanthos* species, i.e., pronounced changes in leaf morphology during plant development.

important than for images on uniform background. A brief summary of the different proposed scenarios (detailed in Chapter 5) is displayed in Figure 1.13.

More formally, let $\mathcal{Y}$ denote the complete set of (fine-grained) categories and let $Y(I)$ denote the true category of image I. Our task is to predict Y. To this end, we investigate two strategies: the first is providing a ranked list of categories (the top-1 estimate is considered for the *baseline case*, i.e., without human intervention), the second is providing a *confidence set* (CS) which depends on I and such that $Y \in CS$ with high probability, say $P(Y \in CS) \geq 1 - \epsilon$. While performance in the former is measured by the rate at which the true category appears among the top k estimates, performance in the latter is essentially measured by the expected size of CS.

Both strategies are anchored by a hierarchical representation of $\mathcal{Y}$. The hierarchy $\mathcal{T}$ serves as a platform for defining features and local discriminant scores X_t, $t \in \mathcal{T}$. We investigate both botanical-based and more generic approaches. In particular, we focus on (1) novel object representations (see Chapter 3) as well as (2) novel classification methods, especially a novel framework for organizing the search; see Chapter 4.

4.1 Object representations

Efficient classification is related to discriminating object description. In this dissertation, we explore improvements in object representation based on specific knowledge domain, and investigate both category-independent and category-dependent features.

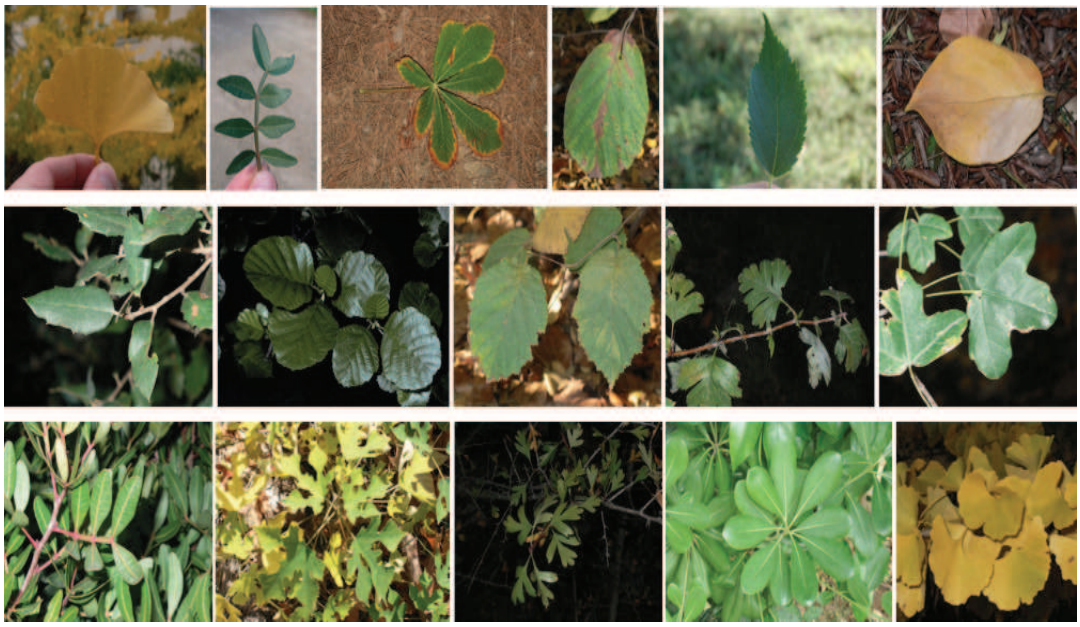


Figure 1.12: Examples of unconstrained photographs. One can photograph a picked leaf (top row), a branch (middle row) or a foliage (bottom row).

First, we focus on a hierarchical representation of botanical *identification keys* (IdKeys) with the aim of mimicking the *process* of identification described by botanists. Such a representation was applied on uncluttered leaf images where a leaf is represented by an ordered set of attributes corresponding to IdKeys; see §3. In particular, these keys encode invariable characteristics, i.e., are independent of the plant species as well as the context, such as geography, climatic conditions, season and instantiation. They could refer to geometric properties not directly observable, such as landmarks, or to predetermined groups of species.

Second, we investigate a more generic representation based on *vantage feature frames* and in which landmarks are considered more like "vantage points" in that orientation plays a role as well; in other words, where the landmarks are in relation to one another. And, whereas these landmarks are the same for each species, it is the local features which permit disambiguation. The vantage feature frames deal with two important aspects: *where to look* and *what to compute*; see §4.

Both representations serve to induce local discriminant functions and classifiers for an efficient identification.

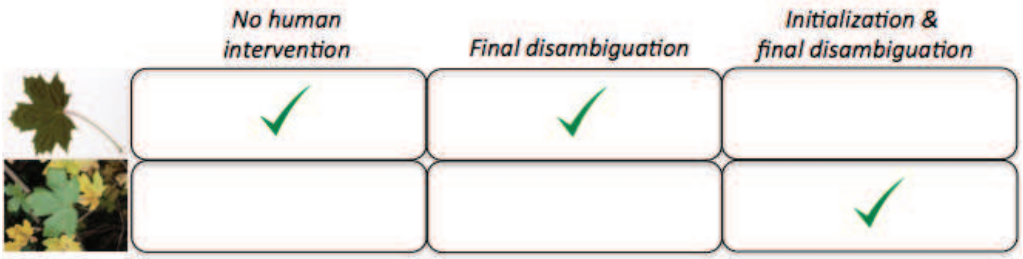


Figure 1.13: We consider three levels of user interaction in identifying the species of leaf images with either uniform and cluttered background: none (the baseline case of a fully automated system returning a point estimate); final disambiguation (the user receives a set of estimates); initialization and final disambiguation (the user also initializes the process by identifying landmarks). The checks indicate the level we require to obtain useful/satisfactory results.

4.2 Classification algorithms

The main idea behind our classification methods in the proposed identification scenarios is to take advantage of a hierarchical representation $\mathcal{T}$ of the data. Examples of hierarchies of leaves are shown in Figure 1.14.

We first build local classifiers using likelihood ratio and local discriminant functions. The hierarchy is then processed breadth-first coarse-to-fine: at each level, all the children of a *positive* node t are retained and tested at the next level. Whereas false positives can be successively pruned, if the true hypothesis is rejected at a node containing it then it cannot be recovered. The *positive* species (i.e., terminal nodes) are finally sorted according to their likelihood ratios.

Then, in analogy with confidence intervals in classical statistics, we build a probabilistic model which enables confidence set (CS) selection. The expected size of the confidence set plays the role of the width of the confidence interval in standard statistics and the posterior probability that the true category belongs to the confidence set plays the role of the confidence level. The idea is to restrict CS candidates to the subsets $\{C_t, t \in \mathcal{T}\}$ and integrate all the evidence from the node scores to compute the posterior probability $P(Y \in C_t | \mathbf{X} = x)$, where $\mathbf{X} = \{X_t, t \in \mathcal{T}\}$. To this end, we require the joint conditional density $p(\mathbf{x}|c)$ of the scores $\mathbf{x} = (x_t, t \in \mathcal{T})$, $c \in \mathcal{Y}$. (We assume the prior distribution $p(c)$ is uniform over species $c \in \mathcal{Y}$.) We use a Gaussian Bayesian network. More details about the motivation and the modeling can be found in Chapter 4.

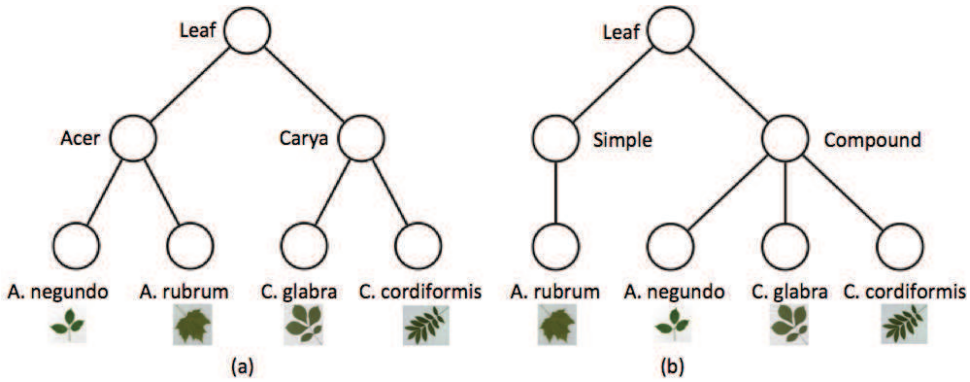


Figure 1.14: Simple hierarchical representations of four species. (a) A semantic hierarchy: the second level represents leaf genera and the third level the species. (b) A hierarchy based on morphological leaf characteristics: the second level represents the leaf type (simple or compound) and the third level the species. A thumbnail from each species is displayed.

4.3 Publications

Following is the list of publications related to our work:

- A. Rejeb Sfar, N. Boujemaa and D. Geman. Identification of Plants from Multiple Images and Botanical IdKeys. In *Proceedings of the ACM International Conference on Multimedia Retrieval*, 2013.

- A. Rejeb Sfar, N. Boujemaa and D. Geman. Vantage Feature Frames For Fine-Grained Categorization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013.

- A. Rejeb Sfar, N. Boujemaa and D. Geman. Vantage Feature Frames For Botanical Species Identification. In *the second workshop on Fine-Grained Visual Categorization, in conjunction with CVPR 2013 conference.*

- A. Rejeb Sfar, N. Boujemaa and D. Geman. Confidence Sets for Fine-grained Categorization and Plant Species Identification. In *International Journal of Computer Vision* (accepted to be published).

5 Outline of the thesis

In this first chapter, we have informally introduced the different issues investigated in this thesis. The remaining part of this dissertation is structured as follows:

- **Chapter 2** reviews the literature that is relevant to the fine-grained categorization, especially the plant species identification, and place our work in the context of related research.
- **Chapter 3** presents two novel object representations. The first representation refers to the IdKey hierarchy which is driven by specific knowledge domain. The second representation is the concept of *vantage feature frames* for fine-grained discrimination.
- **Chapter 4** describes two novel classification frameworks, including the CTF search within the likelihood framework and the model-based approach using a statistical analogy.
- **Chapter 5** details the different identification scenarios that we propose, including fully-automated and semi-automated identifications using either a single or multiple samples.
- **Chapter 6** presents data and experiments used to evaluate proposed techniques and scenarios. Comparisons and critical analyses are provided.
- **Chapter 7** concludes the thesis with a summary of our contributions and some directions for further research.

Chapter 2

Related Work

Our work is related to existing work on fine-grained categorization [37, 78, 75, 90], especially work on plant species identification from leaf images [26, 118, 7, 67, 50]. Different approaches, described in this dissertation, relate to hierarchical classification and classification with class-selective rejection.

This chapter reviews the literature that is relevant to our work. Section 1 provides an overview of related research on fine-grained categorization, especially on some animal species categorization, including birds, insects and dogs. Previous approaches were divided into fully-automated methods (see §1.1) and those using human input (see §1.2). Section 2 describes stat-of-the-art methods on plant identification from leaf images. We discuss both generic (see §2.1) and botanical-based approaches (see §2.2). We also address the problem of cluttered leaf images in previous work in §2.3. Section 3 and Section 4 respectively, present the hierarchical classification and classification with class-selective rejection.

1 About fine-grained categorization

There is a growing body of work investigating identification of birds [37, 109, 123], insects [68, 78], dogs [97, 76, 75], flowers [90, 3] and leaves. Both images and sounds were studied, especially for birds.

1.1 No human intervention

Several fully-automated approaches have been used for fine-grained categorization. Larios et al. [69] describes an automated identification system of stonefly larvae using concatenated histograms of local appearance features. Such histograms involves both region detection and local representation. Shape-based methods, including boundary analyses, have been especially adapted for leaf images using an automatic segmentation [7, 16, 74]. Often, performance is sensitive to the quality of the contour resulting from the segmentation process, which naturally complicates distinguishing between categories with very similar shapes as well as identifying species from cluttered images. Other methods adapt systems for detecting instances of generic object classes [70, 110] by encoding an image as a bag of discrete visual codewords and basing classification on histograms of codeword occurrences; examples include [90, 69]. Some work [126] explore variations such as learning multiple non-redundant codebooks. Other [81, 83] adopt hashing techniques [94] for efficient embedding and matching of high-dimensional feature vectors or [123] use continuous template matching instead of discrete visual codewords for image representation. Again, however, the distinctions among fine-grained categories are sometimes too refined to be captured by variations in bags of visual words, hash tables or random templates.

1.2 Human intervention

To account for such distinctions, an increasing number of studies make use of human input in the identification task. In [13, 109], interactive systems are proposed wherein humans click on bird parts and answer questions about attributes (e.g., "white belly", "red-orange beak", "sharp crown"). The main goal was to efficiently select the most informative questions to pose to the user in order to identify the unknown species as quickly as possible. While [13] incorporates only one type of user input (i.e., binary questions) and uses non-localized computer vision methods based on bag-of-words features, [109] allows heterogenous forms of inputs and incorporates a part-based model.

The authors of [65] propose an interactive mobile application for plant identification which is based on social images of different plant *organs*. The application provides to the user an easy access to a rich botanical knowledge which was collected and revised by

amateur and expert botanists of a social network. In other recent work [33], an online game *Bubbles* is introduced to reveal discriminative features humans use to distinguish between bird species. During the game, the player can choose to reveal full details of circular regions, with a certain penalty. The human selected regions are then used to learn classifiers for closely-related categories.

In [75, 43, 125] annotated training data (e.g., key points and objects parts) are obtained from experts. [75] deals with dog classification and focuses on the dog face location, allowing for a part-based approach and [43, 125] deals with bird classification. In [43], classifiers based on *poselets* [12], i.e., parts of the object from a given viewpoint, are employed to extract part and shape information for building fine-grained models. However, this approach requires 3-D pose annotation, which is based on volumetric primitives that are costly to obtain manually and present other difficulties; instead, the authors of [125] advocate a 2-D rather than 3-D representation in order to reduce the level of annotation required to generate the *poselets*.

Our work on leaves is somewhat similar in that we also make use of specific-domain knowledge and human input; however, we propose novel representations, based on *id-Keys* or *vantage frames*, as well as different identification scenarios based on a pre-defined hierarchical structure. Of course animals and leaves present different kinds of challenges; the former exhibit higher intra-species variation (e.g., birds may be flying, swimming or perched), whereas the latter exhibit more inter-species similarity (e.g., in color, overall shape and internal structure). Note that, for example, bird calls could be also a key point for species identification; so that acoustic classification [1, 14] could be combined to image classification.

2 About leaves

In computational vision, work on plant identification is relatively recent, largely confined to leaves and in a few other cases to flowers [90, 29]. More specifically, most state-of-the-art methods are based on a single leaf analysis with uniform background, including leaf shape, texture, venation and morphological characteristic analyses. Many of these methods were

2.1 Generic approaches

In several previous work, leaves were described using standard computer vision descriptors which can refer to either global or local characteristics.

Global approaches generally represent the whole leaf shape [80, 112, 114], texture [25, 5] or color [6]. For instance, Wang et al. [114] combined different features based on a centroid-contour distance curve to propose a two-stage filtering approach for leaf image retrieval, allowing for reducing the search space. Felzenszwalb and Schwartz [45] proposed a hierarchical shape representation based on a hierarchical description of an object's boundary that performed well on a publicly available leaf dataset (15 Swedish species [104]). This representation is captured by a tree, which they term the *shape-tree* of an object. Wu and Rehg [117] introduced *sPACT* (spatial Principal component Analysis of Census Transform histograms), a new representation for recognizing instances and categories of places or scenes, which performed even better than *shape-tree* on the Swedish leaves [104].

Moreover, Yahiaoui et al. [121] presented a leaf-boundary based approach that attempts to outline foliar properties and achieved good results using the ImageCLEF2011[1] plant identification framework. Yanikoglu et al. [122] employed a variety of shape, texture and color descriptors (117 features in total) to describe a leaf and obtained the best results on the ImageCLEF2012[2] benchmark.

Multi-scale approaches [34, 79, 67] have been also introduced to especially enrich the shape description and make it more robust to contour deformations. For instance, Kumar et al. [67] proposed computing curvature histograms along the contour of the leaf at multiple scales to classify 184 tree species in the Northeastern U.S (not yet publicly available) while introducing the first mobile app *Leafsnap*[3] for identifying plant species. This popular iPhone application which now allows a fair identification of several american plant species by shooting a single picked leaf on an uniform background.

Local approaches compute local features at some landmark points [74, 81, 83] or some regions [27, 96] of a leaf. Landmarks can be either boundary points [8, 74] or salient points [81] of the shape. For example, Ling and Jacobs [74] introduced shape

[1] http://www.imageclef.org/2011/plant
[2] http://www.imageclef.org/2012/plant
[3] http://leafsnap.com

descriptions based on the *Inner Distance*, which they combined with shape contexts [8] (IDSC) to outperform many other approaches on two different leaf datasets. The *inner-distance* is defined as the length of the shortest path between landmark points within the shape silhouette and was used as a replacement for the Euclidean distance to build accurate descriptors for complex shapes like leaves. More specifically, the IDSC represents histograms of distances and angles from sample points in the contour to all other points, along the shortest path inside the leaf shape. The authors of [7] made use of the IDSC to classify larger leaf datasets (more than 150 species) collected in the context of the Smithsonian project (Electronic Field Guide, 2008) and also incorporated it in an earlier version of the mobile app *Leafsnap*.

Mouine et al. extended the shape context method in [81] and presented different methods for plant species identification based on boundary points and Harris points [57]. The key point was to introduce two different sets of points that play different roles in the shape context scheme. They distinguish the *voting* points, which is the set of points used to build the shape context histograms from the *computing* points on where the shape contexts are computed. In [83], the same authors presented a multi-scale shape-based approach, in which they introduced two triangle representations based on local descriptors associated with boundary points (TSLA and TOA) and outperformed previous ImageCLEF2011 results on leaf images with uniform background. The first representation refers to triangles which are described using two side lengths and an angle. The second representation refers to triangles which are described using oriented angles.

2.2 Botanical-based approaches

Domain-specific knowledge is also used to distinguish between similar botanical species. For example, the vein structure could be very interesting to characterize leaves, but the main challenge is to be able to extract it accurately [73] which generally requires high quality of data. Some previous work combined shape with venation features [93, 87] while other work made use of other morphological leaf information. The authors of [60] investigated local detailed shape of the leaf margin. Typically, they focused on detecting the leaf teeth (on the leaf boundary). Du et al. [36] extracted other properties of the

leaf boundary, including aspect ratio, rectangularity, area ratio of convexity, perimeter ratio of convexity, sphericity, circularity and form factor, in order to classify 20 species of plant leaves. Some prior knowledge on simple leaf shape was used to construct a *parametric polygonal leaf template* in [22]. Ten models representing classes of leaf shapes were retained and used for classification. Caballero and Aranda [17] used geometric features, including eccentricity and area to reduce the search space while introducing a novel shape-based leaf descriptor. The authors of [4] obtained good results in the ImageCLEF2012 Plant Identification Task [4] by addressing simple and compound leaves separately using many morphological features and a single leaflet analysis for compound leaves. Also, several approaches have exploited specific well-known landmarks and some measurements for leaf retrieval and plant identification. In [64], landmarks were manually captured and linear and angular measures were derived from the landmark configuration in order to examine relationships between three species of *Acer* genus. One difficulty in such approaches is the automatic extraction of the landmarks. Recently, Mzoughi et al. [84] introduced an automatic method for detecting different leaf parts and used it in [85] to identify scanned leaf images on white background.

Our work is somewhat similar to these morphological approaches in that we also propose to exploit domain knowledge about taxonomy and landmarks in order to build meaningful representation of the object.

2.3 Segmentation and cluttered background

Work on botanical species identification, including most of those mentioned above, generally deals with leaf images on uniform backgrounds. Often, the Otsu thresholding method [91] was used to extract leaf boundary since the image background is homogeneous [114, 36, 16, 81, 83, 104]. However, some methods have used more sophisticated algorithms on images with uniform backgrounds such as Expectation-Maximization with post-processing to remove false positive regions [67].

Only few work addressed the problem of identifying leaf images on cluttered backgrounds which is more likely to be the real-world scenario. To tackle this problem, most of them have designed novel segmentation algorithms to overcome the difficulties

[4] http://www.imageclef.org/2012/plant

posed by a natural background. Obviously, isolating green leaves in an overall not less green environment seems like an other more difficult issue. The authors of [112], considered prior shape information and proposed an automatic marker-controlled watershed segmentation method combined with pre-segmentation and morphological operation to segment leaf images with complicated background. Teng et al. [105] proposed to recover the 3D position of a leaf from different cluttered images with close viewpoints. Then they performed a 2D/3D joint segmentation using 3D distances and color similarity. In [113], an automatic marker-controlled watershed segmentation method is combined with pre-segmentation and morphological operation to segment leaf images with cluttered background based on the prior shape information. Cerutti et al. [21] proposed a two-step active contour segmentation algorithm based on a polygonal leaf model processes the image to retrieve the contour of only simple and palmately lobed leaves. In the case of weed leaves, deformable templates have been used in [77] to segment one single species *Setaria viridis*, providing promising results even with occlusions and overlaps.

Dalcimar et al. achieved the best results on natural leaf photo classification either at ImageClef2011 or ImageClef2012 plant identification tasks [50, 51]. At both tasks, they proposed a shape boundary analysis based on a prior leaf segmentation. To this end, a manual segmentation was performed at the first task [19] while a semi-automatic segmentation was performed at the second task [20]. In the semi-automatic approach, the photo is first automatically segmented by the Mean Shift algorithm. Then, the user needs to mark the location of the leaf and the background on automatically detected regions to guide a merging process. The authors of [4] also achieved good results on natural photos in the ImageClef2012 benchmark. They proposed a feedback-based segmentation scheme using the GrabCut algorithm [98], in which the user makes corrections to the segmented image. Note that all the participants at the ImageClef2012 benchmark who did not use segmentation processes on cluttered images (i.e., photo category) achieved less than $s = 0.2$ as evaluation score[5]; the best score for photo category was $s = 0.51$. One exception is the work of [92] in which no segmentation was performed with all images, including photos, and which achieved a score of $s = 0.32$ for

[5]http://www.imageclef.org/2012/plant

photo category. They used multiple descriptors and visual-patches encoder. More they combined the descriptors, more they obtained better performances.

In this work, we have not attempt segmentation process in the case of leaf images with cluttered background. Instead, we suggest to manually mark some landmarks. We will demonstrate the efficiency of our approach on several kind of leaves.

3 Hierarchical representation and search

Hierarchy is a powerful organizing principle for both representation and search [72, 15, 41, 103]. The idea is to decompose the original problem into more tractable sub-problems sharing more homogenous properties. One monolithic classifier could be then replaced by a hierarchy of classifiers which gather increasingly detailed information about the object under investigation. Many real-world classification problems, are naturally cast as hierarchical classification problems, where the classes to be predicted are organized into a class hierarchy, typically a tree or a Direct Acyclic Graph (DAG). Many of them utilize semantic class hierarchy, including the sharing of training examples across semantically similar categories [47] or combining information from different levels of the semantic hierarchy [127]. Deng et al. [31] consider exploiting the semantic hierarchy in the context of more than $10,000$ categories. More recently, the authors of [32] use hierarchical structures, in large scale recognition, to maximize the information gain while maintaining a relatively small error rate.

In the fine-grained field, only few previous work has taken advantage of a hierarchical structure for categorization or identification tasks. For example, the authors of [60] introduce a hierarchical representation and recognition method of plant species which reflects structural properties of the leaf margin. They detect leaf teeth and consider three hierarchies. The first hierarchy is a representation of global shapes of leaves; the second is of local detailed shapes and the third of teeth. In [86] botanical description are investigated, focusing on leaf arrangement to distinguish between four kinds of leaves (i.e., simple, pinnately compound, palmately compound and compound trifoliate). A hierarchical strategy is followed in order to reduce ambiguity starting from the most different shapes to the closest ones. Both of these methods involve detailed analysis but still be limited in efficiency. The former can not be applied to all kind of leaves. For

example, it does not consider the case of similar simple untoothed species. The latter needs further processing till arriving to the species categories, which is the ultimate target in the identification process. Both of them need a high-quality contour.

Other kind of hierarchies is used in [42] for species identification, i.e., the natural semantic hierarchy which is based on taxonomic groups (such as family and genus). Of course, using such a hierarchy needs specialized domain knowledge about species and taxonomy. In this dissertation, we investigate several hierarchical representation and search strategy. We build a hierarchical representation of leaves based on botanical knowledge, we use pre-defined taxonomic groups, which are defined according to both shared physical and genetic characteristics and finally, we consider purely visual characteristics to automatically build a hierarchy, using an agglomerative clustering on training data.

4 Class-selective rejection

Class-selective rejection [56, 52, 30] is an extension of basic simple rejection [124, 38] in the multi-class case. That is, when an input pattern cannot be reliably assigned to one of the pre-defined classes in a multi-class problem, it is assigned to a subset of classes that are most likely to fit the pattern, instead of simple rejection. Selecting the most promising classes allows to reduce the error rate and to propose a reduced set to another classifier or an expert, which is of great interest in many decision making systems. Examples of class-selective rejection rules include those defined in [55, 56, 59]. The simplest and the most used rule is the *top-n ranking*, in which n takes its values between one and the total number of classes considered. Another popular one is the *constant risk* [55] rule which consists of selecting, for each pattern, the minimum number of best classes so that the accumulated posterior probability exceeds a pre-defined threshold. In [56], Ha defined a new optimality criterion to be the best tradeoff between error rate and average number of classes. An *optimum* class-selective rejection rule was then obtained by solving a discrete convex minimization problem. The authors of [52] addressed the problem of multi-class decision with class-selective rejection and performance constraints. The problem was defined using three kind of criteria: the label sets, the performance constraints, and the average expected loss. More recently, Deng

et al. [32] connected class-selective rejection with hierarchical classification, to restrict the subset of selected classes to internal nodes of a predefined hierarchy. In our work, we also focus on providing the best subset of classes, i.e., the smallest set which contains the true species with high probability. To this end, we use a predefined hierarchy but within a novel probabilistic model-based framework in which internal nodes are considered as *confidence set* candidates.

Chapter 3

Object representation

1 Introduction

Many of the most accurate approaches for fine-grained recognition are based on both specific-domain knowledge and detecting and extracting features from specific part locations of the object. For instance, [43] uses the *poselet* framework [12] and annotated data, obtained from experts, to localize the head and the body of birds, allowing for part-based location-specific feature extraction. The authors of [9] propose a method to build a large set of part-based one-vs-one features for fine-grained categorization based on a dataset of images labeled by class and as well as part locations. In dog breed classification, one may focus only on the dog face and its parts, e.g., eyes and nose [75, 97]. In fact, part-based approaches are naturally suited to fine-grained recognition since the differences between (sub)categories are very fine and objects within the same basic-level category often share the same part structure [106]. In contrast, objects from different basic-level categories, like a car and a person, lack such natural part correspondence.

In this chapter we propose two object representations of a leaf based on domain knowledge about botany and part locations. In Section 2, we first define the leaf parts as in botany. Section 3 introduces a hierarchical approach in which some attributes and well-defined landmarks are found sequentially and adaptively, while section 4 describes a more generic part-based approach using both category-independent and category-dependent features which we also apply on orchid flowers.

2 Leaf definition

In botany, a leaf is defined as a colored, usually green, expansion growing from the side of a stem, in which the sap for the use of the plant is elaborated under the influence of light. The leaf is one of the parts of a plant which collectively constitute its foliage. Usually, a leaf consists of a blade (i.e., the flat part of a leaf), supported upon a petiole (i.e., the small stalk attaching the leaf blade to the stem) which, continued through the blade as the midrib, gives off woody ribs and veins that support the cellular texture. The petiole may be absent in some cases; the leaves are then called "sessile" (i.e., the blade attaches directly to the stem). According to the leaf architecture manual [39], the internal shape of the blade is characterized by the presence of vascular tissue called veins, while the global shape can be divided into three main parts: (1) The basal part, usually the lower 25% of the blade; the base, which connects the blade to the petiole, is situated at its center. (2) The apical part, usually the upper 25% of the blade and centered by a sharp point called the apex (or the tip). (3) The margin is the border or edge of the leaf; see Figure 3.1.

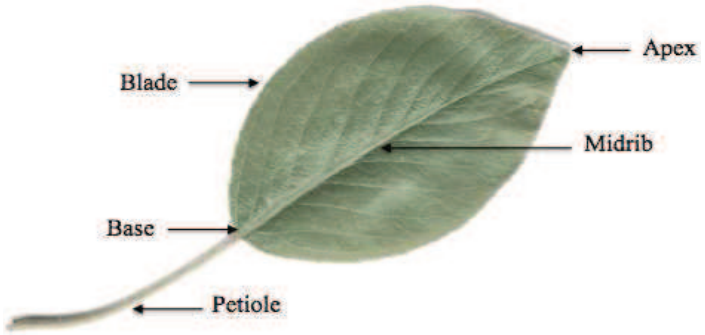


Figure 3.1: The main parts of a leaf.

The blade can be one part or divided into several parts which are called leaflets, and are connected to an extension of the petiole (the rachis). Depending on the blade division, the leaf can be called either "simple" or "compound". However, the "simple" leaf shape may be also formed of lobes, but the gaps between lobes do not reach to the midrib. Also, different "compound" leaves exist in the nature. Figure 3.2 illustrates different kinds of blade divisions (e.g., lobed, palmate, trifoliate, pinnate, etc.).

Figure 3.2: Examples for blade divisions. Simple leaves are displayed in the top row and compound leaves in the bottom row.

Such botanical ingredients might give an interesting alternative for efficient representation of the leaf. However, its shape diversity makes it particularly difficult to characterize all leaves in a unique way (i.e., using a unique basic model) especially for identification purposes. Leaves from the same species can undergo different shape transformations due to local context, such as location, climatic conditions and age, as shown in Chapter 1. Sometimes, only subtle local features in some meaningful regions, such as the basal or the apical parts, can distinguish between similar species.

3 IdKeys

In the manual process, botanists generally select one or several plant organs (e.g., leaves, flowers or fruits) from a single plant and use *identification keys* (IdKeys) [40] which are examined sequentially and adaptively to identify the unknown species. In essence, one is posing and answering a series of questions about plant attributes (e.g, shape, color, distinguishing landmarks, internal structure) with the aim of focusing on the most discriminating features and narrowing down the set of possible species, much like a game of *twenty questions*. In this section, we focus on leaves and introduce a novel representation based on a hierarchy of botanical IdKeys. The bulk of this section appeared in [100].

3.1 Motivation

Exploiting domain-specific knowledge enables automatic systems to capture subtle characteristics, structure the search and mimic the *process* of identification described by botanists. More specifically, a large amount of information about the taxonomic identity of a plant is contained in its leaves. This is due to the fact that leaves are present on the plants for at least several months, which is not generally the case for other organs such as fruits or flowers. This is why most of plant identification work based on image data, including ours, uses leaf image databases.

Moreover, the motivation behind the specific idea of using IdKeys to represent the leaf characteristics, specifically a hierarchical representation of IdKeys, is the analogy with the game of twenty questions where the keys are the attributes we query before making a final guess about the species.

3.2 Hierarchical representation

We represent a leaf by an ordered set of attributes corresponding to IdKeys. In particular, these keys must encode invariable characteristics, i.e., be independent of the context, such as geography, climatic conditions, season and instantiation. They could refer to geometric properties not directly observable, such as landmarks, or to pre-determined groups of species, such as families and genera. More formally, they can be seen as auxiliary hidden variables which facilitate estimating the primary hidden variable, namely the species itself.

We apply the strategy to leaves, but it could be adapted to other botanical organs such as flowers or fruits and even more general biological entities given an appropriate taxonomy and well-defined IdKeys. Note that all organisms present a hierarchical taxonomy (family-genus-species) as well as natural well-defined and named key points.

Let $\mathcal{K} = \{\mathcal{K}_1, \cdots, \mathcal{K}_N\}$ denote the set of keys with $\mathcal{K}_i$ assuming values in Θ_i, and hence $\mathcal{K} \in \Theta = \prod_{i=1}^{N} \Theta_i$. We assume every instance I (a leaf image) of every class Y (the species of I) has a well-defined set of keys $\mathcal{K}(I)$, and that determining $Y(I)$ is facilitated by knowing these keys. In fact, estimating the keys at full resolution may not be feasible and even narrowing down the possible values of key $\mathcal{K}_i$ to a subset $\Theta_i^0 \subset \Theta_i$ still simplifies estimating Y. The ordering of the keys determines the search sequence.

Hence the key hierarchy has N levels $\mathcal{L}_i, i = 1, ..., N$, and estimation of the plausible values Θ_i^0 of key $\mathcal{K}_i$ is conditioned on the previously retained values $\{\Theta_1^0, \cdots, \Theta_{i-1}^0\}$. Finally, we build a classifier for Y itself dedicated to estimated keys.

The search strategy is breadth-first, coarse-to-fine: starting from the root, classifiers are executed sequentially and adaptively, and any node classifier is applied if and only if all ancestor classifiers have been performed and are positive. All details about the classification process will be explained in Chapter 4.

In our case, depending on the leaf type Θ_2 (i.e., simple or compound), we consider five or six IdKeys to simplify the species identification. Three landmarks are considered for simple leaves (centroid, base and apex) and four landmarks for compound leaves (centroid, base, terminal apex and second apex); see Figure 3.1. These landmarks are combined with both the leaf type and the leaf genus to construct the IdKey hierarchy. The full "simple" tree then has six (resp. seven for the "compound" tree) levels, five (resp. six) corresponding to the five (resp. six) keys and the sixth (resp. seventh) to the species. For $\Theta_i, i = 3, \ldots, N$, the possible values are discretized; e.g., the landmark locations have a resolution of 5×5 and are restricted to the boundary points. The number of leaf types is actually two (simple and compound) and the number of genera, $|\Theta_N|$, depends on the dataset.

Let $\Gamma = \prod_{l=1}^{N} \Theta_l \times \mathcal{Y}$, which is the complete set of possible leaf hypotheses or descriptions $(\theta_1, ..., \theta_N, y)$, namely IdKey instantiation and species. Let T denote the full tree graph; see Figure 3.3. Associated with every $t \in T$ we have:

- Γ_t: The set of interpretations (or hypotheses) $\Gamma_t \subset \Gamma$ entertained at node t, ranging from very coarse cells near the root (e.g., restricting only the centroid) to fine cells at the terminal nodes (fully specified descriptions).

- H_t: The hypothesis $I \in \Gamma_t$, where I is an image to be classified. The alternative is $H_{(t)} : I \in \Gamma \setminus \Gamma_t$.

- f_t: A classifier mapping images to $\{0, 1\}$, where $f_t(I) = 1$ (respectively, $f_t(I) = 0$) indicates acceptance (resp., rejection) of H_t.

Each of these ingredients will be explained in more detail in the remainder of this section and the next chapter. Level one corresponds to the centroid of the blade (resp.

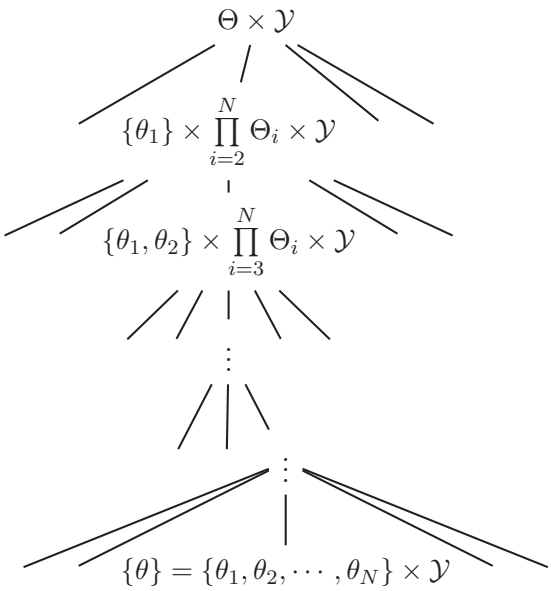


Figure 3.3: Tree-structured hierarchical representation of IdKeys

of the leaflets in the case of a compound leaf), which can be directly calculated from the raw image data once the leaf petiole is removed using the Otsu segmentation algorithm [91] and straightforward post-processing; see Figure 3.4. It should be noted here that we are not concerned by imperfect contours or incomplete petiole removal since our method is robust to such problems. Figure 3.5 illustrates some failure cases.

Each Γ_t at level one is a singleton representing the computed centroid and no test is required. In contrast, the leaf type (level two), the base (level three) and the apex(es) (level four or four and five depending of the results of level two) are all estimated using learned classifiers (which will be defined in Chapter 4). Each landmark detected reduces the number of candidate points for the next detection by excluding its neighborhood from the list of candidates. The whole process is illustrated in Figure 3.6.

Due to the use of pose-indexed features (see §3.3 for more details), only one classifier f_t needs to be learned for each of the $(N-1)$ first levels. In our experiments, we maintain one path through the hierarchy through level $(N-1)$, i.e., we only entertain a single candidate for leaf type and landmarks. For each genus (level N), we learn a dedicated classifier conditioned on the accumulated information. The features are computed in multiple local coordinate systems. Several candidate genera are kept in estimating the

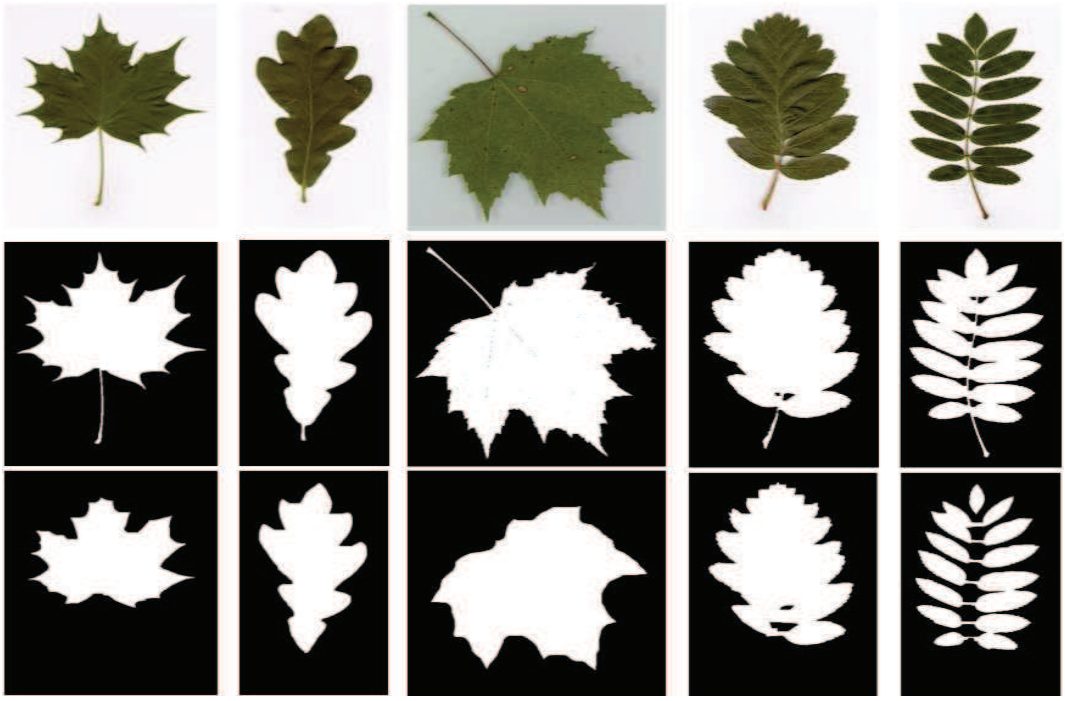


Figure 3.4: Segmentation results with successful petiole removal. Displayed (from the top to the bottom) are an input image, initial segmentation using Otsu algorithm, result after petiole removal.

species, for which we again utilize the same coordinate systems as those used for the genera.

3.3 Feature extraction

As indicated above, the features provided to the learning algorithm are defined in one or more local coordinate systems. We do not use the same frames (i.e., coordinate systems) to estimate the different IdKeys. The motivation is to *focus attention* around each landmark (which is the strategy reported by botanists) and directly extract local features which are invariant to pose, orientation and scale variations, thereby avoiding any need for global image transforms, e.g., geometric normalization. For the leaf type, we use a reference frame which is determined by the estimated centroid and the estimated width of the leaf (radius of the excircle). The axes are parallel to the image borders. For the base and the apex(es), the x-axis is directed towards the centroid of the leaf without

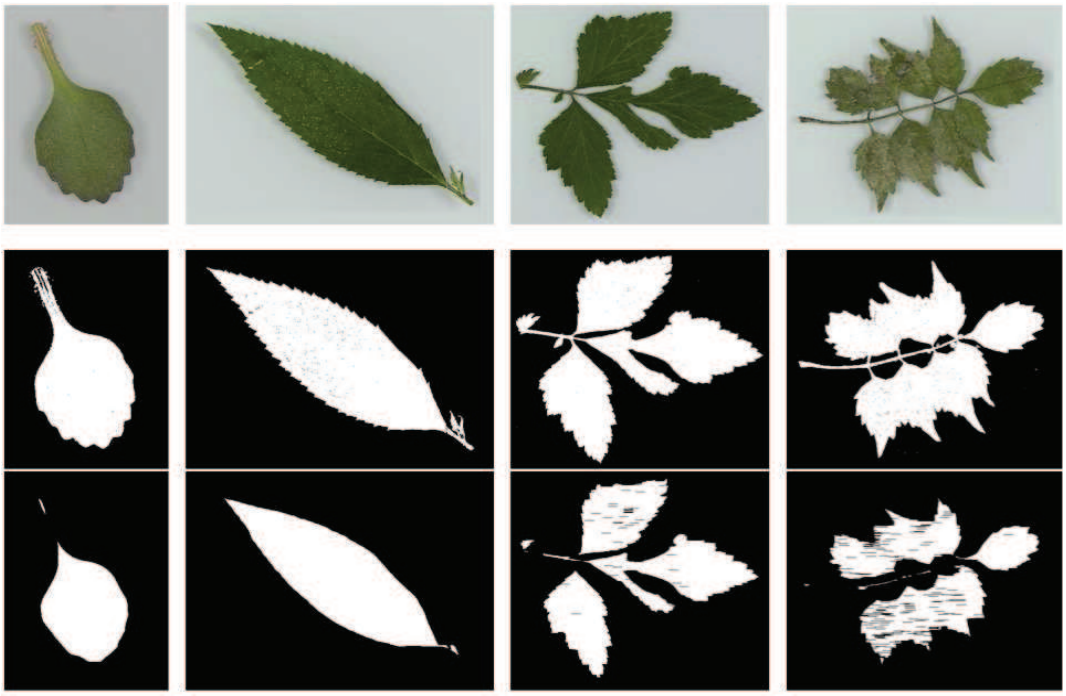


Figure 3.5: Segmentation results with unsuccessful petiole removal. Displayed (from the top to the bottom) are an input image, initial segmentation using Otsu algorithm, result after petiole removal.

the petiole (i.e., centroid of the blade or the leaflets, depending on the leaf type); see Figure 3.7. However, for both the genera and the species, multiple frames are used: two frames for simple leaves, one centered on the apex and the other on the base, and three frames for compound leaves, one centered on the terminal apex, another on the second apex and the last one on the base.

Focusing of this nature is enabled by "pose-indexed" (or "frame-indexed") features Z, introduced in [49] for detecting cats. Although we have many categories of deformable objects, the class of features we use is essentially the same and we refer to [49] for details. Basically, given a frame consisting of two distinguished points and a distinguished scale, there is a candidate feature $Z = Z(w, j)$ for each (local) window w and for each local image property j: the feature Z is just the property histogram in w. Figure 3.8 illustrates (local) windows which were used to extract features for respectively leaf type, base and apex estimation. We use only shape and texture as properties;

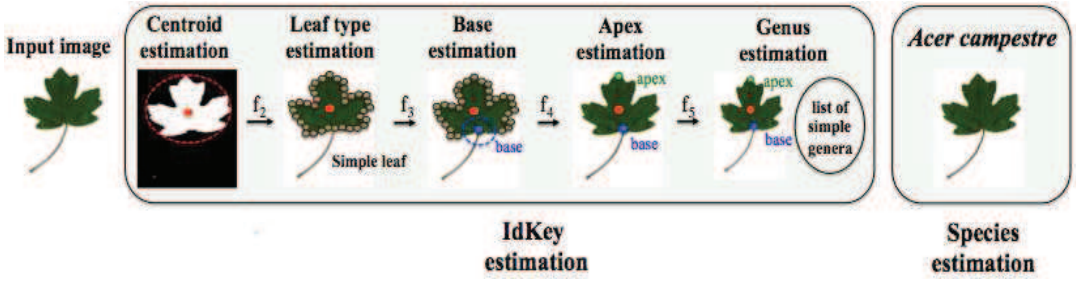


Figure 3.6: Botanical idKeys estimation for species identification.

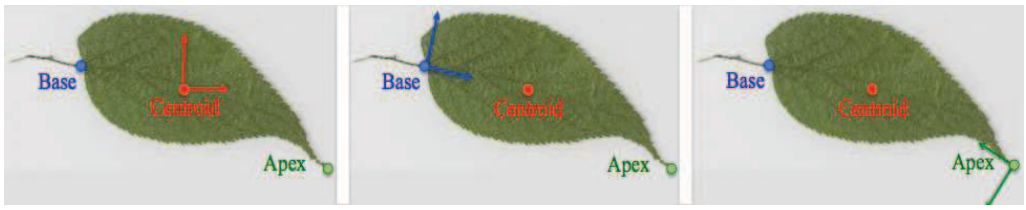


Figure 3.7: Local coordinate systems used for respectively (from left to right) leaf type, base and apex estimations.

specifically, we used Hough, EOH and Fourier histograms [46] as base features. We will see in the next chapter how such features were used to induce classifiers.

4 Vantage feature frames

In this section, another novel object representation, *vantage feature frames*, is introduced for botanical species identification, in particular, and fine-grained recognition, in general. We first discuss the interest and the motivation behind this idea in §4.1. Then, we explain the concept of *vantage feature frame* in §4.2. We apply this representation on both leaves and orchid flowers. The bulk of this section appeared in [101].

4.1 Motivation

Even if idKeys could be generalized and adapted to other objects, it could take considerable effort to define an ordered, coherent set that permits accurate identification. That is why we focus on more generic representation, but still motivated by the strategy used by botanists. Unlike the idKey representation, here we only focus on specific landmarks (i.e., geometric properties) with the aim to learn best frames and best features for each

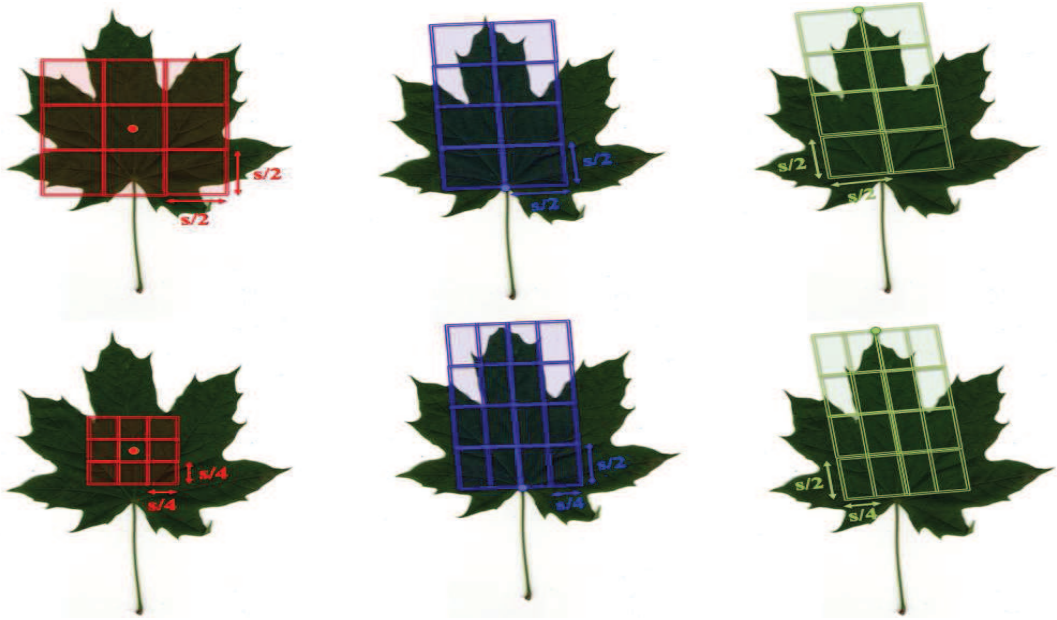


Figure 3.8: Multi-scale (local) windows defined relative to local coordinate systems. Displayed are those used to extract features for respectively leaf type (in red), base (in blue) and apex (in green) estimation. s refers to the approximative width of the leaf.

frame. The idea is to focus attention on visual properties of the object in the vicinity of a small number of distinguished points and whereas these landmarks are the same for each species, it is the local features which permit disambiguation. Both aspects are important: *where to look* and *what to compute*. The vehicle for translating this into a computer vision algorithm is the notion of a *vantage feature frame*.

4.2 Definition

Let $\{C_1, ..., C_n\}$ denote n categories, where a category can refer to a single species or a group of species. In the botanical applications, which motivate this work, there is often useful domain knowledge, typically named landmarks $L = \{l_1, ..., l_K\}$ around which botanists focus in order to separate one species from another (see Figure 3.9). Such landmarks are more like *vantage points* in that orientation plays a role as well, in other words, where the landmarks are in relation to one another. Naturally, species (or some groups of species) tend to have certain signature appearance properties and consequently what to look for in the neighborhood of the landmarks may be species-

dependent. Put differently, the conditional distribution over any large family of generic local features may depend strongly on the species or groups of species. This aspect of the identification process can be encoded by allowing the set of features associated with each landmark to depend on the category. We also want to ensure that the local appearance properties are largely invariant to the orientation and scale of the object.

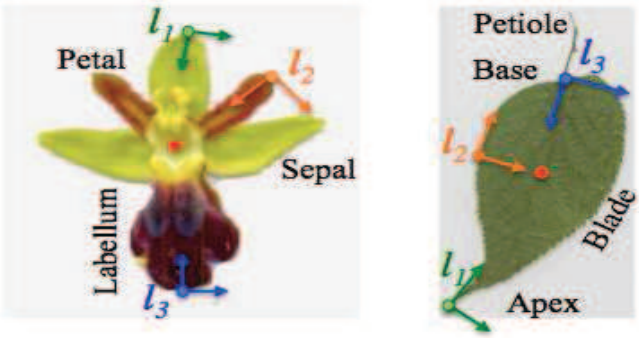


Figure 3.9: Candidate frames for orchids and leaves.

With these considerations in mind, a *vantage feature frame* $\mathcal{F}$ has two components. One, Ω, is geometric and the other, $\mathcal{Z}$, is appearance-based. The geometric component Ω is category-independent and simply a local coordinate system centered at one of the landmarks l. Since we are dealing with images of single objects (e.g., images of leaves) we declare the orientation to be determined by the centroid of the object, that is, the landmark points to the centroid, and the unit distance to be the approximate scale of the object.

The appearance component is a family of *pose-indexed* features, one element of the family for each category: $\mathcal{Z} = \{\mathcal{Z}_1, \ldots, \mathcal{Z}_N\}$, where $\mathcal{Z}_t$ is the set of local features to compute in frame $\mathcal{F}$ for a specific category. Again, the reason for dedicated features is that there is so much variability in the presentation of leaves in the neighborhood of landmarks that some features are far more discriminating than others, and the discriminating ones can depend as well on the vantage point. For example, the discriminating features around the leaf base for estimating a particular group of species might be different from those around the apex for estimating another group. Obviously, to be useful the frame must be reliably detected and the features must be discriminating. As will be seen in ensuing sections, we provide algorithms for learning discriminating ones, de-

tecting them online and pooling the features computed in these frames to identify the categories.

4.3 Learning the frames

Learning the most discriminating frames from scratch would evidently be a major challenge, and we do not attempt this. As indicated above, by leveraging domain knowledge, we begin with a list of candidate frame origins $l_1, ..., l_K$. There will be frames associated with a subset of these. Moreover, since we are dealing with images of single objects (e.g., scanned images of leaves) we declare the orientation of the frame to be determined by the centroid of the object, that is, the landmark points to the centroid, and the unit distance to be the approximate scale of the object. The choice of landmarks or vantage points is performance-based. Assume we are given a classifier for each set of vantage feature frames; our particular choice for leaves is described in §4.6 and for orchid flowers in §4.7. Given $|\mathcal{L}| = K$ candidate landmarks, there are then 2^{K-1} possible set of coordinate systems. Evaluating them one-by-one might be infeasible, in which case one might adopt a greedy strategy: the efficiency of each candidate could be measured by the improvement in the overall classification rate obtained by adding the corresponding frame to the existing list of frames.

For leaves and orchids only three "universal" landmarks $\mathcal{L} = \{l_1, l_2, l_3\}$ have been suggested by botanists; they are described in §4.6 and §4.7 and illustrated in Figure 3.9. For each of the $2^3 - 1 = 7$ combinations of frames, we estimated the classification accuracy using cross-validation. Feature extraction and classification are described in §4.5 and chapter 4 respectively. It should be noted that for this learning process the locations of the landmarks were determined by manually annotating the training data. As a result, the errors that are inevitably made in automatically detecting the landmarks are not taken into account in choosing the best set of frames. One might expect that the more frames the better the performance, and hence using all three would be optimal. However, this was not the case; Table 3.1 shows the recognition rates for the seven possible combinations of frames used for simple leaves. The best performance is obtained with two frames corresponding to l_1 and l_3.

Set of coordinate systems	l_1	l_2	l_3	l_1, l_2	l_1, l_3	l_2, l_3	l_1, l_2, l_3
Recognition rate	0.75	0.72	0.73	0.76	**0.8**	0.77	0.78

Table 3.1: Cross-validated recognition rates for leaves (Smithsonian data §1) for each of seven possible sets of frames sets with centers l_1, l_2, l_3. The best result (in bold) is obtained with two frames centered at the base l_1 and apex l_3.

4.4 Detecting the frames

The first step in classifying an image is to estimate the location, orientation and scale of each frame. As indicated above, the orientation is determined by the centroid, which is directly computed from the raw image data after a segmentation process using the Otsu algorithm [91]. The scale is taken to be the radius of the bounding circle. The landmarks are detected by dedicated classifiers trained on manually annotated images. Since we are only using landmarks on the object boundaries (as determined by the segmentation process), we restrict the search to a sample of boundary points to minimize the computation. In addition, after detecting each landmark, we exclude the boundary points in its neighborhood from the list of candidates; see Figure 3.10.

In order to detect each vantage point, a classifier (see Chapter 4) based on SVM scores is built from positive and negative training examples. Positive images are annotated by the landmark considered and negative images are randomly annotated. The features for SVM learning are defined in the local coordinate system centered on the candidate landmarks (i.e., the x-axis is directed towards the centroid as described above). More specifically, the feature extraction follows the same process described earlier for IdKeys (see §3.3).

4.5 Learning the features

The appearance-based component is category-dependent. Whereas we use the same class of features to learn landmark detectors, we construct a separate binary classifier for each category C_t for distinguishing that category from all others and which employs a learned subset of features $\mathcal{Z}_t$. Hence, we select a category-dependent subset of features $\mathcal{X}_t$ and only these are used to train classifiers.

Specifically, we first estimate the probability distribution of each feature Z under

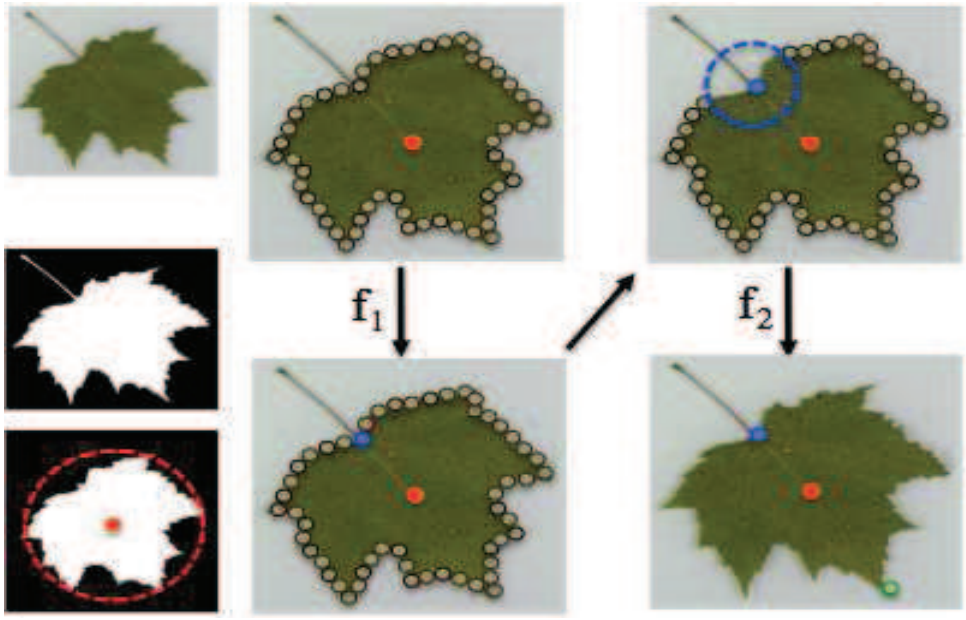


Figure 3.10: A test leaf image is first segmented. Then the petiole is removed in order to compute the centroid (red point) as well as the approximate bounding circle of the leaf blade (red dashed circle). The base (blue point) and the apex (green point) are estimated using learned classifiers (f_1, f_2). The proposed locations for both landmarks are restricted to the boundary points. The neighborhood of the first landmark detected is excluded from the list of candidate points for the next detection (blue dashed circle).

both hypotheses $Y \in C_t$ and $Y \notin C_t$ (where Y is the actual species of the image be classified) from the positive and negative examples. For each distinct category, images belonging to that category are positive and all others negative. For feature $Z(w, j)$, denote the two distributions by $p^+_{w,j}$ and $p^-_{w,j}$ and let $d_{w,j} = |p^+_{w,j} - p^-_{w,j}|$ be the difference in the L1 norm. Then $\mathcal{Z}_t$ consists of the features with the M largest differences. Figure 3.11 illustrates the recognition rate for leaf genera for various M. Selecting category-dependent features increases recognition performance and decreases computation. For instance, we achieve over 75% recognition rate of leaf genus while considering only the first genus returned and using between about 500 and 2500 category-dependent features against only 67% without any selection i.e, $M = 5808$ (see Figure 3.11).

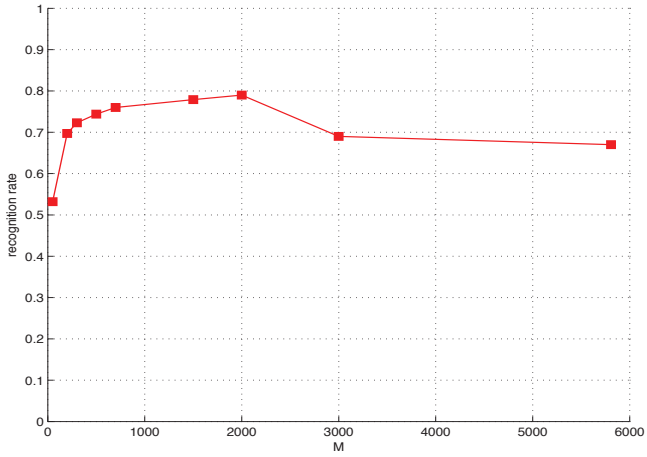


Figure 3.11: Recognition rates for leaf genera from the Smithsonian data (see §1) while considering only the first genus returned and using M selected features.

4.6 Case of leaves

To analyze leaves, experts usually focus on the apex, the base and the leaf margin. These are illustrated in Figure 3.9. Let l_1 denote the leaf apex (respectively, the central sepal for orchids), l_2 the first intersection point between the perpendicular to the apex-base line throughout the centroid of the blade and the leaf boundary and l_3 the leaf base as shown in Figure 3.9. The centroid of a leaf corresponds to the center of mass of the blade (resp. of the leaflets in the case of a compound leaf); the leaf petiole is removed before computing the centroid (see Figure 3.10). As for the segmentation process, again, we are not concerned by imperfect contours or incomplete petiole removal since our method is robust to such problems. Figure 3.10 illustrates the vantage point detection process for a leaf image, namely the leaf base and the leaf apex detection. Random sample of test images with the estimated vantage points for different type of leaves (e.g., toothed, lobed, concave, convex, symmetric, asymmetric) and different datasets are displayed in Figure 3.12. More details about such datasets as well as more qualitative and quantitative results will be shown in Chapter 6.

Figure 3.13 illustrates examples from the most discriminating local windows w for four species. The original set of w used for the selection represents all windows defined earlier for landmark detection, i.e., those which were illustrated in Figure 3.8 for apex and base estimation.

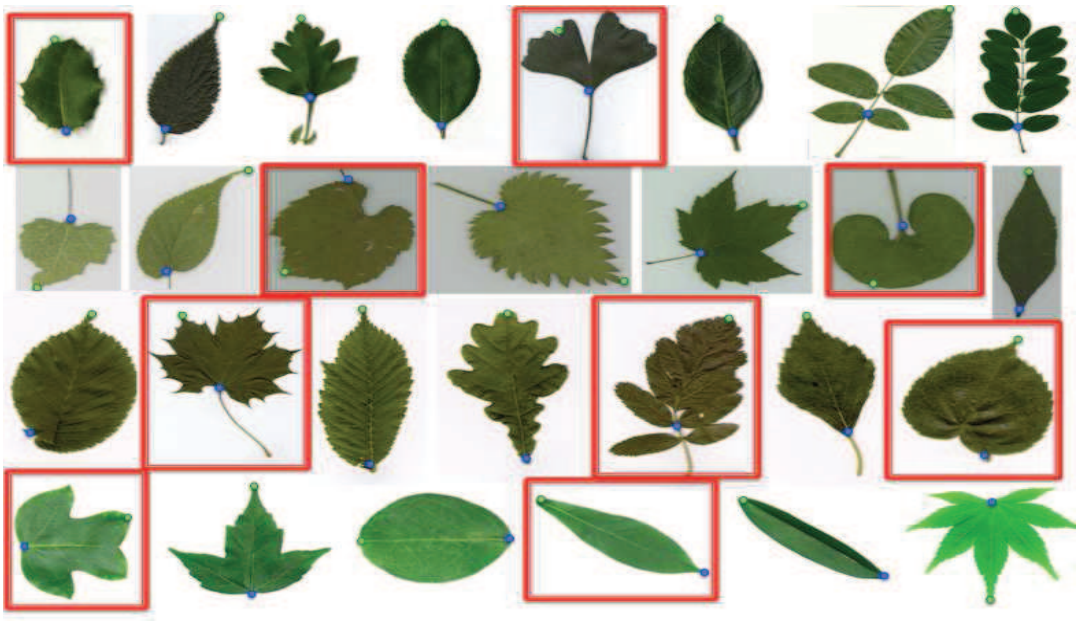


Figure 3.12: Random sample of test images with the estimated vantage points for four different leaf datasets. False detections are framed with a red box. Note that the entire detection process is considered erroneous if any vantage point is not accurately detected.

4.7 Case of orchid flowers

An orchid specialist generally focuses on the sepals, petals and the labellum. These are illustrated in Figure 3.9. As with leaves, let l_1 denote the central sepal for orchids, l_2 the petal on the right of l_1 and l_3 the bottom of the orchid labellum as shown in Figure 3.9. Again, we are not concerned by imperfect contours after the segmentation process. Figure 3.15 illustrates examples from the most discriminating local windows w for four orchid species. As with leaves, quantitative results will be shown in Chapter 6.

5 Summary

The different object representations described through this chapter are designed to *separate* species using their most discriminating characteristics. The hierarchy of IdKeys, introduced above, focuses on botanical properties. It is dedicated to leaves but could be extended to other objects using the same principle. The Vantage Feature Frame is a more generic representation which investigates both aspects, category-independent

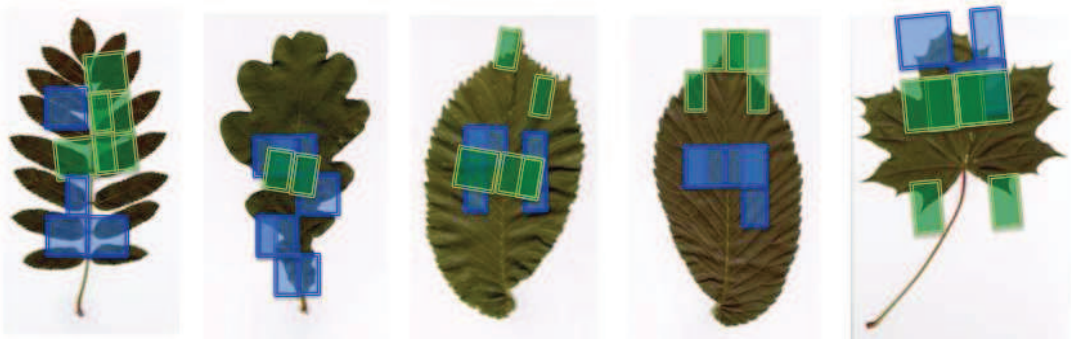


Figure 3.13: Examples of the five most discriminating local windows for different species. Blue boxes refer to local windows relative to the coordinate system centered in the leaf base and green ones refers to those relative to the coordinate system centered in the leaf apex.

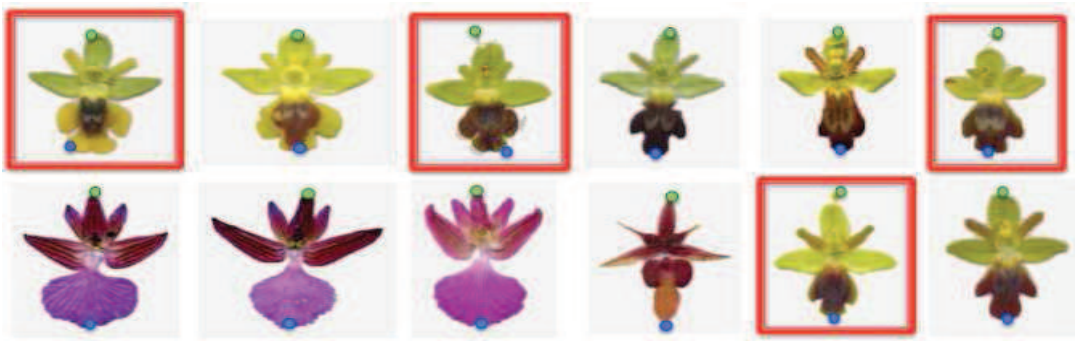


Figure 3.14: Random sample of test images with the estimated vantage points for orchid flowers. False detections are framed with a red box. Note that the entire detection process is considered erroneous if any vantage point is not accurately detected.

and category-dependent description. The large inter-class similarity as well as the large intra-class variability between fine-grained categories, especially botanical species, makes such a local detailed image description very important in order to induce efficient classifiers.

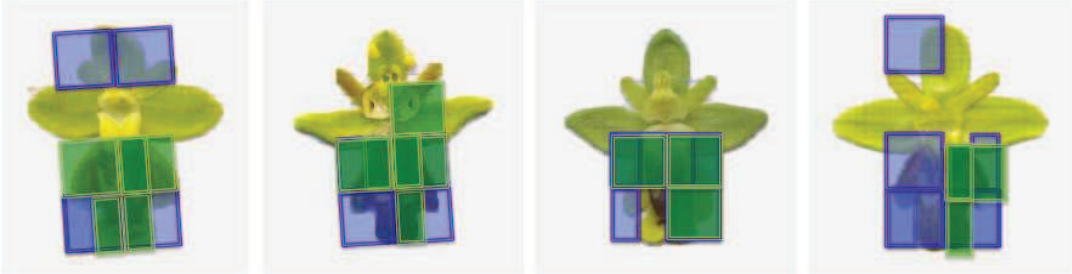


Figure 3.15: Examples of the five most discriminating local windows for different species. Blue boxes refer to local windows relative to the coordinate system centered in the bottom of the orchid labellum and green ones refers to those relative to the coordinate system centered in the central sepal.

Chapter 4

Classification

1 Introduction

Multi-class classification is one of the core problems in many applications. It refers to assigning each of the observations into one of possibly many categories. In this work we investigate two ways to address this problem. We first focus on a coarse-to-fine approach using a likelihood ratio framework; see §4. Second, we introduce a model-based approach, focusing on selecting a *confidence set* (CS), i.e., a variable-length list of categories which contains the true one with high probability (e.g., 99% CS); see §5. Both approaches are based on a hierarchical representation of the full set of categories. This tree of subsets indexes a graded family of categories of varying sizes as well as local discriminant functions for deciding between each subset and all others combined. These ingredients are described respectively in §2 and §3.

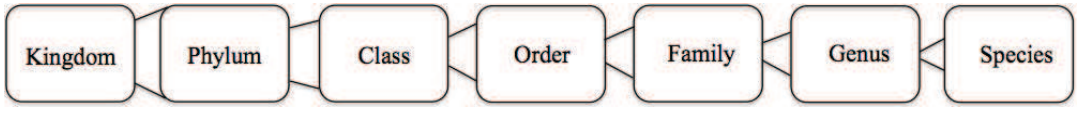


Figure 4.1: Biological classification. There are seven main taxonomic ranks defined by the international nomenclature codes: kingdom, phylum/division, class, order, family, genus, species.

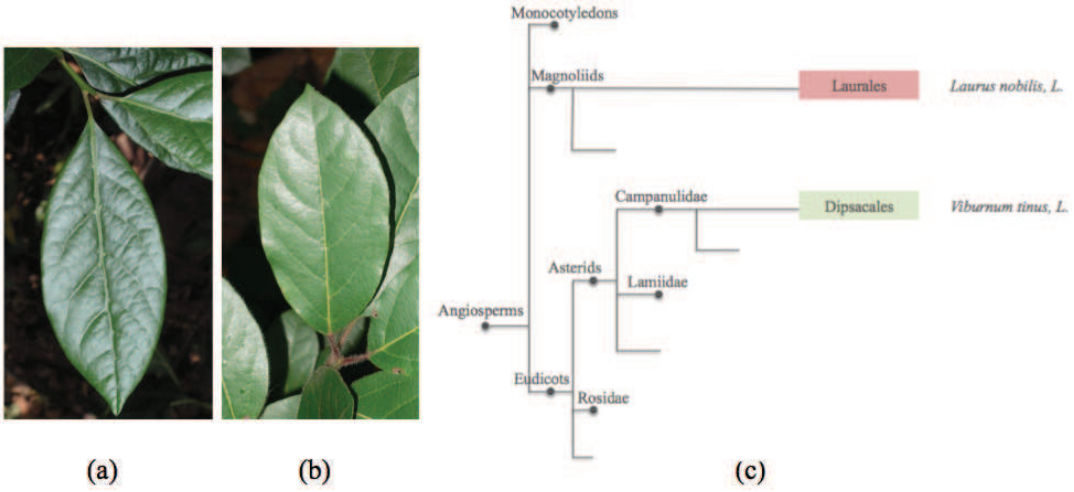


Figure 4.2: Leaves of (a) *Laurus nobilis, L.* and (b) *Viburnum tinus, L.* species and (c) their respective positions in the APG classification. Note the visual similarity between the two leaf shapes despite the large distance between the species in the hierarchy (the first common ancestor of the two species is the root node).

2 Hierarchy construction

Botanical species are naturally organized in a hierarchical taxonomy; see Figure 4.1. Different taxonomic systems are used for different kind of plants. For instance, the APG III system [53] is currently recognized by most botanists for flowering plant (angiosperms) classification and is mostly a molecular-based system of plant taxonomy, as shown in Figure 4.2. Hierarchical representation can allow for classification algorithms or cost metrics that penalize misclassification errors according to "closeness" in the hierarchy. For example, it may not be as problematic which fir tree it is as long as we do not confuse it with other non-related trees. Hierarchical representation is hardly unique; there are a great many ways to recursively decompose the data. For instance, the hierarchy can be manually designed using pre-defined biological groups or automatically built using morphological characteristics; see Figure 1.14. While the former is defined according to both shared physical and genetic characteristics and should be provided by experts, the latter is based on purely visual characteristics and can result from a hierarchical clustering on training data. We use both kinds of hierarchies for classification.

More specifically, we first use a pre-defined two-level taxonomic hierarchy, especially with IdKeys: the first level represents the genera and the second represents the species. Then we apply an agglomerative clustering principle to automatically generate a visual-based hierarchy.

Agglomerative procedures are probably the most widely used of the hierarchical clustering techniques (a useful review of the standard techniques has been given in [61]). They produce a series of partitions of the data: the finest consists of single-member 'clusters' and the coarsest consists of a single group containing all individuals. Variations are based on domain knowledge about botanical species and landmarks, but the principle is quite general: a tree-structured hierarchy is recursively constructed bottom-up by successively merging similar groups. We treat each species as a singleton cluster at the outset and then successively merge (or agglomerate) pairs of clusters until all the clusters have been merged into a single cluster that contains all species using Ward's criterion [115] and the Euclidean norm. The dissimilarity between two clusters is given by

$$dissim(r,s) = \frac{n_r * n_s}{n_r + n_s} \times ||\bar{Z}_r - \bar{Z}_s||^2, \tag{4.1}$$

where r and s denote two specific clusters, n_r and n_s denote the sizes of the two clusters, $\bar{Z}_r$ and $\bar{Z}_s$ denote the centers of gravity of the clusters and $||.||$ is the Euclidean norm. Local features are used to compute the centers of gravity of the clusters. More specifically, texture and shape-based features are used to characterize each leaf image and are defined in two local coordinate systems, one centered on the leaf base and the other on the leaf apex. Thus, we use, the same local features described in §3.3 to characterize species.

Visualizing this tree-structured hierarchy provides a useful summary of the data, i.e., an overview of the visual similarities and relationships among species based on both the basal and the apical parts. Figure 4.3 represents a dendrogram [48] that illustrates the nested grouping of the species produced by a hierarchical clustering on 50 botanical species. Note that many clusters obtained could be matched with morphological classes defined by botanists themselves. In particular, two large, natural clusters are formed at the first level of the hierarchy: one cluster consists of compound and lobed leaves (on the left) and the other cluster of simple leaves (on the right). Species with lobed leaves

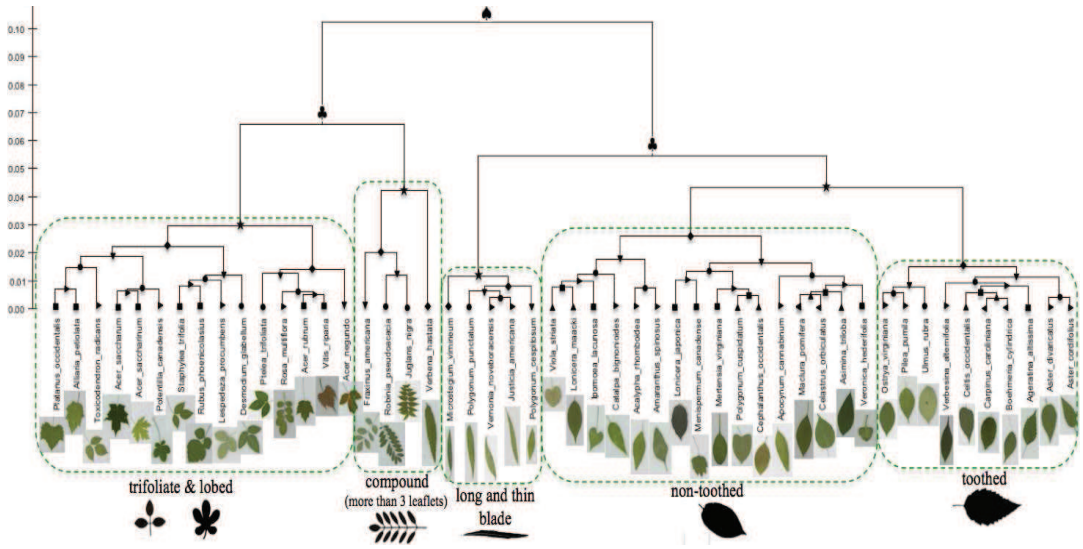


Figure 4.3: A dendrogram representing a hierarchical clustering of 50 species of Smithsonian leaves. Displayed are the nested groupings of species, similarity levels at which groupings change, and a thumbnail from each species. Many clusters match morphological classes.

(with 3 lobes) merge with those with trifoliate leaves (with 3 leaflets) while species with compound leaves containing more leaflets are grouped together. Also, toothed leaves are separated from non-toothed leaves. Therefore, such a hierarchical clustering could even help botanists in classifying large amounts of newly collected leaves by suggesting coarse morphological categories.

3 Discriminant functions

The hierarchical representation will serve as a platform for the classification algorithm. To this end, we learn local discriminant functions at each node $t \in \mathcal{T}$. The framework is largely classifier-independent in that any learning algorithm could be chosen to induce such discriminant functions from the training data. We have chosen to train a Support Vector Machine (SVM) classifier for $Y \in C_t$ versus $Y \notin C_t$. The SVM is a widely used classification method, introduced in [11], which belongs to the general category of *kernel methods* [102]. SVMs can be either linear or nonlinear. Some nonlinear kernels include

polynomial, radial basis function (RBF) and sigmoid. We use RBF kernel

$$K(x, x') = \exp(-\gamma||x - x'||^2).$$

Let X_t be the SVM score associated with the node t and learned from training data. We approximate the probability distribution of X_t by a Gaussian density Φ_t using estimated mean μ and variance σ^2. Figure 4.4 shows examples of the empirical distributions of X_t (i.e., histograms obtained from real data) for different nodes and categories. Clearly the shape varies considerably and certainly has higher variation or entropy than a Gaussian. We use the Gaussian nonetheless since we have sufficient data to reliably estimate the mean and variance and the Gaussian has maximum entropy for a fixed mean and variance.

4 Coarse-to-fine search and likelihood framework

Coarse-to-fine (CTF) classification is an efficient way of organizing object recognition in order to systematically exploit shared attributes and the hierarchical nature of the data. Here, we apply CTF search within a likelihood framework.

4.1 Coarse-to-fine search

The basic structure of a CTF search is a nested representation of the space of hypotheses and a corresponding hierarchy of (binary) classifiers with a steady progression from very general classifiers with low resolution in coarse-grained categories to those dedicated to fine-grained categories. When properly designed, the fine classifiers are rarely evaluated.

We use breadth-first, CTF strategy. That is, starting from the root, the classifiers are executed sequentially and adaptively, and a classifier is executed if and only if all ancestor classifiers are performed and are positive. This promotes extremely efficient computation since, generally, large subsets of hypotheses are simultaneously pruned. It should be emphasized that a CTF hierarchy is not a decision tree. In fact, unlike a decision tree, during a CTF search a data point may traverse many branches and may reach no leaves, i.e., a tested image may arrive at no leaves or more than one leaf in the tree.

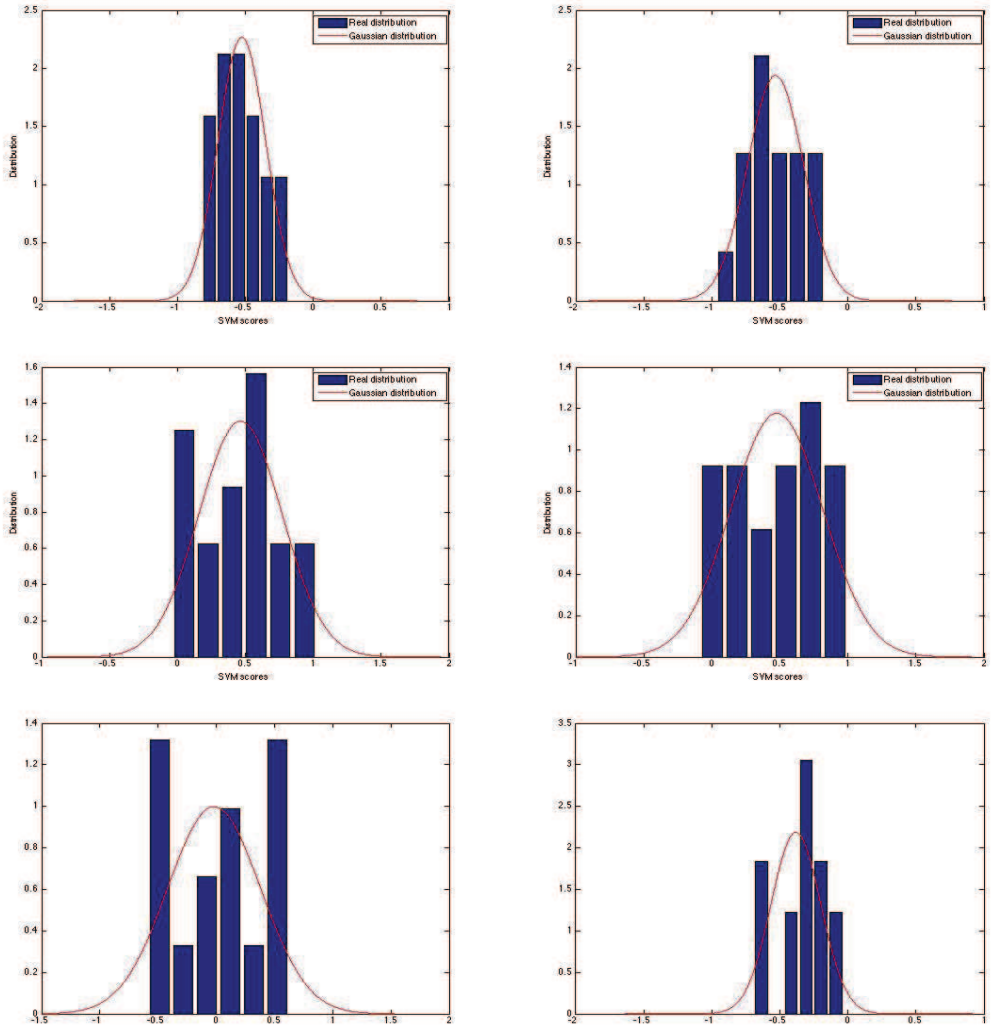


Figure 4.4: The empirical distributions of local SVM scores of different hierarchical nodes of Figure 4.3. Each distribution is approximated by a Gaussian density (in red) with the estimated means and variances.

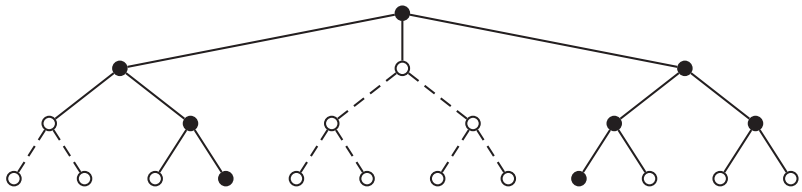


Figure 4.5: Example of a three-level hierarchy using three IdKeys. Two of the possible values of the first key are kept (full nodes at the first level). Then we keep three possible values for the second key. Note that we are not keeping here all the combinations of these two and three values. However, we are keeping exactly three combinations (corresponding to the paths formed by the full nodes except the root). Finally we keep only two paths which correspond to two triplets (each triplet corresponds to the retained values of the full nodes that form the path). Only these paths are considered for the species identification task.

For example, Figure 4.5 illustrates the search based on three Idkeys. At the end of the search, only two paths, i.e., two elements of the full Idkey space, are considered for the species identification task. At each node t of the hierarchy, let $f_t = f_t(I)$ be the binary classifier which is designed to separate each node t and all others combined based on the features extracted from I. At each level, all the children of a *positive* node t (i.e., one for which $f_t(I) = 1$) are retained and tested at the next level. Whereas false positives can be successively pruned, if the true hypothesis is rejected at a node containing it then it cannot be recovered. In the case of the IdKey representation, described in the previous chapter, only the classifiers for species which belong to the retained genera are performed. Finally, those species for which $f_t(I) = 1$ are then sorted according to their likelihood ratios (see 4.2).

4.2 Likelihood ratios

In order to induce f_t from the training data at node t, we consider a likelihood framework using local discriminant functions, i.e., SVM scores X_t. Let $\Phi_t(x_t)$ be the (Gaussian) density of X_t. The corresponding classifier f_t is then based on the likelihood ratio as follows:

$$L_t(I) = \frac{\Phi_t(X_t(I)|Y \in C_t)}{\Phi_t(X_t(I)|Y \notin C_t)} \tag{4.2}$$

Table 4.1: Example of 5-fold cross validation to set ρ_t for an internal node. All of $\rho_t = \{-3.5, -3, -2.5\}$ achieve the highest true positive rate. The largest value is kept, i.e., $\rho_t = -2.5$.

ρ_t	-3	-2.5	-2	-1.5	-1	-0.5	0	0.5	1	2	3
Fold 1	1.0	1.0	0.98	0.98	0.98	0.96	0.94	0.94	0.91	0.91	0.89
Fold 2	1.0	1.0	1.0	1.0	0.98	0.98	0.98	0.98	0.98	0.98	0.96
Fold 3	1.0	1.0	1.0	1.0	0.97	0.97	0.94	0.88	0.85	0.79	0.66
Fold 4	1.0	1.0	0.98	0.98	0.97	0.95	0.94	0.93	0.9	0.9	0.88
Fold 5	1.0	1.0	0.99	0.98	0.96	0.96	0.94	0.94	0.91	0.91	0.91
Average	**1.0**	**1.0**	0.99	0.99	0.97	0.96	0.95	0.93	0.91	0.9	0.86

Given $L_t(I)$, let

$$f_t(I) = \begin{cases} 1 & \text{if } \log(L_t(I)) > \rho_t \\ 0 & \text{else} \end{cases} \tag{4.3}$$

Here, ρ_t is a learned node-dependent threshold used to control the false negative rate, that is to allow only a very small number of instances in which $Y \in C_t$ but $f_t(I) = 0$ (missed detections). This can be accomplished at the expense of (temporary) low specificity (i.e., a high false positive rate), but this is a favorable tradeoff in our context.

In our setting, k-fold cross-validation is used during the training stage to set ρ_t for each node of the hierarchy. The training set is randomly partitioned into k equal size subsets. Of the k subsets, a single subset is retained as the validation data for testing, and the remaining $k-1$ subsets are used as training data. The cross-validation process is then repeated k times, with each of the k subsets used exactly once as the validation data. Each time, the true positive rate is computed on the validation data using different values for ρ_t. For each node, we retain the largest value that permits the highest true positive rate in average. The candidates for ρ range from -5 to 5 by step of 0.5, and k ranges from three to five depending on the dataset sizes. Table 4.1 shows an example of the true positive rates for different values of ρ_t. In this case, we retain $\rho_t = -2.5$. Negative values promotes low missed detection rates.

The same likelihood framework is used for IdKey estimation. In this case, we just have to consider Γ_t instead of C_t while defining the likelihood ratio L_t. However, for $l = 1, ..., (N-1)$ (i.e., the leaf type and landmark levels) all the nodes t at level l share the same classifier f_l. Also, only a single estimate is retained, namely the one

corresponding to the node t at which the likelihood ratio L_t is maximized. In contrast, for the genera and species, the classifier is in fact node-dependent. *All* the estimated genera ($f_g(I) = 1$) are considered for species identification.

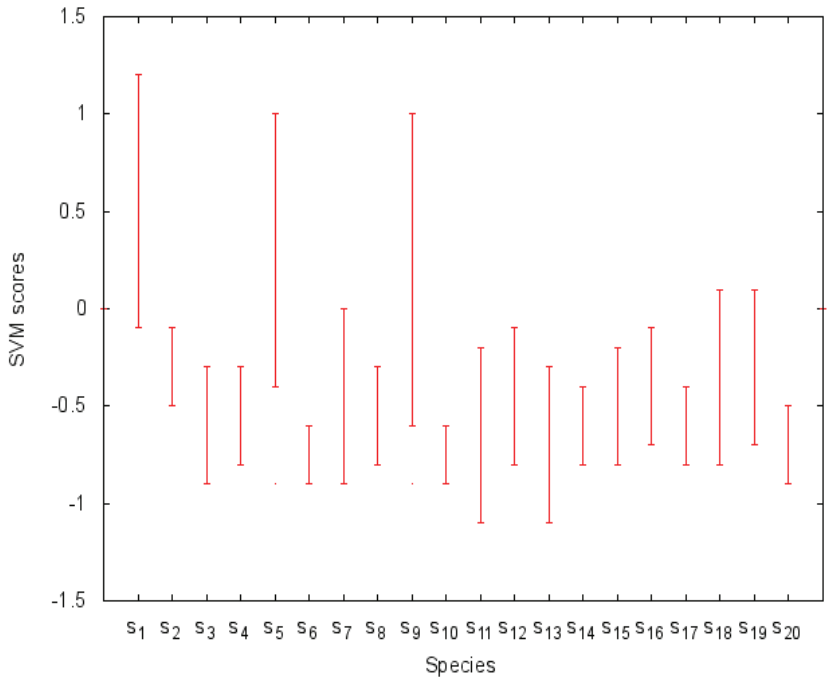


Figure 4.6: SVM score ranges for twenty different species. Often scores are on different scales.

Several work in the literature used SVMs to produce probabilities either for posterior probability-based decision making [58, 95] or likelihood-based decision-making [88, 107]. Here, the motivation behind using likelihood ratios based on SVM scores instead of simply SVM scores alone is that SVM scores which are associated to different C_t might naturally occur on different scales and thus comparisons among them could be arbitrary. For example, figure 4.6 illustrates the SVM score ranges for a random set of species. Note that score ranges could be very different. Also, the mapping of SVM score to a likelihood ratio takes into account the distribution under both hypotheses. In particular, this mapping is *not* monotone, i.e, does not preserve the ordering of SVM scores across a level. This is illustrated in Figure 4.7, which shows two pairs of distributions for two categories C_1 and C_2. The dashed red and blue lines correspond respectively to the

SVM score distribution of images in C_1 and in the complement of C_1. The solid red and blue lines correspond respectively to the SVM score distribution of images in C_2 and in the complement of C_2. For a test image I, we show in black the SVM score $X_1(I)$ associated to C_1 as well as the densities of $X_1(I)$ under both hypothesis: $Y \in C_1$ and $Y \notin C_1$ (i.e., $\Phi_1(X_1(I)|Y \in C_1)$ and $\Phi_1(X_1(I)|Y \notin C_1)$). Likewise, we show in green the SVM score $X_2(I)$ associated to C_2 as well as the densities of $X_2(I)$ under both hypothesis: $Y \in C_2$ and $Y \notin C_2$ (i.e., $\Phi_2(X_2(I)|Y \in C_2)$ and $\Phi_2(X_2(I)|Y \notin C_2)$). I would be classified as C_1 using SVM scores, since $X_1(I) \geq X_2(I)$ (see the black and the green point at the x-axis), but as C_2 using likelihood ratios.

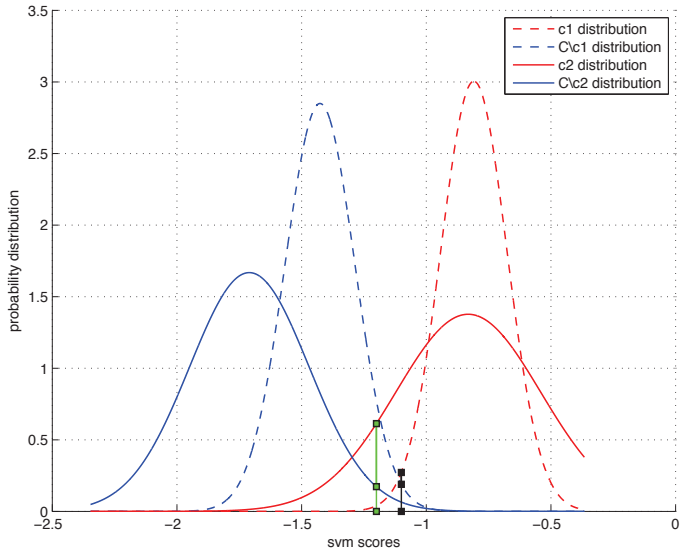


Figure 4.7: Comparison between the SVM score distributions of two different categories. Both distributions are approximated by Gaussians.

5 Confidence sets

In classical (frequentist) statistics, a confidence interval CI [89] is a data-dependent interval estimate of a single population parameter. For example, the CI provides an indication of the precision of a point estimate such as maximum likelihood. Precision corresponds to the length of the CI and the confidence level can be interpreted as the fraction of times in repeated experiments that this random interval would contain the

true parameter. A confidence set (CS) refers to an extension of confidence intervals to a multidimensional parameter [24]. Bayesian CI's and CS's [71] extend these notions to Bayesian statistics wherein a prior distribution over parameters combined with a data model leads to a posterior distribution over parameters given the observations, interpreting the confidence level as a posterior probability. In analogy with these classical tools, we propose a semi-automated system which outputs a CS, a variable-length list of categories which contains the true one with high probability, rather than providing a point estimate or ranking all candidates.

One straightforward way to generate a CS is model-based: delineate a feature vector $Z = Z(I)$ and *model* for the joint distribution $p(z, c)$ of Z and Y. Provided with this, and given an image I (and hence z) the natural recipe for assembling $\widehat{C}(z)$ would be to compute the posterior distribution $p(c|z)$ over categories and aggregate the masses starting from the largest one, say $p(c_1|z) \geq p(c_2|z) \geq \cdots$, until the cumulative probability passes $1 - \epsilon$. That is, $\widehat{C} = \{c_1, ..., c_k\}$ where $p(c_1|z) + \cdots + p(c_{k-1}|z) < 1 - \epsilon$ and $p(c_1|z) + \cdots + p(c_k|z) \geq 1 - \epsilon$. In principle, the CS can then be any subset of categories. We propose a different strategy which is anchored by a hierarchical representation of $\mathcal{Y}$ and which drastically reduces the space of candidate CS's.

5.1 Statistical model

In this framework, the hierarchy considered is a binary tree which consists of a recursive partitioning of $\mathcal{Y}$. The hierarchy serves as a platform for defining features and for selecting confidence sets. The construction of the hierarchy is based on hierarchical clustering of training data as described in §2. This structure provides a natural family of (visually) closely-related categories with diverse sizes. This argues for *restricting* $\widehat{C}$ to the subsets $\{C_t, t \in \mathcal{T}\}$. In the standard case of returning a single estimate, the selection is restricted to the terminal nodes which are individual species. As before, the data for selecting $\widehat{C}$ is a discriminant function on $\mathcal{T}$. Since the choice of $\widehat{C}$ depends only on the scores $\mathbf{X} = \{X_t, t \in \mathcal{T}\}$, we will sometimes write $\widehat{C}(\mathbf{X})$ to emphasize the dependence on the data.

The modeling is naturally done at the level of $\mathbf{X}$ and Y, thereby integrating all the evidence from the node scores. Let $p(\mathbf{x}, c)$ be a model for the joint distribution

$P(\mathbf{X} = \mathbf{x}, Y = c)$. In order to specify $p(\mathbf{x}, c)$ we fix a prior $p(c)$ over categories (usually uniform); hence the key ingredient is the conditional data distribution $p(\mathbf{x}|c), c \in \mathcal{Y}$. (Note that the score at the root is meaningless since all categories belong to C_{root} and consequently this node can be ignored in what follows.) The components of $\mathbf{x}$ are real-valued and indexed by the tree $\mathcal{T}$; hence the dimension of $\mathbf{x}$ is basically twice the number of categories. The model we use for $p(\mathbf{x}|c)$ in our application is a Bayesian network over Gaussian variables and will be described in detail in §5.2. In brief, the two children t_1 and t_2 of t_{root} serve as roots of the BN, which then has the form:

$$p(\mathbf{x}|c) = f(x_1|c)f(x_2|c) \prod_{t \in \mathcal{T} \setminus \{t_1, t_2\}} f_t(x_t|x_{t\text{-}}, c) \tag{4.4}$$

Here t- denotes the parent of t; $f(x_1|c)$ and $f(x_2|c)$ are the marginal densities of scores X_{t_1}, X_{t_2} given $Y = c$, both assumed univariate normal; and $f_t(x_t|x_{t\text{-}}, c)$ is the conditional density of X_t given $\{X_{t\text{-}} = x_{t\text{-}}, Y = c\}$. Since we are assuming $(X_t, X_{t\text{-}})$ is bivariate Gaussian given $Y = c$, the form of the conditional density follows immediately. Again, the details for our application to plants, including parameter estimation, appear in §5.2.

5.2 Bayesian network

A Bayesian network (BN) [63] is a graphical structure that encodes probabilistic relationships among a set of random variables via a directed acyclic graph (DAG). The structure of a DAG is defined by two sets: (1) the set of vertices (nodes) V which represent random variables, (2) the set of directed edges (arcs) E which connect pairs of nodes, representing direct dependence among the variables. Figure 4.8 illustrates un example of a DAG.

A BN enables an effective representation and computation of the joint probability distribution over a set of random variables. It reflects a simple conditional independence statement based on the Markov property. Namely that the conditional probability distribution at a node given all non-descendants depends only on parents. That is, there are no direct dependencies in the system being modeled which are not already explicitly shown via arcs.

More formally, given a directed acyclic graph $G = (V, E)$ and $X = (X_v)_{v \in V}$, a set

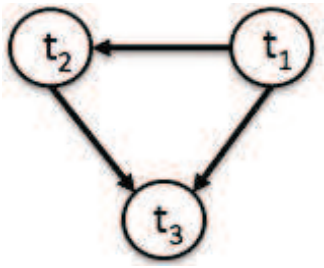


Figure 4.8: Example of a directed acyclic graph (DAG).

of random variables indexed by V. X is a BN with respect to G if its joint probability distribution can be written as a product of the individual distributions, conditional on their parent variables:

$$p(x) = \prod_{v \in V} p(x_v | x_{p(v)}),$$

where $p(v)$ is the set of parents of v (i.e., those vertices pointing directly to v via a single edge). For example, a BN associated to the DAG of Figure 4.8 would define a unique joint probability distribution over its nodes $\{t_1, t_2, t_3\}$:

$$p(x_1, x_2, x_3) = p(x_3|x_1, x_2)p(x_2|x_1)p(x_1),$$

where $\{X_1, X_2, X_3\}$ is the set of random variables respectively indexed by $\{t_1, t_2, t_3\}$.

In our setting, we are modeling X as a BN given each species $Y = c$. The underlying DAG (directed acyclic graph) is the binary tree $\mathcal{T}$, with arrows from parents to children. This structure enables modeling a relatively simple BN in which each node has a single parent (except the root) and two children (except terminal nodes). For example, in the simple BN of Figure 4.9, t and $p(t)$ are respectively the parent and the grand parent node of both $c_1(t)$ and $c_2(t)$. Let $\{X_t, X_{p(t)}, X_{c_1(t)}, X_{c_2(t)}\}$ be the random variables respectively indexed by $\{t, p(t), c_1(t), c_2(t)\}$. X_t should then *separate* $X_{p(t)}$ and $X_{c_1(t)}$ (resp. $X_{c_2(t)}$), meaning that the grand parent and the child should be conditionally independent given the parent X_t.

Here, using a Gaussian Bayesian network is motivated by several observations. First, we have already considered the individual SVM scores X_t as Gaussians: see §3. Second, whereas there are significant (conditional) correlations among many pairs of variables

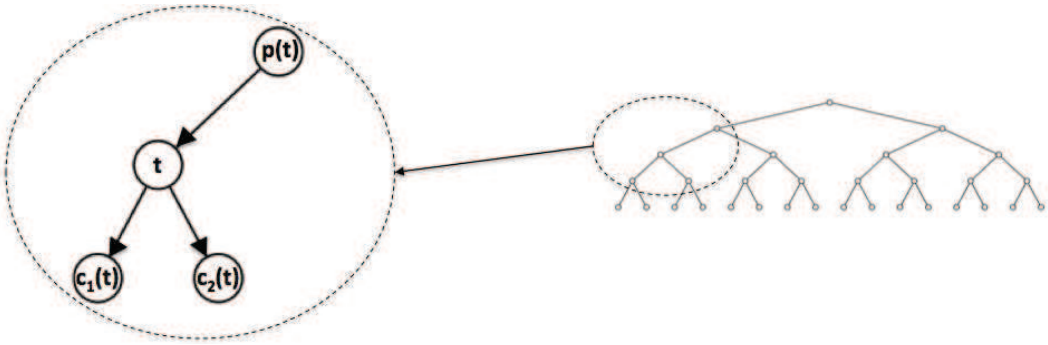


Figure 4.9: Illustration of the Markov property on a tree structure. The tree encodes independence assumptions, by which each variable is independent of its non-descendants given its parent in the tree. There are no direct dependencies which are not already explicitly shown via arcs: (i) the grand parent $X_{p(t)}$ and the child $X_{c_1(t)}$ (resp. $X_{c_2(t)}$) are conditionally independent given the parent X_t. (ii) The children $X_{c_1(t)}$ and $X_{c_2(t)}$ are conditionally independent given X_t. $\{X_t, X_{p(t)}, X_{c_1(t)}, X_{c_2(t)}\}$ are random variables corresponding to nodes $\{t, p(t), c_1(t), c_2(t)\}$.

X_t, X_s given the species Y, clearly we must control the complexity of the joint distribution since we do not have sufficient data to reliably estimate all the order $|\mathcal{T}|^2$ parameters involved in a full multivariate Gaussian parameterization. The motivation for the Bayesian network is that the largest of the (absolute) correlations tend to be between parents and children; see Figure 4.10.

Also, while BN assumes that there is conditional independence between grand parent and children, as well as between children, given the parent (given a species), unconditionally, the (absolute) correlation could be significant. We have first compared $corr(X_{p(t)}, X_{c(t)}|Y)$, a measure of independence between grand parent $p(t)$ and child $c(t)$ (given a species Y) and $corr(X_{p(t)}, X_{c(t)}|X_t, Y)$, a measure of independence between the grand parent and the child given the parent (given a species); an example is shown in Figure 4.11. Second, we have compared $corr(X_{c_1(t)}, X_{c_2(t)}|Y)$, a measure of independence between children (given a species) and $corr(X_{c_1(t)}, X_{c_2(t)}|X_t, Y)$, a measure of independence between children given their parent (given a species). As shown in both Figures 4.11 and Figure 4.12, the difference between correlations is generally large, i.e., $corr(X_{p(t)}, X_{c(t)}|Y) \gg corr(X_{p(t)}, X_{c(t)}|X_t, Y)$ and $corr(X_{c_1(t)}, X_{c_2(t)}|Y) \gg corr(X_{c_1(t)}, X_{c_2(t)}|X_t, Y)$ which further motivated the use of BN.

It should be emphasized, here, that all the measures of independence, given above,

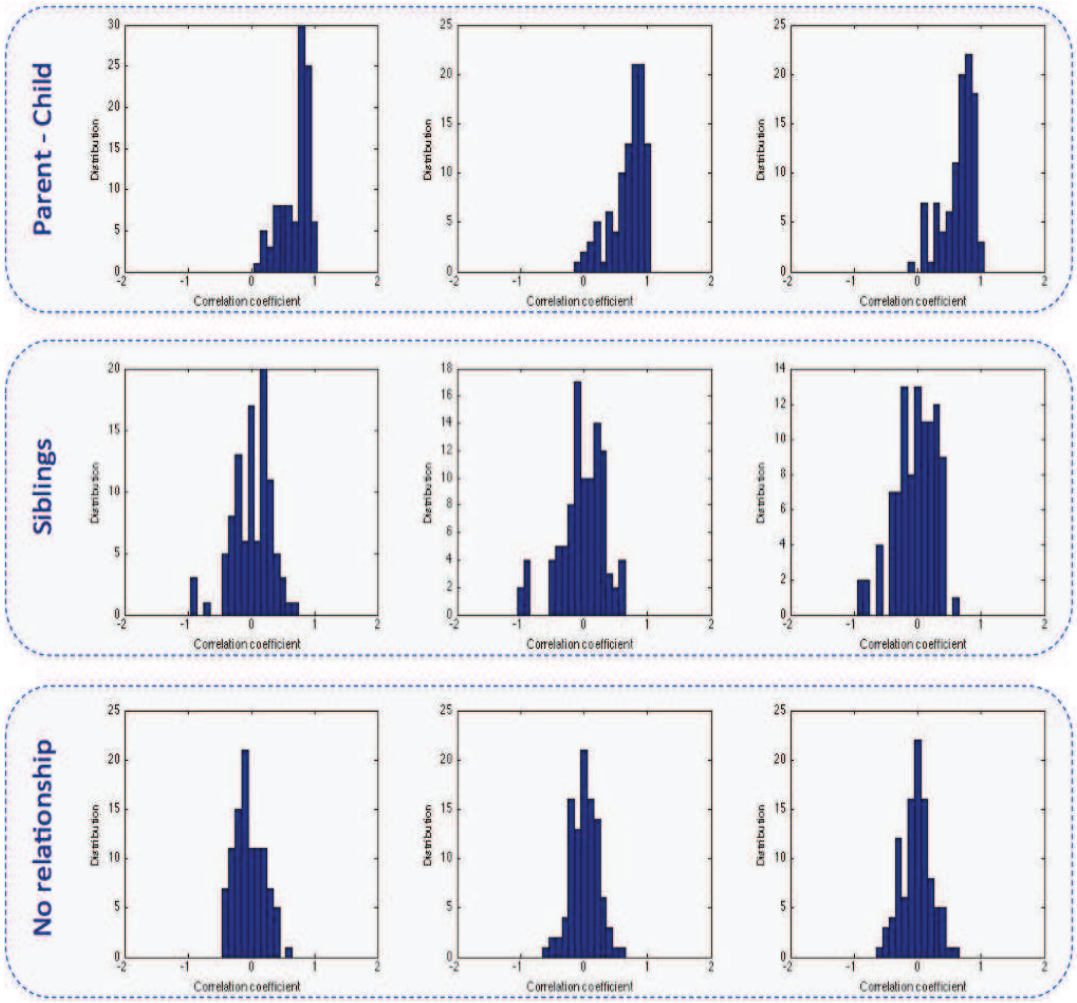


Figure 4.10: Histograms of correlation coefficients between differently situated pairs of nodes in the hierarchy for three different species. Top row: correlation coefficients between a node and its parent. Middle row: between siblings (nodes of a same parent node). Bottom row: between nodes from different halves of the hierarchy (no common ancestor except the root). Note that the largest of the (absolute) correlations tend to be between parents and children.

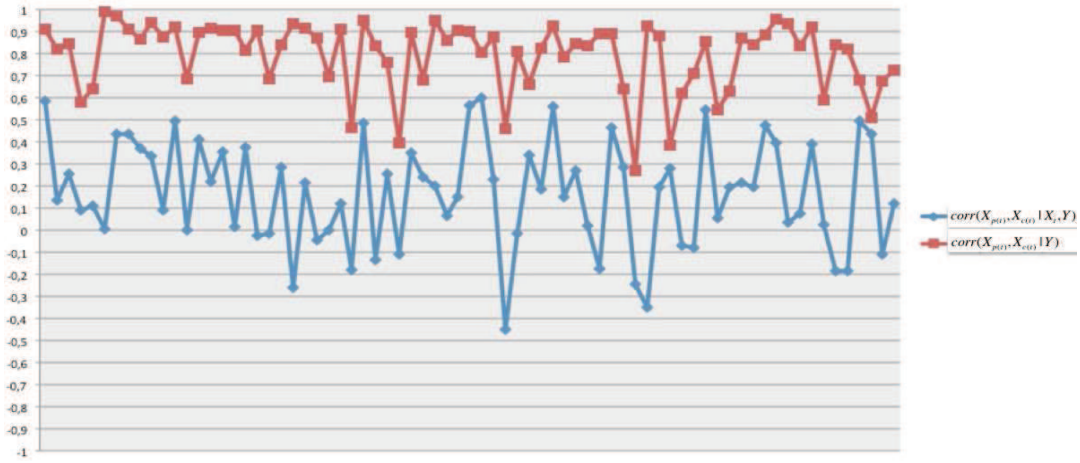


Figure 4.11: Example of correlation coefficients between a grand parent $X_{p(t)}$ and a child $X_{c(t)}$ for different species. In red are conditional correlations given the parent X_t and in blue are the (absolute) correlations, unconditionally to X_t. Note that the difference between both kinds of correlations is generally very large.

have been given by correlation since we assume normal random variables. In general, random variables may be uncorrelated but highly dependent. Generally, the correlation coefficient detects only linear dependencies between two variables (X, Y). However, in the special case of a multivariate normal distribution (i.e, the pair (X, Y) has multivariate normal distribution), any two or more of its components that are uncorrelated are independent.

With this Gaussian Bayesian network we must estimate three parameter (mean, variance, correlation with parent) for each non-root node and two parameters (mean and variance) for nodes t_1 and t_2. Consequently, the densities $f(x_1|c)$ and $f(x_2|c)$ in Equation (4.4) are univariate normal. The densities $f_t(x_t|x_{t-}, c)$ are obtained by recalling that if U_1, U_2 are jointly normal with means and standard deviations $\mu_1, \mu_2, \sigma_1, \sigma_2$ and correlation coefficient ρ, then $f(u_1|u_2)$ is normal with mean $\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2 - \mu_2)$ and variance $(1 - \rho^2)\sigma_1^2$. Hence

$$\begin{aligned} f_t(x_t|x_{t\text{-}}, c) &= \frac{1}{\sigma_t^c\sqrt{2\pi(1-\rho_t^c)}} \\ &\times \exp\left\{-\frac{(x_t - \mu_t^c - \rho_t^c\frac{\sigma_t^c}{\sigma_{t\text{-}}^c}(x_{t\text{-}} - \mu_{t\text{-}}^c))^2}{2(1-\rho_t^{c2})\sigma_t^{c2}}\right\} \end{aligned}$$

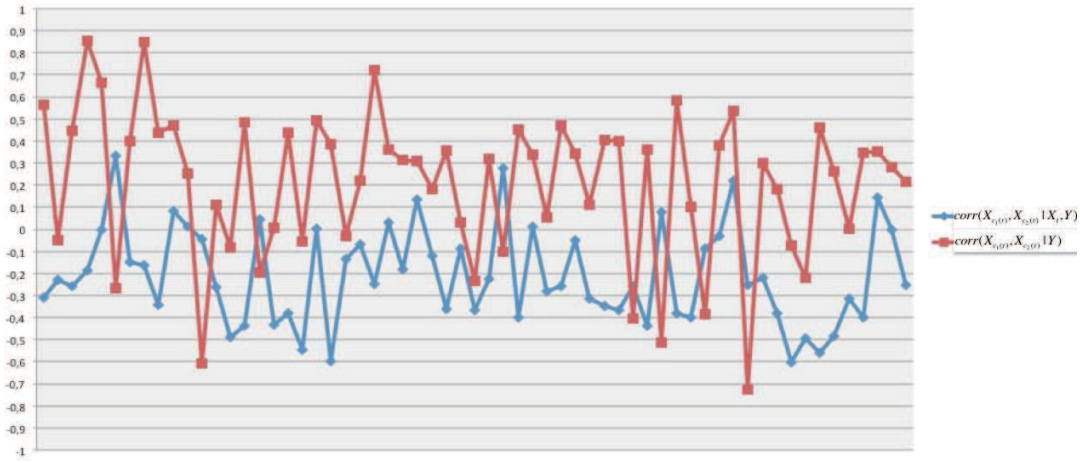


Figure 4.12: Example of correlation coefficients between two children $X_{c_1(t)}$ and $X_{c_2(t)}$ for different species. In red are conditional correlations given the parent X_t and in blue are the (absolute) correlations, unconditionally to X_t. Note that the difference between both kinds of correlations is generally very large.

where the superscripts indicate the dependence on the species c. Computing $p(\mathbf{x}|c)$ and thus $P(Y \in C_t|\mathbf{X} = \mathbf{x})$ is then straightforward.

5.3 Constructing the confidence set

The first step in CS selection is to compute the posterior probabilities $P(Y \in C_t|\mathbf{X} = \mathbf{x})$ for each $t \in \mathcal{T}$. This is straightforward given our model:

$$P(Y \in C_t|\mathbf{X} = \mathbf{x}) = \sum_{c \in C_t} P(Y = c|\mathbf{X} = \mathbf{x}) \tag{4.5}$$

$$= \frac{\sum_{c \in C_t} p(\mathbf{x}|c)}{\sum_{c \in \mathcal{Y}} p(\mathbf{x}|c)} \tag{4.6}$$

Now define

$$B(\mathbf{x}) = \{t \in \mathcal{T} : P(Y \in C_t|\mathbf{X} = \mathbf{x}) \geq 1 - \epsilon\}.$$

Obviously we can assume $\epsilon < 0.5$; in practice, we take values such as 0.05 and 0.01. It is then easy to see that for every $\mathbf{x}$, the set of nodes $B(\mathbf{x})$ is a non-empty path in $\mathcal{T}$ originating at one of the two roots t_1, t_2 and generally terminating before a terminal node is reached. The natural definition of $\widehat{C}$ is then the *smallest* set C_t in the tree which

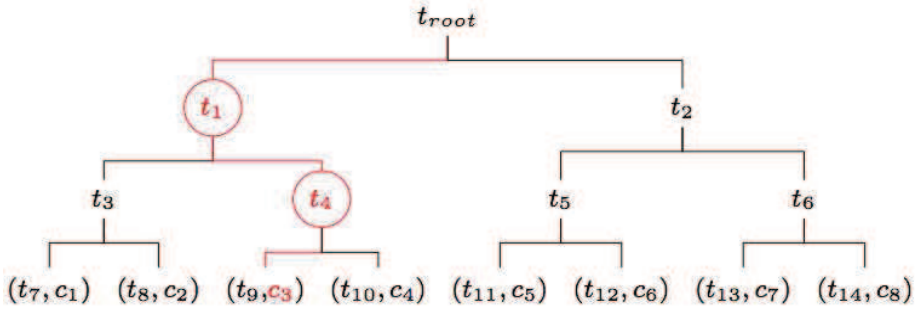


Figure 4.13: Example of $\mathcal{T}$ illustrating the key objects for 8 categories $(c_1, ..., c_8)$. $\mathcal{T}$ contains 14 nodes (not counting the root t_{root}), labeled $(t_1, ..., t_{14})$. Associated with each node t is a set of categories C_t, e.g., $C_{t_1} = \{c_1, c_2, c_3, c_4\}$. Here, the true category is $Y = c_3$, $B(\mathbf{x}) = \{t_1, t_4\}$, (red circles) which are on the true path (in red). So, $T(\mathbf{x}) = t_4$ and $\widehat{C}(\mathbf{x}) = \{c_3, c_4\}$.

satisfies the constraint. Specifically,

$$\widehat{C}(\mathbf{x}) \doteq C_{T(\mathbf{x})},\ T(\mathbf{x}) = \arg\min_{t \in B(\mathbf{x})} |C_t|.$$

Equivalently, $T(\mathbf{x})$ is the deepest node in $B(\mathbf{x})$. The corresponding *confidence level* for the given data is then

$$p(\mathbf{x}) = P(Y \in \widehat{C}(\mathbf{x}) | \mathbf{X} = \mathbf{x})$$

and the average confidence level is

$$Ep(\mathbf{X}) = P(Y \in \widehat{C}(\mathbf{X})).$$

Given the definition of $B(\mathbf{x})$, it follows that $Ep(\mathbf{X}) \geq 1 - \epsilon$.

Figure 4.13 illustrates the concepts above for a simplified hierarchical structure $\mathcal{T}$ of 8 categories $(c_1, ..., c_8)$. Here $T(\mathbf{x}) = t_4$ is the deepest node in $B(\mathbf{x}) = \{t_1, t_4\}$, and the resulting confidence set is the $\widehat{C} = \{c_3, c_4\}$.

The efficiency of this algorithm will be demonstrated in a variety of experiments in Chapter 6, both in terms of comparing with other methods as well as generating high confidence sets.

5.4 Relationship to non-Bayesian confidence sets

In classical (frequentist) statistics, there is no r.v. Y, only a family of probability distributions $\{p(\mathbf{x}|c)\}$ indexed by a parameter $c \in \mathcal{Y}$. A $100(1-\epsilon)\%$ confidence set for the true parameter c^0 is a random set (i.e., data-dependent) which contains c^0 with probability $1-\epsilon$. For a continuous real-valued parameter, an interval is often centered at a point estimate $\hat{c}$ such as the maximum likelihood estimator $\hat{c}_{ML} = \arg\sup_c p(\mathbf{x}|c)$.

Following this recipe we would begin with the maximum likelihood estimator $\hat{c}_{ML}$, which coincides with the MAP estimator $\arg\max_c P(Y = c|\mathbf{X} = \mathbf{x})$ in the Bayesian case when the prior is uniform. The tree provides a neighborhood structure: a natural way to "center" the CS at $\hat{c}_{ML}$ is to consider the subsets of categories along the path from $\hat{c}_{ML}$ to the root. However, given such a set $\widehat{C}(\mathbf{x})$ of categories containing $\hat{c}_{ML}(\mathbf{x})$, computing $P(c^0 \in \widehat{C}(\mathbf{X}))$ would require knowing the distribution of the ML estimator under c^0, which appears difficult. The Bayesian argument gives this in an *average* sense. (Note, however, that the CS constructed in the previous section does not necessarily contain the MAP estimator, but nearly always does in practice.)

6 Summary

In this chapter, we have presented two classification methods, both based on a hierarchical structure of the data. The first method adopts a CTF strategy, where classification proceeds systematically from coarse-grained to fine-grained characterizations, to finally output the set of "positive" leaf nodes. It uses a likelihood ratio framework based on local (SVM) scores. The second considers all hierarchical nodes (internal and leaf nodes) as potential candidates , i.e., *confidence set* candidates, and aims to output the one whose expected size is minimized subject to containing the true estimate with high probability. To this end, an analogy with confidence intervals in classical statistics is considered and a model-based strategy is used.

Chapter 5

Identification scenarios

1 Introduction

Different levels of interactive identification can be considered depending on the difficulty of the leaf data. The particular scenarios we consider were previously introduced in Figure 1.13. Here, section 2 describes an automated system while sections 3 and 4 introduce semi-automated alternatives. All of these scenarios consider a single leaf image. In section 5, we propose an identification process based on multiple images.

2 Baseline scenario

The baseline scenario is the standard one with no human intervention: given an image of a leaf, usually scanned against a flat background, the system automatically provides a single estimate of the true species. Given the different ingredients introduced previously, we propose to use either the IdKey representation or the vantage feature frames within a coarse-to-fine search. In such a framework, a tested image may arrive at no leaves, one or more than one leaf in the tree. In the last case, the species returned by the system is the one which gives the highest likelihood ratio L_t. The whole scenario is illustrated in the top row of Figure 5.1. There is no human intervention neither at the beginning nor at the end of the task. As shown in Figure 1.13, we apply this fully-automated scenario only on scanned leaves, in order to be able to ensure a fair automatic IdKey or vantage point detection. However, even with scanned leaves, the utility of this approach

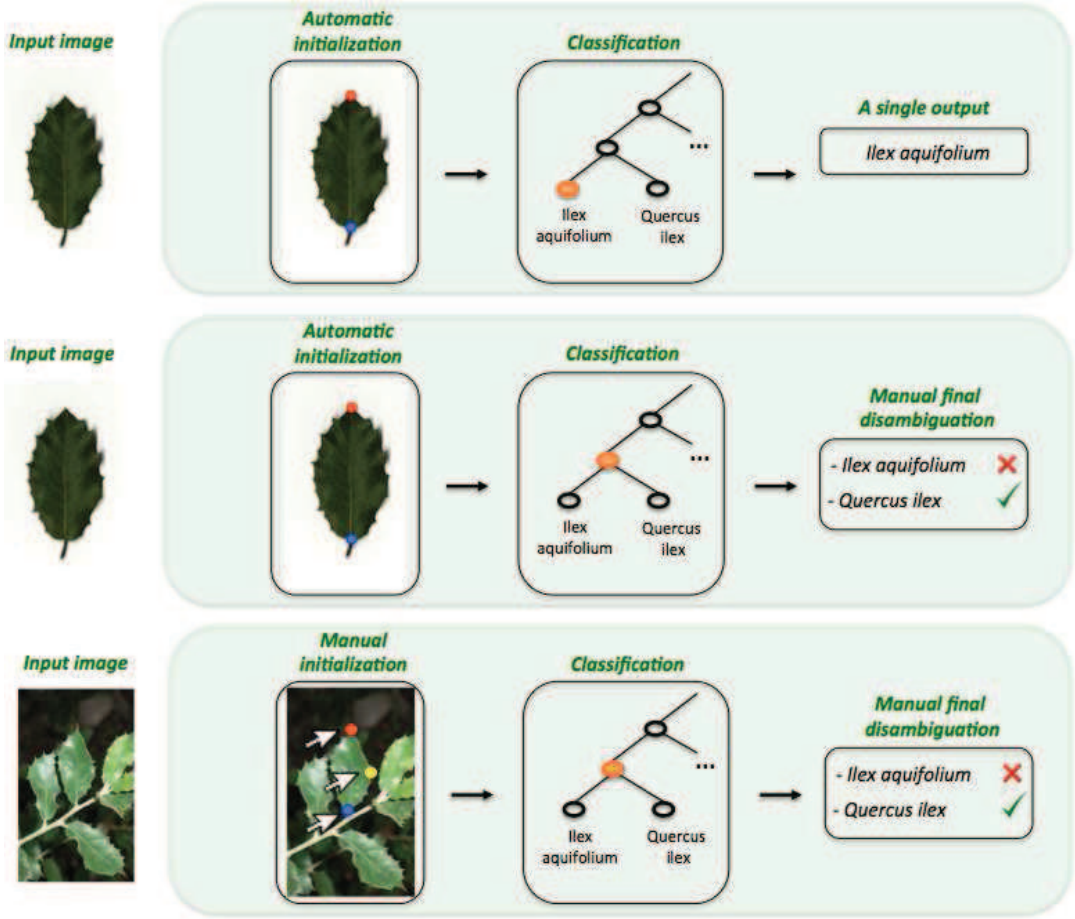


Figure 5.1: Given an input leaf image, three scenarios are proposed. A fully-automated identification, i.e., baseline case (top row) and a semi-automated identification with human intervention only at the end of the process, i.e., final disambiguation (middle row) or at both the beginning and the end for initialization and final disambiguation (bottom row).

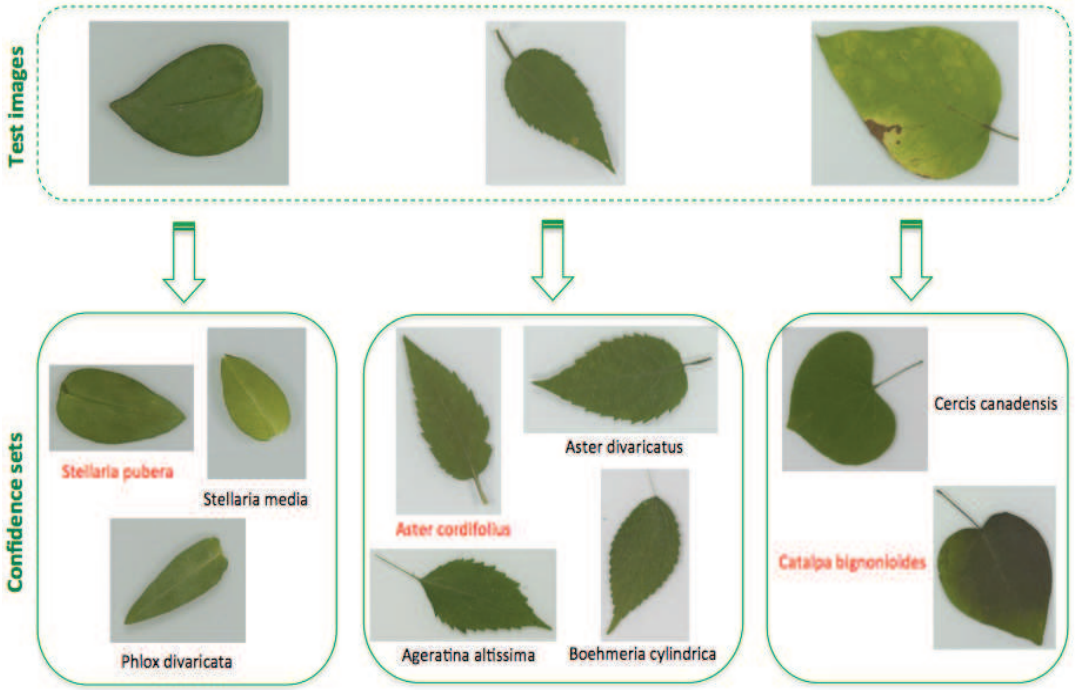


Figure 5.2: A sample of test scanned leaves with non-singleton confidence sets (CS) of species while a hierarchical clustering was used to get the CS candidates. For each CS, a training image from each species is displayed. For each test image, the red species is the true one. The CS are visually coherent.

is questionable due to relatively high error rates with large databases which contain very similar species and display high variability within the same species. This motivates a design of semi-automated systems.

3 Final disambiguation

Given a leaf image, vantage feature frames are first automatically detected and category-dependent features are extracted as described in §4. Then, given the scores $\mathbf{x}$ for the leaf image being processed, the Bayesian network model and fixing ϵ, we compute the sub-path $B(\mathbf{x})$ of $\mathcal{T}$ and finally provide $\widehat{C}(\mathbf{x})$ to the user. The type of user intervention will then depend on the needs and skills of the user. The novice user may simply accept $\widehat{C}$ as it stands or use reference material to narrow it down. A more skilled user may be able to identify it if it resides in the set or recognize that it does not. Of course, the smaller

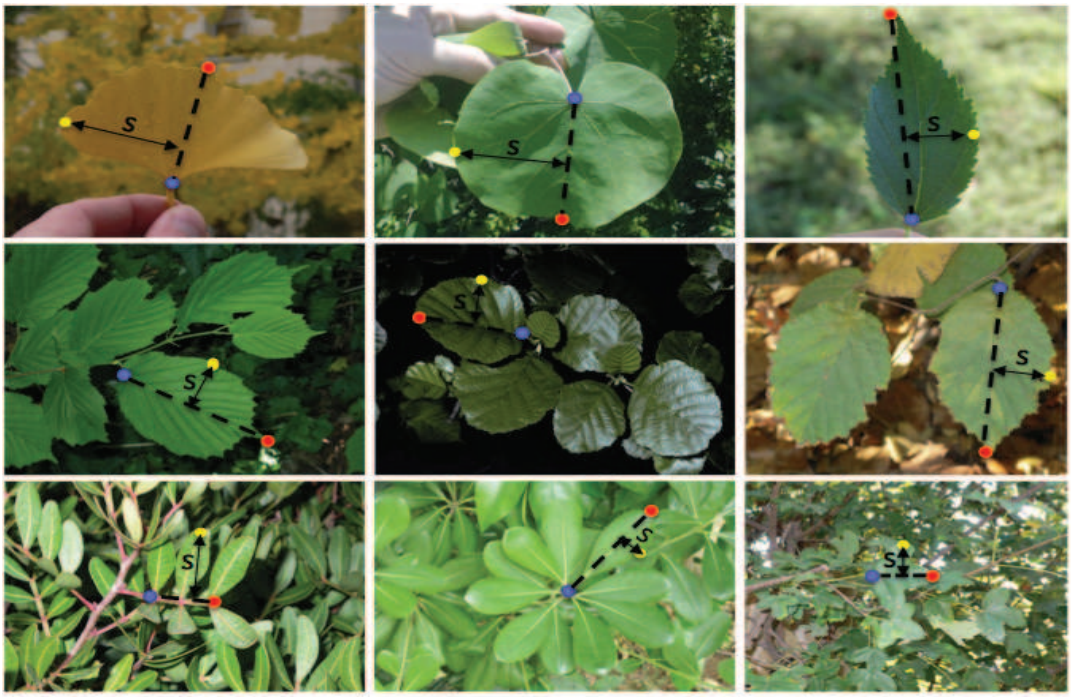


Figure 5.3: Examples of leaf photographs manually marked. For each image, displayed are the leaf base (blue point), the leaf apex (red point) and a third boundary point (yellow point). The approximate width (marked as s at each image) of the leaf is defined as the distance between the yellow point and the apex-base line.

the confidence set, the more informative and useful it is. When a clustering principle is used to build the hierarchy of CS candidates, we guarantee a visually coherent set to the user; examples are shown in Figure 5.2. Several experiments in Chapter 6 will demonstrate the efficiency of such a scenario on leaf images with uniform background (e.g. scanned leaves).

Using a corse-to-fine classification within a likelihood framework, we can also provide a set of estimates with the top k highest likelihood ratio L_t. However, in this case, we can not ensure with a high probability that the true species is among the estimates. Of course, the higher is k, the higher is the probability but the less useful is the system.

4 Initialization

For leaf images with a cluttered background, automatic detection of vantage points requires a very efficient segmentation algorithm (robust to background noise and texture), which is not the case for the algorithm we use (Otsu) or any we are aware of. In particular, it could be exceedingly difficult to automatically (and accurately) extract a single leaf boundary from a branch or foliage image; see Figures 5.3 (middle and bottom rows). Moreover, returning a $P\%CS$ is of little value in applications if either $|CS|$ is very large or $P \ll 100$. The minimal intervention we can imagine is asking the user to mark several landmarks; providing a faithful segmentation is another possibility but we are able to obtain good results without this level of intervention.

We ask the user to mark the two terminals of the main vein of the leaf, the base and the apex, as well as a third boundary point which will be used to approximate the width of the leaf (see Figure 5.3). The centroid of the leaf is defined as the mid-point of the apex-base line. Local features are then extracted in two coordinate systems, one centered on the base and the other on the apex as described in Chapter 3. The same classification process as in the previous scenario §3 is used to provide the user with a CS. A summary of this scenario is provided in Figure 5.1 (bottom row). Some examples of the CS returned are shown in Figure 5.4.

5 Multiple images

In a botanical field scenario where the basic unit of observation is a plant, botanists can examine different samples of leaves from the same plant in order to determine the species. In fact, one sample alone might not capture sufficient information for accurate identification.

Using multiple-image queries rather than a single leaf image can then improve the identification accuracy by taking advantage of the added information and the complementaries of different leaf appearances

First, we apply one of the identification scenarios described above to ensure both quick and effective species estimation for each image query $\mathcal{Q}_j$. The set of estimates is then ranked (depending on the strategy chosen, we can use likelihood ratios or posterior

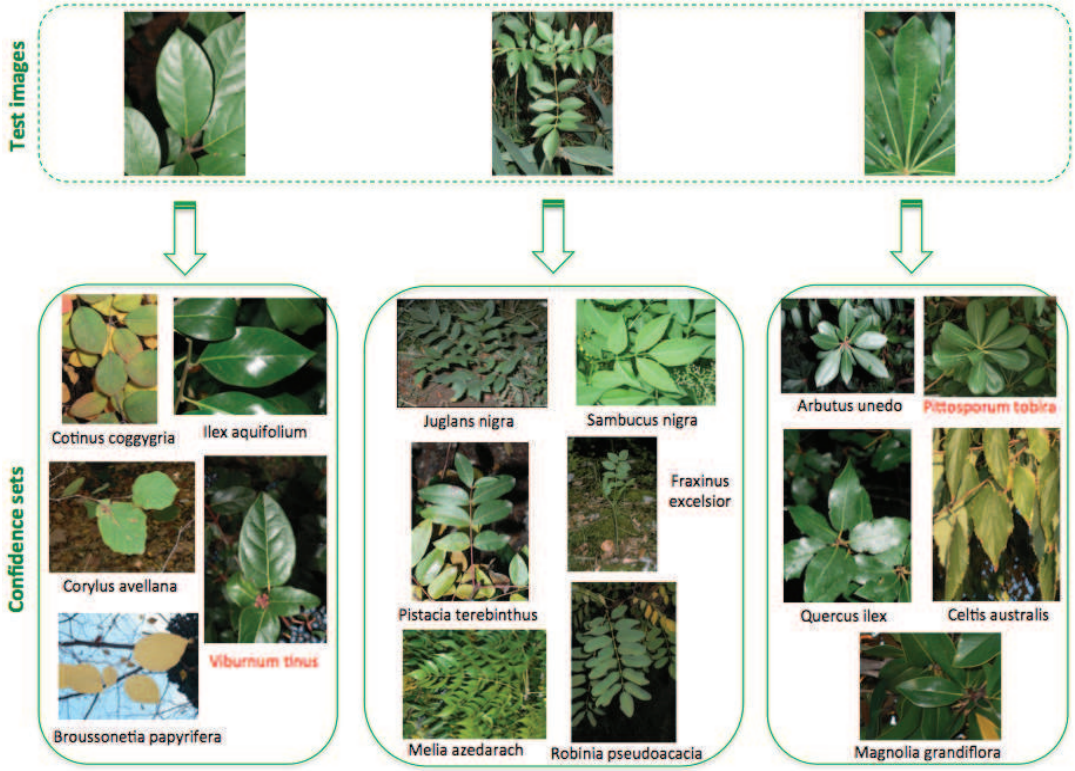


Figure 5.4: A sample of test leaf photos (cluttered background) with non-singleton confidence sets (CS) of species. For each CS, a training image from each species is displayed. For each test image, the red species is the true one.

masses for ranking). Then, the individual ranked estimates are collated into a single set of estimated species for $\mathcal{Q}$. More specifically, they are combined by scoring each species by the number of its occurrences at the first r ranks.

Let $\widehat{Y}_r$ denote the combined set of estimates at rank r for all the images of $\mathcal{Q}$. The r^{th} species returned for $\mathcal{Q}$, denoted $\widehat{y}_r$, is then defined recursively as the most frequent species in the union of $\{\widehat{Y}_1, ..., \widehat{Y}_r\}$ other than $\widehat{y}_1, ..., \widehat{y}_{r-1}$. An illustrative example using a hierarchical estimation of IdKeys is shown in Figure 5.5. This aggregation of results improves the identification accuracy, as confirmed on several leaf datasets as described in Chapter 6. It is also independent of the process used to estimate the species of each query image. The same framework could then be used on other applications such as using multiple organ queries.

For educational and decision-support purposes, one can also imagine to illustrate each estimated species by a varied set of representative images in order to provide useful information about the unknown plant (represented by $\mathcal{Q}$) and its most closely related species. In our setting, we illustrate each returned species with both the most similar and the most different training leaf on average to the query. These additional images illustrate both the intra-class variability and inter-class similarity (e.g., see the two first estimated species for the first query, i.e., plant ID = 212, in Figure 5.6) and consequently could be useful in studying similar species, for instance helping botanists discover new relations or distinctions between or within different taxonomic groups.

Let $\mathcal{M}_\mathcal{Q}$ denote the feature vector obtained by averaging the individual feature vectors over all $\mathcal{Q}_j$, so $\mathcal{M}_\mathcal{Q}$ represents the compound query $\mathcal{Q}$. For each rank r, let Z_{I_T} denote the feature vector of the training image I_T from the same species as $\widehat{y}_r$. Let

$$d_{\mathcal{Q},I_T} = ||\mathcal{M}_\mathcal{Q} - Z_{I_T}||$$

be the difference in $L2$ norm between $\mathcal{Q}$ and I_T. Then, the most similar and the most different training images to $\mathcal{Q}$ consist of those which respectively minimize and maximize $d_{\mathcal{Q},I_T}$. Some examples are illustrated in Figure 5.6.

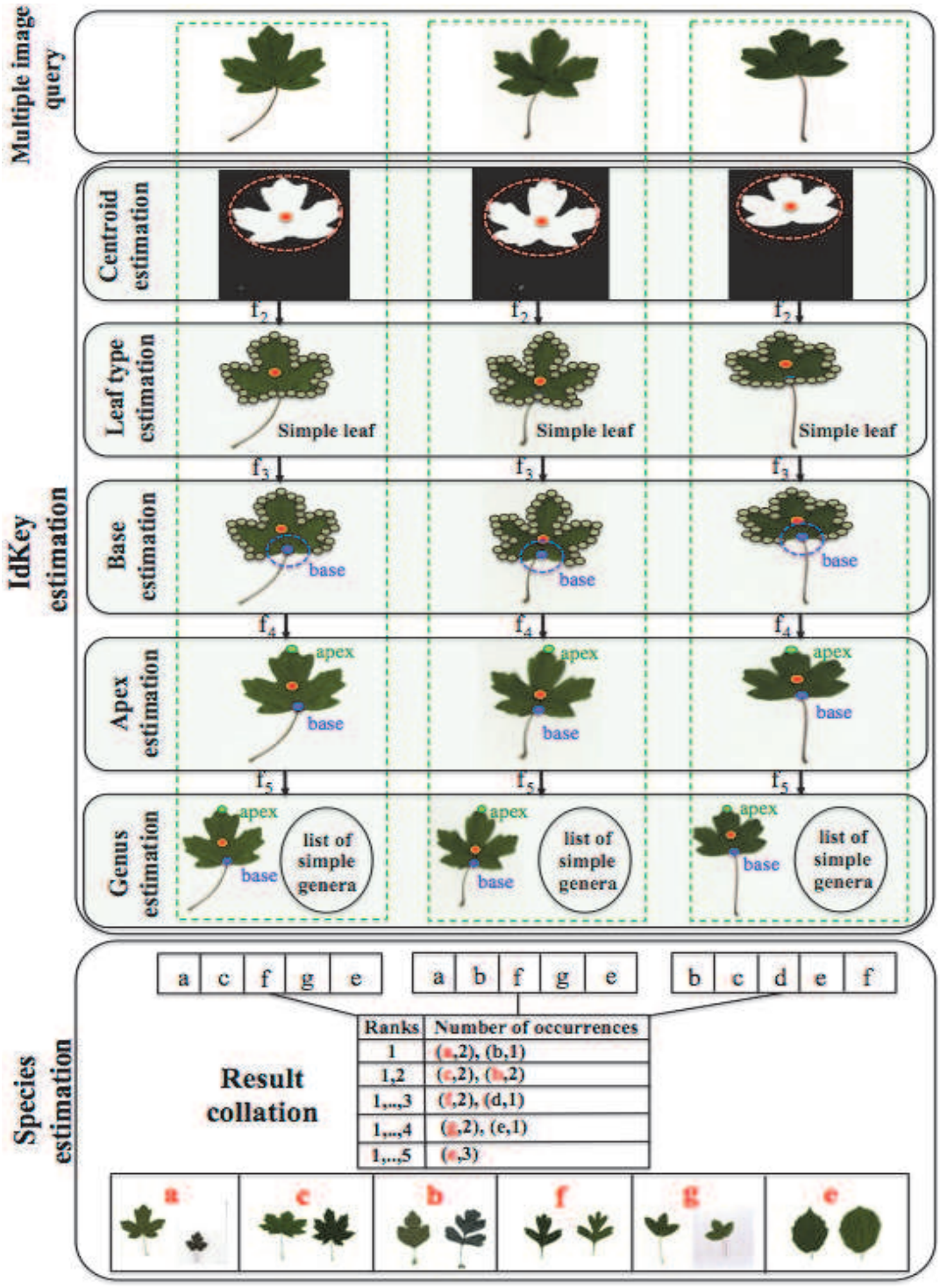


Figure 5.5: Species identification using botanical IdKeys and multiple image queries.

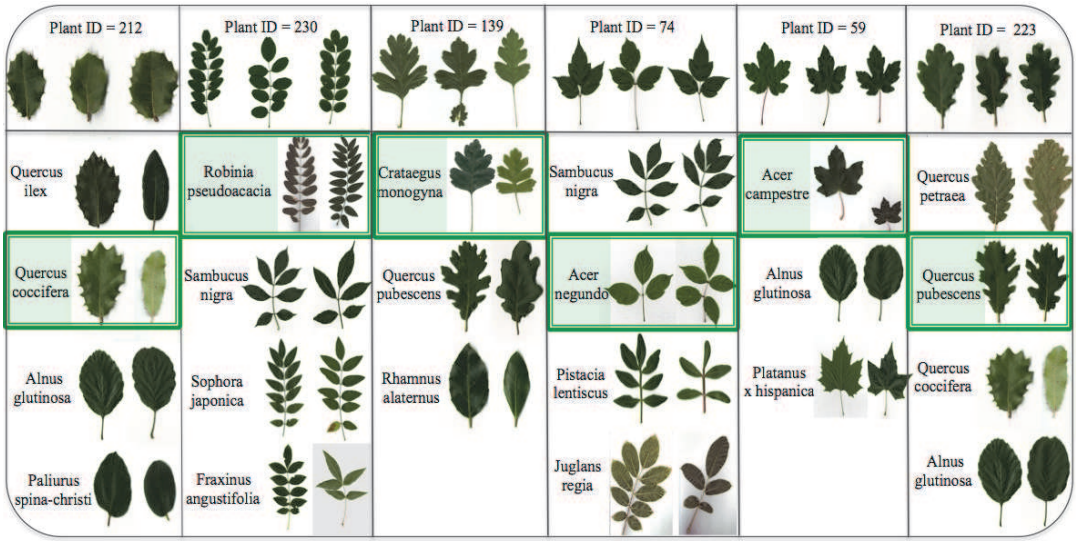


Figure 5.6: Examples of plant identification with multiple scanned leaf images, using a hierarchical estimation of IdKeys within a likelihood framework. Each column shows one example. For each example the top row shows the query plant which is represented by three images, while rows 2-5 show the top four species returned (when they exist). Each species returned is displayed with both the most similar (on the left) and the most different (on the right) training image on average to the query and which belong to that species. Correct species are framed with a green box.

6 Summary

This chapter presented different identification scenarios (i.e., both fully-automatic and semi-automatic scenarios) using the different descriptions and classification processes introduced in previous chapters. In each scenario, a complete identification process is detailed. Educational and decision-support applications could then be considered. Either a single or multiple leaf samples could be used. Here, an example of a natural result aggregation is presented. Other state-of-the art methods could also be adopted.

Chapter 6

Experiments

1 Datasets

We considered five challenging leaf datasets from different geographical areas. Four of them consist of images of single leaves on a white background (Swedish [104], Flavia [119], ImageClef 2011 [50] and Smithsonian [7] datasets). The last one consists of unconstrained photographs of leaves (ImageClef 2011 photo category [50]). We also apply our approaches on a dataset of orchid flowers[1].

1.1 Swedish leaves

This has 1125 scanned leaf images containing 75 images from each of 15 different Swedish plant species; see Figure 6.1. This dataset was the first publicly available leaf data, introduced by the authors of [104] for research. Although it contains relatively few varieties of species (e.g., a single palmate leaf and only two compound species), we chose it in order to be able to compare our work with various approaches, including generic shape classification approaches [117, 45, 74] which were applied on leaves. Following all previous work on this dataset, we randomly select 25 training images from each species and test on the remaining images in order to evaluate our performance.

[1] Courtesy of Roland Martin and Errol Vela
[2] http://www.imageclef.org/2011/Plants
[3] http://www.tela-botanica.org/site:botanique

Figure 6.1: Samples from the Swedish dataset. One image from each species is shown.

Figure 6.2: Samples from the Flavia dataset. One image from each species is shown.

1.2 Flavia leaves

The Flavia dataset is composed of 1907 scans of leaves. It consists of 32 species with 50-60 observations in each species. As we can see in Figure 6.2, this dataset contains only simple leaves. It was introduced in [119] and was used to evaluate some leaf classification algorithms [111, 35, 54]. Following [119], we used 10 leaves from each of 32 species to evaluate our performance, so that a total of 320 leaves are used for testing the algorithms and the remaining leaves for training.

1.3 Smithsonian leaves

This leaf database has 6717 scanned leaf images containing 202 different species (148 simple species and 54 compound ones) from the Northeastern U.S area. The number of exemplars per species varies from 2 to 63. These images were provided by the Smithsonian botanical institution within the framework of the US National Herbarium.

Figure 6.3: Samples from the Smithsonian dataset. One image from each species is shown.

One particularity of these data is that the images present various poses and orientations of leaves as well as different structures of basal and apical parts as shown in Figure 6.3. Thus, good performance on such a dataset suggests robust and effective landmark estimation. We use two-thirds of the images for training and one-third for testing.

1.4 ImageClef leaves

Used in the ImageCLEF2011 plant identification task[2], this dataset contains three categories of images: scans of leaves acquired using a flat-bed scanner, scan-like leaf images acquired using a digital camera and free natural photos. The complete leaf collection consists of 71 species from the French Mediterranean and was constructed through a citizen science initiative conducted by Telabotanica[3], a French social network of amateur and expert botanists (more details can be found in [50]). As a result, the task it represents is quite close to the conditions encountered in a real-world application. Here, we focus on both scans and photos; see Figure 6.4 and 6.5.

1. *Scanned leaf subset*: This category has 3070 scanned leaves. We consider the same training and test sets as in the ImageCLEF2011 benchmark, i.e., 2349 images for training and 721 test images. In particular, the training leaves were collected from 151 plants and those of the test set from other 55 plants.

2. *Natural photo subset*: This category has 1469 unconstrained photographs of leaves. Again, the same training and test sets as in the benchmark are considered, i.e.,

Figure 6.4: Samples from the ImageClef2011 leaf scans. One image from each species is shown.

930 images for training and 539 test images. Each image can represent either a single leaf, a branch or a foliage as shown in Figure 1.12 (see the first Chapter). In particular, the training leaves were collected from 269 plants and those of the test set from 99 other plants. Not all the species were considered for testing. Only samples from 26 species were available for testing using 40 training species.

1.5 Orchid flowers

There are 1610 images representing 23 species of a relatively rare orchid flower family provided by the "Mediterranean Orchid Society" (*Société Méditerranéenne d'Orchidologie*).[1]

2 Experiments and analyses

In this section, we evaluate the proposed approaches on the different datasets. Also, we compare our results with previous work on plant identification.

2.1 IdKey estimation

As previously described, the IdKey estimation is hierarchical. First, we estimate the leaf type, i.e., simple or compound. Depending on this first key value, we compute either

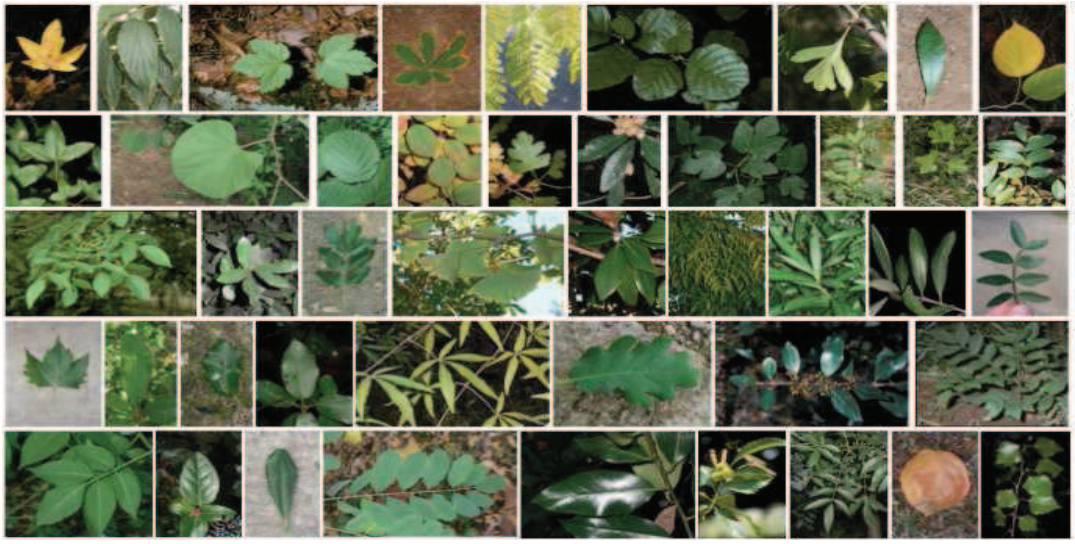


Figure 6.5: Samples from the ImageClef2011 leaf photos. One image from each species is shown.

Table 6.1: Accuracy of IdKey estimation on several leaf datasets.

IdKeys	Leaf type (simple or compound)	Botanical points
Smithsonian leaves	98.4%	90.4%
Swedish leaves	99.2%	95%
ImagClef2011 leaves	95.8%	92.4%

three or four landmarks. Three landmarks are considered for simple leaves (centroid, base and apex) and four landmarks for compound leaves (centroid, base, terminal apex and second apex).

We achieve over 95% accuracy for leaf type estimation on different leaf datasets and over 90% for botanical landmark detection as shown in Table 6.1, and thereby confirm efficient discrimination between simple and compound leaves as well as reasonable invariance to shape and structure. The euclidian distance was used to evaluate the landmark detection $D_l = d(\widehat{l} - l)$ where $\widehat{l}$ is the estimated landmark location, l is the true landmark and d is the euclidian distance. The entire detection process is considered erroneous if any point is not accurately detected, i.e., $D_l \geq 10$ pixels for any landmark l.

In each histogram in Figure 6.7, each bin i represents the percentage of tested images,

Figure 6.6: Samples from the Orchids dataset. One image from each species is shown. Note that the color is not a discriminative feature; many differently colored orchids could belong to the same genus or species.

from the Swedish dataset, in which $D_l \in [x_i, x_{i+1}[$, where $\{x_i\}$ goes from 0 pixels to 100 pixels by steps of 10 pixels. It should be noted that 10 pixels represents 2.5% of the average width of the tested images. In fact, the tested images have different sizes: the average width is about 400 pixels and the average height is about 380 pixels. We notice that for both the terminal apex and the base, the distance D_l between the estimated point and the actual point is fewer than 10 pixels about 95% of the time. The few cases in which $D_l \geq 100$ pixels (last bin), often correspond to a confusion between the two points, i.e., the actual base is identified as the terminal apex and the actual terminal apex is identified as the base.

2.2 Vantage point detection

We evaluate vantage point detection for both leaves and orchid flowers. We recall that vantage points refer to the apex and the base points for leaves and to the tip of the central sepal and the bottom of the labellum for orchid flowers. We achieve over 90% accuracy for vantage point detection either for leaves or Orchid flowers as shown in Table 6.2. Figure 6.8 shows vantage point detection results for orchids and different type of leaves (e.g., toothed, lobed, concave, convex, symmetric, asymmetric). We used the same metric described above to evaluate the detection accuracy.

An accurate point detection is important for a correct identification. Over 80% of

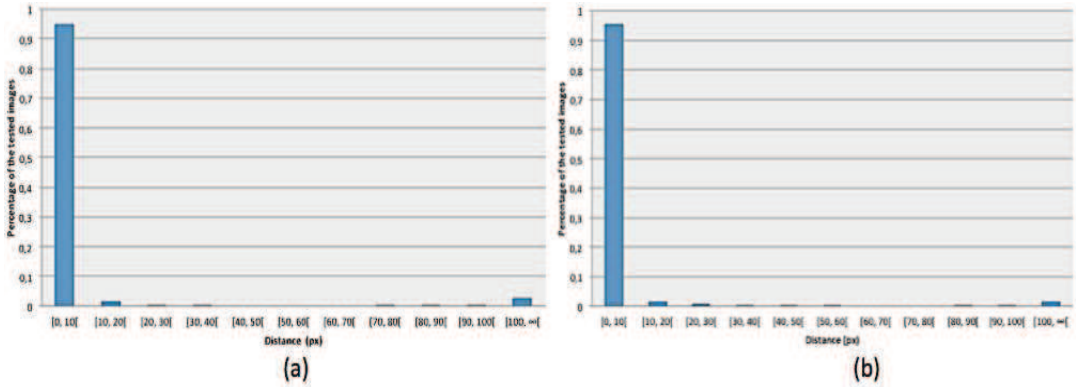


Figure 6.7: Botanical landmark estimation on the Swedish dataset. (a) Histogram of the distances between the estimated terminal apex and the actual one. (b) Histogram of the distances between the estimated base and the actual one.

Dataset	**Detection Rate**
Smithsonian leaves	92%
Swedish leaves	96%
ImageCLEF leaves	93%
Orchids	95%

Table 6.2: Accuracy of vantage point detection on several leaf datasets.

leaf images in which point detection was considered erroneous were incorrectly identified, especially those in which the system mixes up the base and the apex; examples include those of Figure 6.9.

2.3 Coarse-To-Fine (CTF) classification

Here, we evaluate the CTF strategy within the likelihood framework. We use both the IdKey representation and the Vantage Feature Frames (VFF).

We provide the rate on the holdout test data at which the true species coincides with our top estimate ("top-1") and appears among our top five estimates ("top-5") using both single and multiple-image queries.

For both the Smithsonian and Swedish subsets, a multiple-image query represents random images from the same species since there is no information about the plants used in these datasets. However, for ImageCLEF2011, we re-organized the testing data

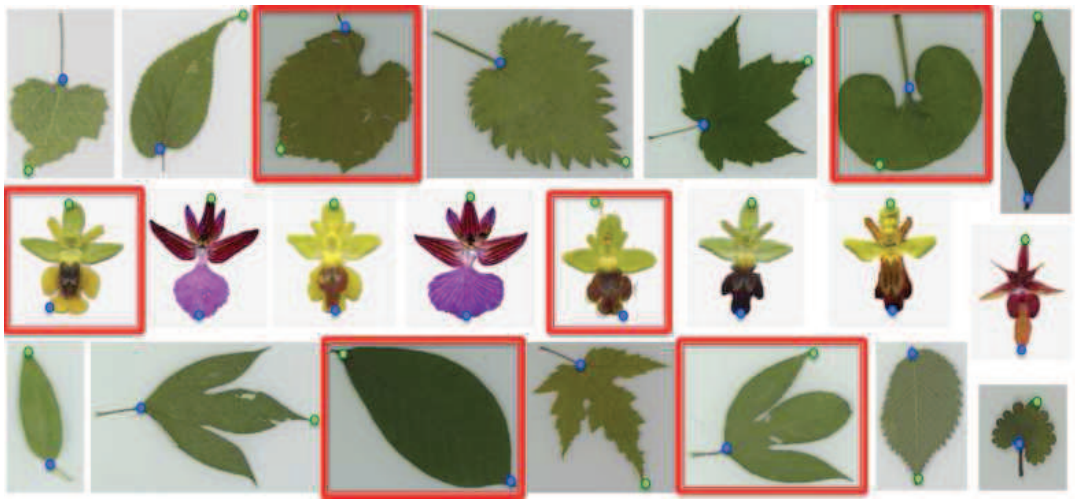


Figure 6.8: Random sample of test images with the estimated vantage points for both Smithsonian leaves and orchids. False detections are framed with a red box. Note that the entire detection process is considered erroneous if any vantage point is not accurately detected.

to extract the different testing plants and to be able to evaluate the efficiency of our approach in the conditions of a real-world application. Note that there is no plant used in both the testing and training sets. Rather than using all the test images, we consider only random images of different plants, i.e., we compute the recognition rate for the 55 plants of this dataset. Each query represents a single or multiple image(s) of the same plant.

Swedish Data: We first compare our results with the state-of-the-art results on this dataset, including those of the IDSC [74], the Shape-Tree [45] and sPACT [116] methods which all use single image queries. We achieve the best performances using the IdKeys approach (also using single image queries) with over 98% accuracy for the top-1 estimate using either IdKey or VFF representation; see Table 6.3. However, a "flat classification" using one-vs-all SVM's and VFF representation (F-SVM in Table 6.3) yields only 93.3% accuracy while only considering the species with the highest SVM score. Using a hierarchical model is clearly of value for this dataset.

Smithsonian Data: Table 6.4 reports the recognition rates for both the top-1 and the top-5 responses with CTF classification using the IdKey representation as well as different numbers of images per query. The correct species appears in the top-5 over

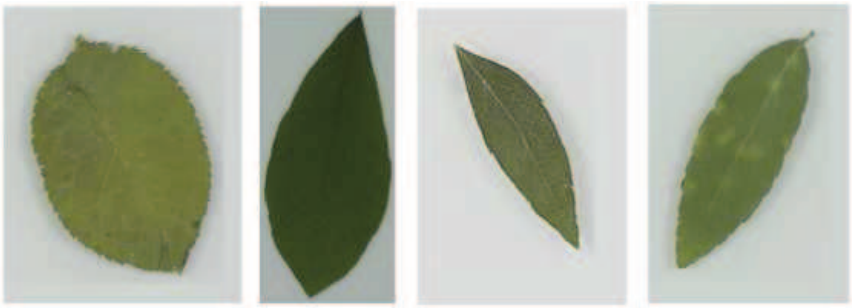


Figure 6.9: Example of leaves for which the system mixes up the base and the apex. Note that those leaves have either a very small or no petiole and reveal a very similar shape in their both extremities.

Table 6.3: Results of different methods on the Swedish data.

Methods	Perf. (top-1)
IdKeys + 5 images/Query	**100%**
IdKeys	**98.4%**
VFF	**98.4%**
sPACT [116]	97.92%
TSLA [82]	96.53%
Shape-Tree [45]	96.28%
SPTC+DP [74]	95.33%
IDSC+DP [74]	94.13%
F-SVM	93.3%
SC+DP [74]	88.12%
Söderkvist [104]	82.40%

95% of the time considering multiple-image queries. In particular, we achieve 78.5% accuracy for the top-ranked species using a single image per query and over 90% using 10 images.

Table 6.5 reports the recognition rates for both the top-1 and the top-5 responses while considering a CTF classification using the Vantage Feature Frames. The VFF representation improves the performance of the CTF search by achieving 92% of accuracy (instead of 88.4% using IdKeys) while considering the top-5 estimates for a single test image. VFF are not only more generic but could also be more efficient than the hierarchical representation of IdKeys which was dedicated only to leaves.

Orchid Data: To the best of our knowledge, there is no previous work on this family of

Table 6.4: Performance of our CTF classification using IdKeys on the Smithsonian data.

Nb images/Query	1	3	5	10
Perf. (top-1)	78.5%	85.3%	89.1%	**90.6%**
Perf. (top-5)	88.4%	95.3%	**96.7%**	95.1%

Table 6.5: Performance of our CTF classification using VFF on the Smithsonian data

Nb images/Query	1	3	5	10
Perf. (top-1)	79.6%	86.2%	89.7%	**90.2%**
Perf. (top-5)	92%	94.6%	**95.2%**	94.8%

flowers. We applied the Vantage Feature Frame approach within a CTF search on this data to demonstrate how it could be readily applied to a different type of closely-related botanical species. We achieve 81% accuracy for the top-ranked species ($n = 1$) and 97% for $n = 5$ as shown Table 6.6.

ImageCLEF2011 Data: Finally, we apply our approach in a real-world context using a set of leaf images from the same plant for each query thanks to the additional information provided with this dataset. Table 6.7 shows our performance when considering different numbers of images per query for both the top-1 and the top-5 responses. In particular, using multiple-image queries improves the identification performance, reaching over 95% accuracy for the top-5 estimates. The results on a random sample of test plants is shown in Figure 5.6. Of particular note is first the high similarity between leaves of some species and the variation in appearance of leaves within others; for example, note the visual difference between the leaf images of the first plant (plant ID=212) and those from the training set. This is essentially due to the use of different plants and conditions to collect training and test data.

n	**1**	**2**	**3**	**4**	**5**
Orchid data	81%	92%	94%	96%	97%

Table 6.6: Recognition rates using Vantage Feature Frames on Orchid flowers.

Table 6.7: Performance of CTF classification using IdKeys on the ImageClef2011 data

Nb images/Query	1	3	5	10
Perf. (top-1)	61.8%	65.5%	74.5%	**78.2%**
Perf. (top-5)	83.6%	**98.2%**	96.4%	96.4%

2.4 Confidence sets

We evaluate the confidence set based classification. For ease of notation, we label three scenarios:

- **CS0:** The confidence set CS is generated by ranking the posterior probabilities and accumulating species until the total mass exceeds $1 - \epsilon$; see §5.
- **CS1:** The process is automatically initialized and the $\widehat{C}$ used is described in §3. Note that the size of $\widehat{C}$ is necessarily at least as large as the CS returned by **CS0**.
- **CS2:** The process is manually initialized and the $\widehat{C}$ used is described in §4.

We will also refer to the *baseline cases* where the confidence set is restricted to a singleton by considering only the species in CS with the highest posterior mass.

To evaluate the performance of the proposed framework, we first provide the rate on the holdout test data at which the true species appears among the list of estimates, and second analyze the size of the response.

In order to be able to compare our performance with that of other methods, we will also adopt other evaluation metrics: (1) the accuracy rate among the top k estimates for the Swedish, Flavia, and Smithsonian datasets, (2) the evaluation metric[1] used for the ImageCLEF2011 plant identification task, for the ImageCLEF photo subset, which allows us to compare our performance with that of all the task participants. Such a metric refers to a *normalized classification rate* evaluated on the first species returned for each test image while taking into account the individual plant and the author (more details about the metric definition and the participants can be found in [50]). In all the following experiments , we use $\epsilon = 0.01$.

Before focusing on the identification results, we recall that the CS classification approach leverages certain sets of species from a predefined hierarchy based on a hierarchical clustering as described in §2. Two hierarchies are illustrated in Figures 6.10 and

Table 6.8: Comparison between CS0 and CS1 on the Swedish dataset.

Scenario	Accuracy	Average size of the response
CS0	99.2%	1.2
CS1	99.5%	1.3

6.11 for the Swedish and the Flavia datasets, respectively. Note the visual similarity of the grouped species which makes the confidence sets visually coherent.

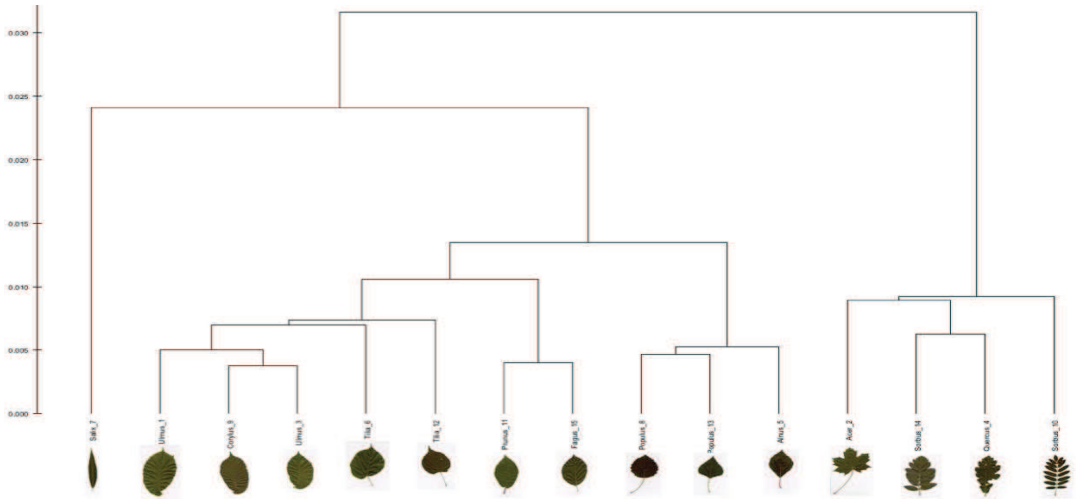


Figure 6.10: A dendrogram representing a hierarchical clustering of Swedish species. Displayed are the nested groupings of species, similarity levels at which groupings change, and a thumbnail from each species.

Swedish Data: We use this dataset to evaluate CS1. As shown in Table 6.8, the correct species belongs to the CS returned at 99.5% of the time while applying the CS1 scenario and at 99.2% while applying CS0. Figure 6.12 illustrates the distribution of the size of the CS while applying both scenarios. The average size is less than 1.5; see Table 6.8. Both of CS1 and CS0 do achieve near-perfect results while returning a single estimate at most of the time. By construction, both strategies are equivalent when only one estimate is returned. Note that one advantage of the proposed approach compared with CS0 is that CS1provides visually coherent sets for the user; see Figure 5.2. An additional advantage will be demonstrated on the Smithsonian and the ImageClef data.

In order to be able to compare CS1 with previous work using the same evaluation framework, we use the baseline strategy. That is, we provide only the species in CS

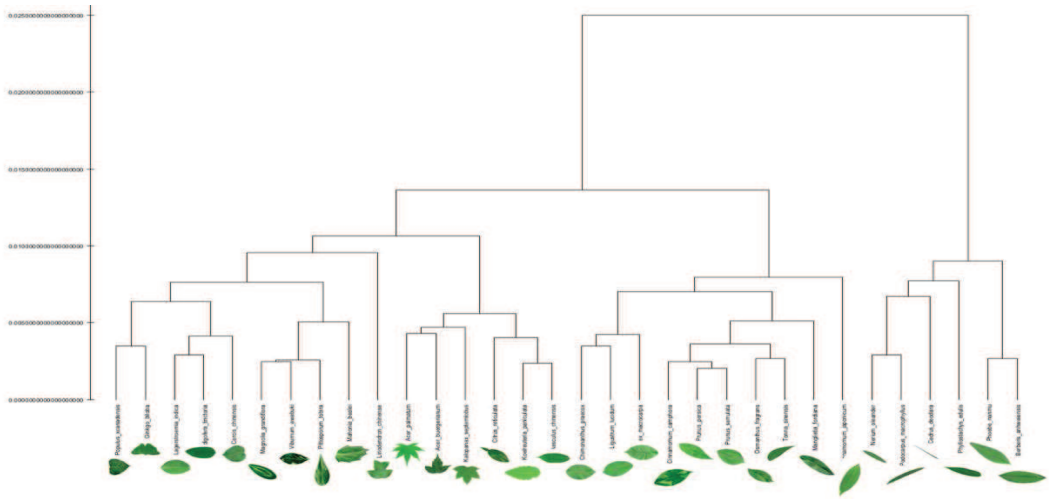


Figure 6.11: A dendrogram representing a hierarchical clustering of Flavia species.

Table 6.9: Comparison between CS0 and CS1 on the Flavia dataset.

Scenario	Accuracy	Average size of the response
CS0	97.1%	1.6
CS1	98.1%	2.2

with the highest posterior mass. We achieve better performance than CTF classification while considering the top-1 estimate (98.7% vs. 98.4% as described in Table 6.3).

Flavia Data: As with the Swedish leaves, we do achieve near-perfect results: using CS1, the average size of the response is 2.2 for an accuracy rate of 98.1% as shown in Table 6.9. Figure 6.13 illustrates the distribution of $\widehat{C}$ while applying both CS0 and CS1. We have $|\widehat{C}| = 1$ 87.5% of the time. Also, we notice from Table 6.9 that CS1 outperforms CS0 in accuracy. It does so at the expense of providing slightly more estimates on average.

Finally, we use the same evaluation framework as in [119] to enable a direct comparison with some previous methods. We consider only a single estimate. As shown in Table 6.10, we outperform other methods, including the SVM-based flat classifier (F-SVM), by returning the species in $\widehat{C}$ with the highest posterior mass.

Smithsonian Data: With CS1, we achieve 92% accuracy while returning about 5 estimates on average; the accuracy with CS0 is 90%. As shown in Figure 6.14, we do return a single estimate about 93.2% of the time. In order to be able to compare CS1

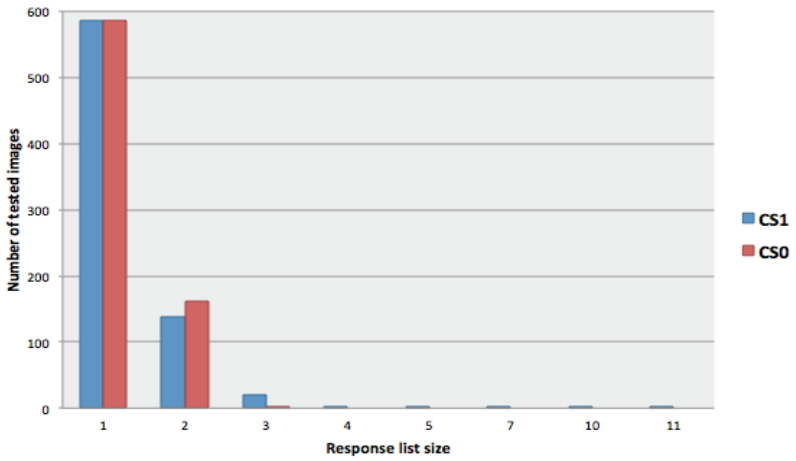


Figure 6.12: The distribution of $|\widehat{C}|$, the size of the CS returned, for both methods of constructing the CS when testing on the Swedish leaves.

Table 6.10: Different results on the Flavia data while considering a single estimate (top-1).

Methods	Accuracy
CS1 - *baseline case*	**97%**
F-SVM	94%
RBFNN [35]	94%
MLNN [35]	94%
1-NN [54]	93%
MMC [111]	92%
BPNN [111]	92%
RBPNN [54]	91%
PNN [119]	90%

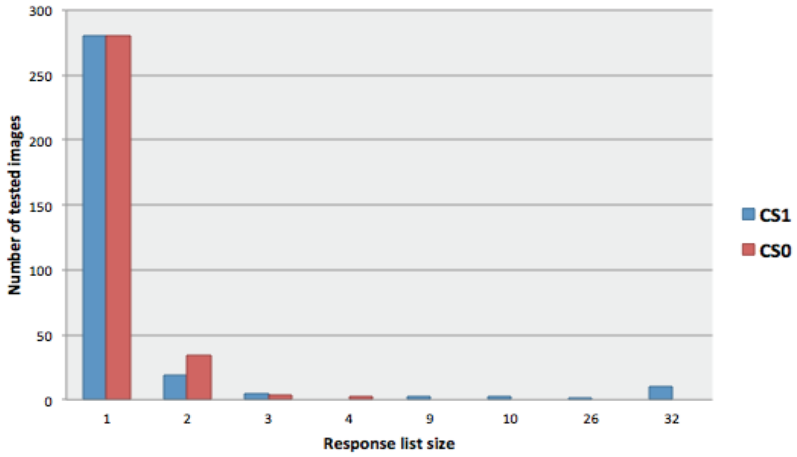


Figure 6.13: The distribution of $|\widehat{C}|$, the size of the CS returned, while testing on the Flavia leaves.

with CTF classification on such a dataset, we rank the list of species in $\widehat{C}$ for each test image, using their posterior masses. In this case, we also achieve about 90% accuracy for the top response compared with 79.6% for using CTF strategy within the VFF description where achieving about 90% accuracy required using the top five responses.

Whereas using both strategies, the CS is estimated by the model to capture the true species with very high probability, this of course does not necessarily occur in practice due to errors in estimating the true posterior distribution. To illustrate this, Figure 6.15 shows the distribution of the posterior masses of the true species on the Smithsonian leaves. Note the high value (at least 0.9) for the majority of the tested images; in this special case, CS0 is equivalent to CS1 as both achieve perfect results. However, CS1 is more efficient when the true species has low mass under the model; the CS1 strategy can recover from such a catastrophic error in estimation due to the way the CS is constructed as long as there are species with non-trivial posterior masses which are visually similar to the true one. In Figure 6.15, 10.4% of the images for which the posterior probability of the true species is less than 0.1 were missed by CS0 but not by CS1, but never vice-versa.

ImageCLEF Data: Finally, we apply our approach in a real-world context using unconstrained photographs. For this subset, we focus on CS2, using human input to mark some landmarks at the beginning of the process as explained in §4. First, we

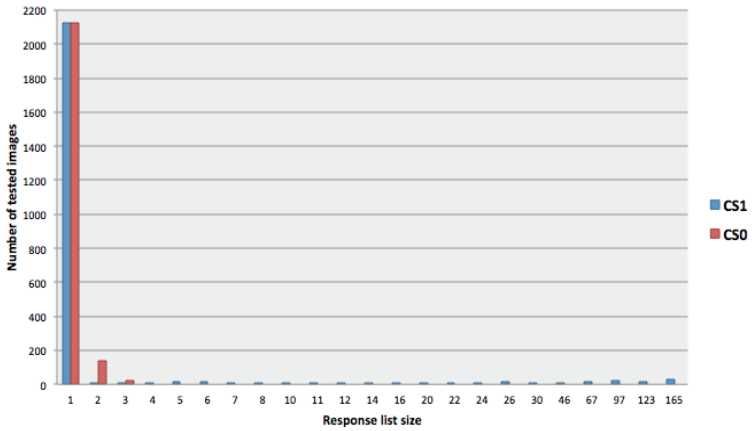


Figure 6.14: The distribution of $|\widehat{C}|$, the size of the CS returned, while testing on the Smithsonian leaves.

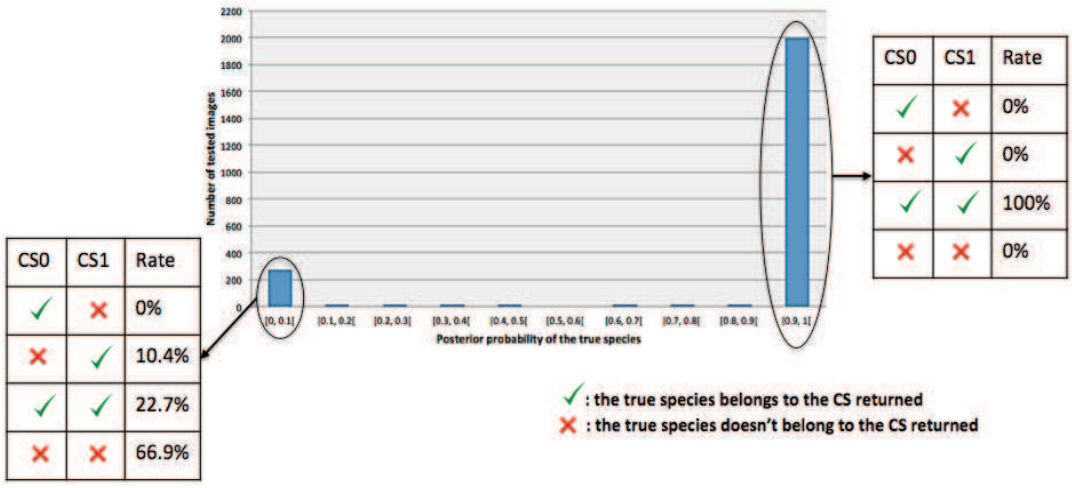


Figure 6.15: The histogram (in blue) of the posterior masses on the true species for the leaves in the Smithsonian dataset. The two tables compare the performance of CS1 and CS0 at the two extremes, i.e., when the posterior mass on the true species is very low and very high. In the former case (drastic estimation error), the CS1 strategy is able to recover (generate a CS with the true species) but CS0 does not for 10.4% of the images, but there are no images for which the opposite occurs, i.e., CS0 succeeds but CS1 does not.

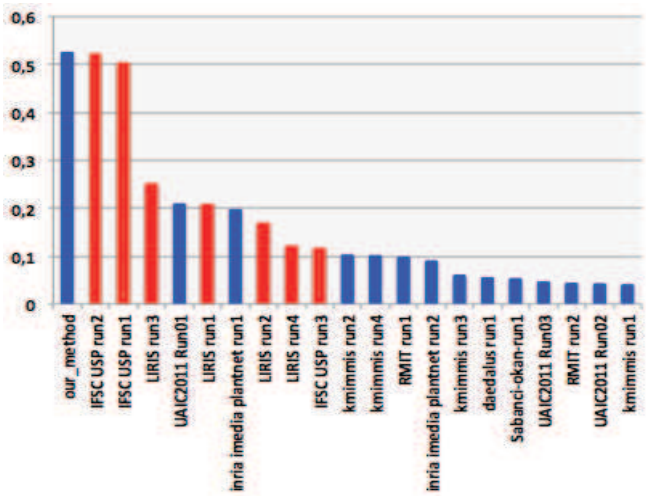


Figure 6.16: Classification scores on the leaf photos of the ImageCLEF2011 dataset. In red are the scores of the methods which use segmentation and in blue are the scores of those which do not use segmentation.

compare our method with the entries to the ImageCLEF2011 plant identification task on the photo category. (Again, we rank the list of species in $\widehat{C}$ using their posterior masses.) In this task, each entry was assigned a normalized classification score s^1 as explained in §1. Figure 6.16 shows the scores of all the submitted runs of the eight participants; details about the participants can be found in [50]. We achieve the best score: $s = 0.525$. More specifically, two groups can be formed among the participants: the methods which use segmentation process (in red) and those which do not use segmentation (in blue). One can notice a relatively big gap between these two groups in terms of performance, i.e., there is a difference of about ± 0.3 between the best scores of the two groups; see Figure 6.16. We outperform all the previous work on such data, including segmentation-based methods. Note that the best score ($s = 0.523$ for "IFSC UPS run2") among the participants was obtained using a *manual* segmentation which is not feasible in real-word application.

Figure 6.17 illustrates the distribution of $|\widehat{C}]$ while applying CS2 to the ImageCLEF2011 photos. About 50% of the time we find $|\widehat{C}| \leq 10$. However, we only achieve 58.4% accuracy due to the difficulty of this task compared with identifying leaves on an uniform background; evidently, the posterior probabilities are poorly estimated. Figure

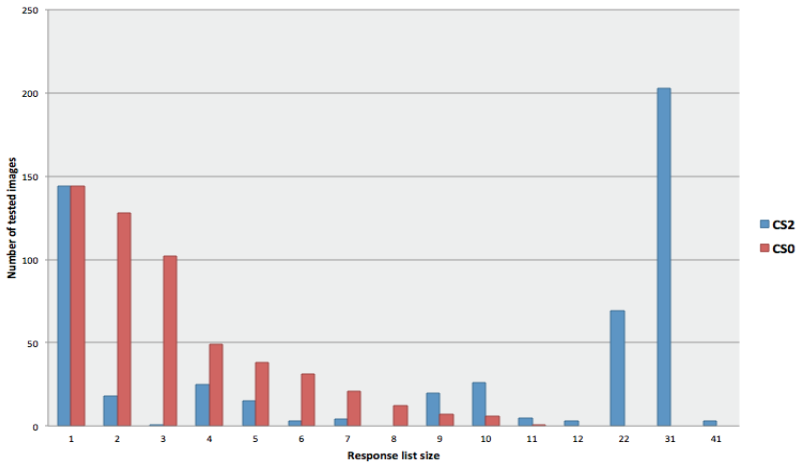


Figure 6.17: The distribution of $|\widehat{C}|$, the size of the CS returned, while testing on ImageCLEF2011 leaf photos. The blue histogram is CS2 and the red is CS0, both with manual landmark identification.

6.18 shows the distribution of the posterior masses of the true species on the ImageClef2011 photos. In contrast with Figure 6.15, note the low value (less than 0.1) of this mass for the majority of the tested images, which accounts for the even lower accuracy of CS0 strategy, namely 38.4%. The superior performance of CS2 occurs because for 32.5% of the images for which the posterior mass on the true species is less than 0.1, the CS generated by CS0 does not contain the true species but the one generated by CS2 does.

Moreover, additional issues are revealed from a more detailed analysis and which would explain the relatively low accuracy rate (comparing to other data). Figure 6.19 illustrates the different accuracies obtained per species. We completely fail to recognize those which have only few training samples (between zero and six); see the red boxes in Figure 6.19. Note that four tested species do not appear among the training species and these represent about 12% of the test images. Also, using different image types (i.e., leaf, branch and foliage photos) has made the task more challenging, especially since the number of samples per image type is not balanced. For example, one has only very few foliage images to predict a picked leaf image from the same species. However, we manage to recognize species from different image types with approximatively "equivalent" performance, especially for branch and foliage photos as shown in Table 6.11.

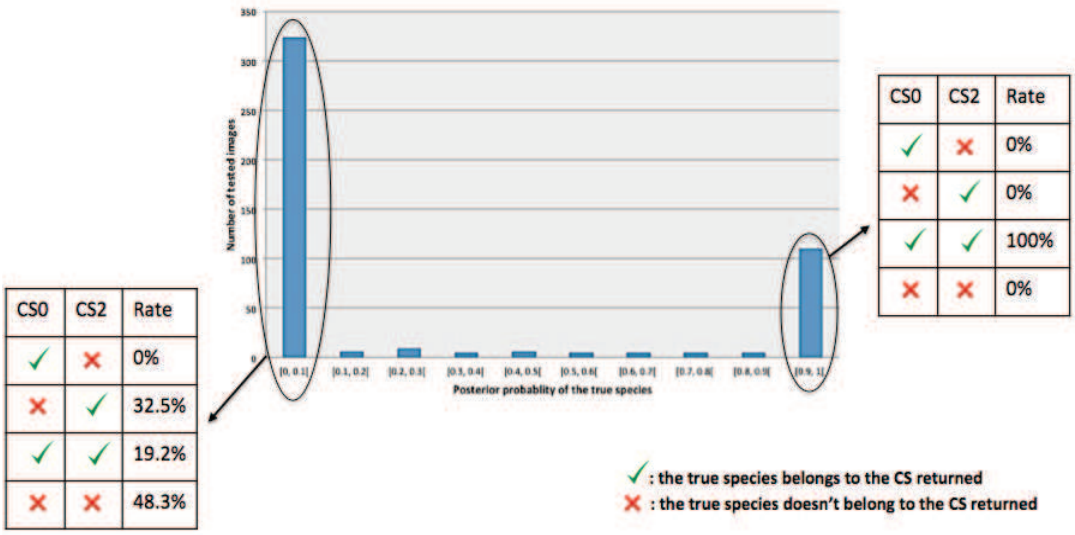


Figure 6.18: The histogram (in blue) of the posterior masses on the true species for the ImageClef2011 photos. The two tables compare the performance of CS2 and CS0 at the two extremes, i.e., when the posterior mass on the true species is very low and very high. Among the low ones, the the CS2 strategy succeeds and the CS0 strategy does not in 32.5% of the cases, but never the opposite.

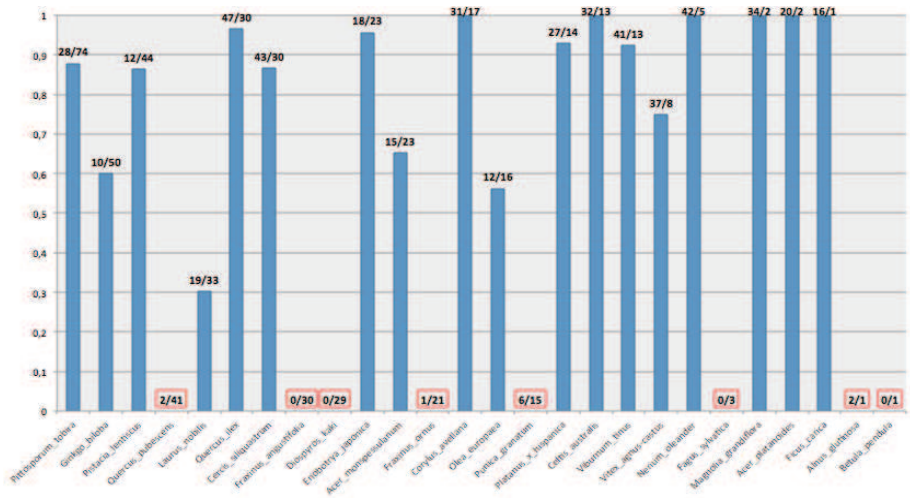


Figure 6.19: Illustration of the performance per species on the ImageCLEF photo subset. Each bin is labeled by two numbers separated by a slash. The first one refers to the number of training samples in the species considered and the second one refers to the number of testing samples.

Table 6.11: Performance of CS2 on different image types of ImageCLEF2011 photos.

Image type	Accuracy
Leaf	53.4%
Branch	60.3%
Foliage	61.5%

Figure 6.20: Random sample of incorrectly identified imageCLEF2011 leaf photos.

More generally, the quality of the photographs affects the performance. Of course, a well-photographed leaf would be easier to identify and also well-photographed training samples would lead to a better learning algorithm. For example, a close-up photo where the leaf covers a large part of the picture is sharp whereas the background is optically blurred due to a short deep-of-field would provide more useful visual content than a picture which is globally blurred or in which the leaf is out of focus, too damaged (e.g., dry leaves), too small or/and the background is predominant with a sharp visual content like grass or foliage of other plants, etc. Figure 6.20 illustrates some cases of failure.

Finally, experiments on the ImageCLEF photo subset demonstrate the efficiency of the proposed scenarios using multiple leaf images of an unknown plant. Note that we use here only the imageCLEF data since they are the only leaf images for which we know the plant identity thanks to the additional annotation provided with this dataset; see §1.

We improved to 74.5% accuracy (a gain of 16.1%) and the normalized ImageCLEF2011 score reaches $s = 0.626$ (a gain of about 0.01).

3 Summary

This chapter describes the performance of different identification scenarios using several leaf databases of differing difficulty. We have compared and analyzed different results. We have outperformed several state-of-the art methods and baselines, including a flat classification based on one-vs-all SVM's. The key point is that we have investigated both description and classification algorithms. We believe that both of these ingredients should be efficient to achieve near-perfect results, especially for fine-grained categorization. Each of the proposed algorithms could also be used in different applications. Extensions focus on experiments on larger datasets or other kind of images, in order to test, for example, the performance of the *confidence set* based classification.

Chapter 7

Conclusion

1 Summary of contributions

Stored images of biological objects are accumulating at a staggering rate due to new sensor technologies, their expanding use in web-based search engines and growing demands for web-based services in traditional sciences such as botany. These developments have been accompanied by an increasing demand for the automated analysis and more fine-grained discrimination of these data. In this dissertation, we investigated fine-grained issues, focusing on determining botanical species from leaf images, and considered the whole chain of an identification process, including object description and representation as well as classification algorithms and identification scenarios.

In Chapter 3, we introduced a novel object representation which is derived from domain knowledge in the form of IdKeys, and built a hierarchical representation of botanical keys which is dedicated to leaves. The keys are determined sequentially and prime the identification of the species. The main point is to refine the leaf description and thus narrow down the set of possible estimates. To obtain higher efficiency and a generic representation, we also investigated discriminating feature frames which are centered on botanical vantage points. The different characteristics of these frames, namely, the geometric and the appearance-based components, combine to provide the cues needed to distinguish between closely-related objects such as leaves or orchid flowers.

Using such descriptions, we began with a baseline identification scenarios, i.e., providing the user with a single estimate or a ranked list of the top-k estimates. For this, we have adopted a coarse-to-fine classification strategy based on a likelihood ratio framework using local (SVM) scores. The key point was to be able to exploit shared attributes and the hierarchical nature of the data (see §4).

In order to improve the accuracy and thus the utility of our identification system, we introduced in §5 the concept of confidence sets in analogy with confidence sets in classical statistics. The idea was to output a set of categories which contains the true one with high probability, rather than a point estimate (or a ranked list). Our approach was model-based and Bayesian. The expected size of the confidence set plays the role of the width of the confidence interval in standard statistics and the posterior probability that the true category belongs to the confidence set plays the role of the confidence level. We have also investigated the use of multiple leaf images to identify a plant in order to take advantage of the complementaries of different leaf appearances, there by further improves the recognition rates.

Both fully-automated and semi-automated identifications have been explored as described in Chapter 5. We have shown that different levels of interactive identification could be considered depending on the difficulty of the data. The user could participate either at the end or the beginning of the process in order to ensure an accurate and useful system, especially for the most difficult cases. For example, we proposed manual initialization of the process by asking the user to mark some well-defined landmarks in order even attempt to identify leaves against cluttered backgrounds (i.e., natural photos) without a burdensome and error-prone segmentation process. Also, providing the user with a visually coherent confidence set of categories makes the system usable by either an amateur or a botanist. While the amateur may simply accept the system output or use reference material to narrow it down, the botanist may be able to identify the correct species if it resides in the set or recognize that it does not. In both cases, such an automated system can dramatically speed up identification and classification.

2 Future work

Automatic landmark definition

Either for defining IdKeys or vantage points, we have taken advantage of the specific domain knowledge in botany. So far, landmark definition has been manual. The IdKey hierarchy has been manually crafted and the vantage point candidates have been provided by experts. Automatically determining candidate landmarks for constructing the vantage feature frames would make our object representation system more generic and practical for fine-grained categories other than botanical species, removing the need of expert intervention which could be expensive in some fields. Learning such landmarks of interest from scratch would evidently be a major challenge that we would like to attempt.

Cluttered photos

We believe that an important contribution of our work was to reveal the efficiency of our approach on natural leaf photos compared with the state-of-the-art methods. However, while applying our methods on photos, we have realized that, unlike scans, different types of images can be taken for a single organ. In the case of leaves, one can photograph a picked leaf, a branch or foliage, which can add further challenges to automated systems. As shown in Chapter 6, even if we outperformed the current state-of-the-art, we did not reach very high accuracy as with scans. Further improvements are necessary to increase the recognition rates on unconstrained photographs of leaves. An interesting idea might be to separate the different types of photos and process them independently (and maybe differently). For example, it is obvious that it could be extremely difficult to identify an image of a single leaf using images of dense foliage.

Multiple organs

Using multiple leaf images per plant can, obviously, improve the recognition rates. Also, considering images of different organs as well as leaves (e.g., flowers and fruits) could potentially improve recognition and render fine-grained categorization of plants of further interest to amateurs and botanists alike. One perspective is to extend our

framework to support other organs and design a procedure for resolving ambiguities by a form of improvised, on-line learning.

Non-botanical objects

So far, we have only considered botanical objects, especially leaves. Further work remains to be carried out towards extending our framework to support other kinds of biological objects. A short-term extension of this work is to test our classification framework on other fine-grained categories but using appropriate features and hierarchical representations of the data.

Bibliography

[1] M. A. Acevedo, C. J. Corrada-Bravo, H. Corrada-Bravo, L. J. Villanueva-Rivera, and T. M. Aide. Automated classification of bird and amphibian calls using machine learning: A comparison of methods. *Ecological Informatics*, 4(4):206–214, 2009.

[2] A. Angelova and S. Zhu. Efficient object detection and segmentation for fine-grained recognition. In *CVPR*, 2013.

[3] A. Angelova and S. Zhu. Efficient object detection and segmentation for fine-grained recognition. In *CVPR'13: IEEE Conference on Computer Vision and Pattern Recognition*, Portland, OR, June 23-28, 2013.

[4] A. Arora, A. Gupta, N. Bagmar, S. Mishra, and A. Bhattacharya. A plant identification system using shape and morphological features on segmented leaflets: Team iitk, clef 2012. In *CLEF (Online Working Notes/Labs/Workshop)*, 2012.

[5] A. Backes and O. Bruno. Plant leaf identification using color and multi-scale fractal dimension. In A. Elmoataz, O. Lezoray, F. Nouboud, D. Mammass, and J. Meunier, editors, *Image and Signal Processing*, volume 6134 of *Lecture Notes in Computer Science*, pages 463–470. Springer Berlin Heidelberg, 2010.

[6] B. S. Bama, S. M. Valli, S. Raju, and V. A. KUMAR. Content based leaf image retrieval (cblir) using shape, color and texture features. *Indian Journal of Computer Science and Engineering*, 2(2):202–211, 2011.

[7] P. Belhumeur, D. Chen, S. Feiner, D. Jacobs, W. Kress, H. Ling, I. Lopez, R. Ramamoorthi, S. Sheorey, S. White, and L. Zhang. Searching the world's herbaria: A system for visual identification of plant species. In *ECCV*, pages 116–129, 2008.

[8] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(4):509–522, 2002.

[9] T. Berg and P. N. Belhumeur. POOF: Part-based one-vs-one features for fine-grained categorization, face verification, and attribute estimation. In *Proc. Conf. Computer Vision and Pattern Recognition (CVPR)*, June 2013.

[10] A. Bosch, A. Zisserman, and X. Munoz. Image classification using random forests and ferns. In *ICCV*, 2007.

[11] B. E. Boser, I. M. Guyon, and V. N. Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory*, pages 144–152. ACM Press, 1992.

[12] L. Bourdev and J. Malik. Poselets: Body part detectors trained using 3d human pose annotations. In *ICCV*, 2009.

[13] S. Branson, C. Wah, B. Babenko, F. Schroff, P. Welinder, P. Perona, and S. Belongie. Visual Recognition with Humans in the Loop. In *ECCV*, Sept. 2010.

[14] F. Briggs, B. Lakshminarayanan, L. Neal, X. Z. Fern, R. Raich, S. J. Hadley, A. S. Hadley, and M. G. Betts. Acoustic classification of multiple simultaneous bird species: A multi-instance multi-label approach. *The Journal of the Acoustical Society of America*, 131(6):4640–4650, 2012.

[15] M. Burl and P. Perona. Using hierarchical shape models to spot keywords in cursive handwriting data. In *CVPR*, pages 535 –540, June 1998.

[16] C. Caballero and M. C. Aranda. Plant species identification using leaf image retrieval. In *Proceedings of the ACM International Conference on Image and Video Retrieval*, CIVR '10, pages 327–334, New York, NY, USA, 2010. ACM.

[17] C. Caballero and M. C. Aranda. Plant species identification using leaf image retrieval. In *Proceedings of the ACM International Conference on Image and Video Retrieval*, pages 327–334. ACM, 2010.

[18] M. Carvalho, F. Bockmann, D. Amorim, C. Brando, M. de Vivo, J. Figueiredo, and al. Taxonomic impediment or impediment to taxonomy? a commentary on

systematics and the cybertaxonomic-automation paradigm. *Evolutionary Biology*, 34:140–143, 2007. 10.1007/s11692-007-9011-6.

[19] D. Casanova, J. B. Florindo, and O. M. Bruno. Ifsc/usp at imageclef 2011: Plant identication task. In V. Petras, P. Forner, and P. D. Clough, editors, *CLEF (Notebook Papers/Labs/Workshop)*, 2011.

[20] D. Casanova, J. B. Florindo, W. N. Gonçalves, and O. M. Bruno. Ifsc/usp at imageclef 2012: Plant identification task. In *CLEF (Online Working Notes/Labs/Workshop)*, 2012.

[21] G. Cerutti, L. Tougne, J. Mille, A. Vacavant, and D. Coquin. Understanding leaves in natural images - a model-based approach for tree species identification. *Computer Vision and Image Understanding*, 117(10):1482–1501, 2013.

[22] G. Cerutti, L. Tougne, A. Vacavant, and D. Coquin. A parametric active polygon for leaf segmentation and shape estimation. In *Advances in Visual Computing*, pages 202–213. Springer, 2011.

[23] L.-B. Chang, Y. Jin, W. Zhang, E. Borenstein, and S. Geman. Context, computation, and optimal roc performance in hierarchical models. *International Journal of Computer Vision*, 93(2):117–140, 2011.

[24] N. R. Cook. *Confidence Intervals and Sets*. John Wiley and Sons, Ltd, 2005.

[25] J. Cope, P. Remagnino, S. Barman, and P. Wilkin. Plant texture classification using gabor co-occurrences. In G. Bebis, R. Boyle, B. Parvin, D. Koracin, R. Chung, R. Hammound, M. Hussain, T. Kar-Han, R. Crawfis, D. Thalmann, D. Kao, and L. Avila, editors, *Advances in Visual Computing*, volume 6454 of *Lecture Notes in Computer Science*, pages 669–677. Springer Berlin Heidelberg, 2010.

[26] J. S. Cope, D. Corney, J. Y. Clark, P. Remagnino, and P. Wilkin. Plant species identification using digital morphometrics: A review. *Expert Systems with Applications*, 39(8):7562 – 7573, 2012.

[27] J. S. Cope, P. Remagnino, S. Barman, and P. Wilkin. Plant texture classification using gabor co-occurrences. In *Advances in Visual Computing*, pages 669–677. Springer, 2010.

[28] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *CVPR (1)*, pages 886–893, 2005.

[29] M. Das, R. Manmatha, and E. Riseman. Indexing flower patent images using domain knowledge. *IEEE Intelligent Systems*, 14:24–33, 1999.

[30] J. J. del Coz, J. Díez, and A. Bahamonde. Learning nondeterministic classifiers. *Journal of Machine Learning Research*, 10:2273–2293, 2009.

[31] J. Deng, A. C. Berg, K. Li, and L. Fei-Fei. What does classifying more than 10,000 image categories tell us? In *Proceedings of the 11th European conference on Computer vision: Part V*, ECCV'10, pages 71–84, Berlin, Heidelberg, 2010. Springer-Verlag.

[32] J. Deng, J. Krause, A. Berg, and L. Fei-Fei. Hedging your bets: Optimizing accuracy-specificity trade-offs in large scale visual recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Providence, RI, USA, June 2012.

[33] J. Deng, J. Krause, and L. Fei-Fei. Fine-grained crowdsourcing for fine-grained recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2013.

[34] C. Direkoğlu and M. S. Nixon. Shape classification via image-based multiscale description. *Pattern Recognition*, 44(9):2134–2146, 2011.

[35] J.-X. Du, D. Huang, X. Wang, and X. Gu. Shape recognition based on radial basis probabilistic neural network and application to plant species identification. In *ISNN (2)*, pages 281–285, 2005.

[36] J.-X. Du, X. Wang, and G.-J. Zhang. Leaf shape based plant species recognition. *Applied Mathematics and Computation*, 185(2):883–893, 2007.

[37] K. Duan, D. Parikh, D. Crandall, and K. Grauman. Discovering localized attributes for fine-grained recognition. In *CVPR*, pages 3474–3481, 2012.

[38] R. El-Yaniv and Y. Wiener. On the foundations of noise-free selective classification. *Journal of Machine Learning Research*, 11:1605–1641, 2010.

[39] B. Ellis. *Manual of leaf architecture*. Cornell paperbacks. Published in association with the New York Botanical Garden, 2009.

[40] T. Elpel. *Botany in a Day: The Patterns Method of Plant Identification*. Thomas J. Elpel's herbal field guide to plant families of North America. Hops Press, 2004.

[41] X. Fan and D. Geman. Hierarchical object indexing and sequential learning. In *Proceedings of the Pattern Recognition, 17th International Conference on (ICPR'04) Volume 3 - Volume 03*, ICPR '04, pages 65–68, Washington, DC, USA, 2004. IEEE Computer Society.

[42] R. Farrell, O. Oza, N. Zhang, V. I. Morariu, T. Darrell, and L. S. Davis. Birdlets: Subordinate categorization using volumetric primitives and pose-normalized appearance. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 161–168. IEEE, 2011.

[43] R. Farrell, O. Oza, Z. Zhang, V. Morariu, T. Darrell, and L. Davis. Birdlets: Subordinate categorization using volumetric primitives and pose-normalized appearance. In *ICCV*, pages 161–168, 2011.

[44] P. Felzenszwalb, R. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part-based models. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(9):1627–1645, 2010.

[45] P. Felzenszwalb and J. Schwartz. Hierarchical matching of deformable shapes. In *CVPR*, pages 1–8, june 2007.

[46] M. Ferecatu. *Image retrieval with active relevance feedback using both visual and keyword-based descriptors*. PhD thesis, Université de Versailles SaintQuentin-en-Yvelines, 2005.

[47] R. Fergus, H. Bernal, Y. Weiss, and A. Torralba. Semantic label sharing for learning with many categories. In *ECCV (1)*, pages 762–775, 2010.

[48] A. Fernández and S. Gómez. Solving non-uniqueness in agglomerative hierarchical clustering using multidendrograms. *J. Classification*, 25(1):43–65, 2008.

[49] F. Fleuret and D. Geman. Stationary features and cat detection. *Journal of Machine Learning Research (JMLR)*, 9:2549–2578, 2008.

[50] H. Goëau, P. Bonnet, A. Joly, N. Boujemaa, D. Barthelemy, J.-F. Molino, P. Birnbaum, E. Mouysset, and M. Picard. The clef 2011 plant images classification task. In *CLEF (Notebook Papers/Labs/Workshop)*, 2011.

[51] H. Goëau, P. Bonnet, A. Joly, I. Yahiaoui, D. Barthelemy, N. Boujemaa, and J.-F. Molino. The imageclef 2012 plant identification task. In *CLEF (Online Working Notes/Labs/Workshop)*, 2012.

[52] E. Grall-Maes and P. Beauseroy. Optimal decision rule with class-selective rejection and performance constraints. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(11):2073–2082, 2009.

[53] T. A. P. GROUP. An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: Apg iii. *Botanical Journal of the Linnean Society*, 161(2):105–121, 2009.

[54] X. Gu, J.-X. Du, and X.-F. Wang. Leaf recognition based on the combination of wavelet transform and gaussian interpolation. In *Advances In Intelligent Computing*, pages 253–262. Springer, 2005.

[55] S. S. Gupta. On some multiple decision (selection and ranking) rules. *Technometrics*, 7(2):225–245, 1965.

[56] T. M. Ha. The optimum class-selective rejection rule. *IEEE Trans. Pattern Anal. Mach. Intell.*, 19(6):608–615, 1997.

[57] C. Harris and M. Stephens. A combined corner and edge detector. In *Alvey vision conference*, volume 15, page 50. Manchester, UK, 1988.

[58] T. Hastie and R. Tibshirani. Classification by pairwise coupling, 1998.

[59] T. HORIUCHI. class-selective rejection rules to minimize the maximum distance between selected classes. *Pattern Recognition*, 31(10):1579 – 1588, 1998.

[60] C. Im, H. Nishida, and T. L. Kunii. A hierarchical method of recognizing plant species by leaf shapes. In *MVA*, pages 158–161, 1998.

[61] A. K. Jain, M. N. Murty, and P. J. Flynn. Data clustering: a review. *ACM Comput. Surv.*, 31(3):264–323, Sept. 1999.

[62] N. Japkowicz and S. Stephen. The class imbalance problem: A systematic study. *Intelligent data analysis*, 6(5):429–449, 2002.

[63] F. V. Jensen. *An introduction to Bayesian networks*, volume 210. UCL press London, 1996.

[64] R. Jensen, K. Ciofani, and L. Miramont. Lines, outlines, and landmarks: Morphometric analyses of leaves of acer rubrum, acer saccharinum (aceraceae) and their hybrid. *Taxon*, 2002.

[65] A. Joly, H. Goëau, P. Bonnet, V. Bakić, J. Barbe, S. Selmi, I. Yahiaoui, J. Carré, E. Mouysset, J.-F. Molino, et al. Interactive plant identification based on social image data. *Ecological Informatics*, 2013.

[66] A. Khosla, N. Jayadevaprakash, N. Yao, and F. Li. Novel dataset for fine-grained image categorization. 2011.

[67] N. Kumar, P. N. Belhumeur, A. Biswas, D. W. Jacobs, W. J. Kress, I. C. Lopez, and J. V. Soares. Leafsnap: A computer vision system for automatic plant species identification. In *Computer Vision–ECCV 2012*, pages 502–516. Springer Berlin Heidelberg, 2012.

[68] N. Larios, H. Deng, W. Zhang, J. Sarpola, M.and Yuen, R. Paasch, A. Moldenke, D. Lytle, S. Ruiz-Correa, E. Mortensen, L. Shapiro, and T. Dietterich. Automated insect identification through concatenated histograms of local appearance features: feature vector generation and region detection for deformable objects. *Mach. Vis. Appl.*, 19(2):105–123, 2008.

[69] N. Larios, H. Deng, W. Zhang, M. Sarpola, J. Yuen, R. Paasch, A. Moldenke, D. Lytle, S. Correa, E. Mortensen, L. Shapiro, and T. Dietterich. Automated insect identification through concatenated histograms of local appearance features. In *Applications of Computer Vision, 2007. WACV '07. IEEE Workshop on*, pages 26–26, 2007.

[70] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *CVPR*, pages 2169–2178, Washington, DC, USA, 2006. IEEE Computer Society.

[71] P. Lee. *Bayesian statistics: an introduction*. Number v. 2 in A Charles Griffin Book. Oxford University Press, 1989.

[72] F.-F. Li and P. Perona. A bayesian hierarchical model for learning natural scene categories. In *CVPR (2)*, pages 524–531, 2005.

[73] Y. Li, Z. Chi, and D. D. Feng. Leaf vein extraction using independent component analysis. In *SMC*, pages 3890–3894, 2006.

[74] H. Ling and D. Jacobs. Shape classification using the inner-distance. *IEEE Trans. Pattern Anal. Mach. Intell*, 29:286–299, 2007.

[75] J. Liu, A. Kanazawa, D. W. Jacobs, and P. N. Belhumeur. Dog breed classification using part localization. In *ECCV (1)*, pages 172–185, 2012.

[76] O. M. Parkhi, A. Vedaldi, C. Jawahar, and A. Zisserman. The truth about cats and dogs. In *ICCV*, pages 1427–1434, 2011.

[77] A.-G. Manh, G. Rabatel, L. Assemat, and M.-J. Aldon. Weed leaf image segmentation by deformable templates. *Journal of agricultural engineering research*, 80(2):139–146, 2001.

[78] G. Martínez-Muñoz, N. L. Delgado, E. N. Mortensen, W. Zhang, A. Yamamuro, R. Paasch, N. Payet, D. A. Lytle, L. G. Shapiro, S. Todorovic, A. Moldenke, and T. G. Dietterich. Dictionary-free categorization of very similar objects via stacked evidence trees. In *CVPR*, pages 549–556, 2009.

[79] F. Mokhtarian, S. Abbasi, and J. Kittler. Efficient and robust retrieval by shape content through curvature scale space. *Series on Software Engineering and Knowledge Engineering*, 8:51–58, 1997.

[80] F. Mokhtarian and A. K. Mackworth. A theory of multiscale, curvature-based shape representation for planar curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(8):789–805, 1992.

[81] S. Mouine, I. Yahiaoui, and A. Verroust-Blondet. Advanced shape context for plant species identification using leaf image retrieval. In *ICMR*, page 49, 2012.

[82] S. Mouine, I. Yahiaoui, and A. Verroust-Blondet. A shape-based approach for leaf classification using multiscale triangular representation. In *Proceedings of the 3rd ACM conference on International conference on multimedia retrieval*, ICMR '13, pages 127–134, New York, NY, USA, 2013. ACM.

[83] S. Mouine, I. Yahiaoui, and A. Verroust-Blondet. A shape-based approach for leaf classification using multiscaletriangular representation. In *ICMR*, pages 127–134, 2013.

[84] O. Mzoughi, I. Yahiaoui, and N. Boujemaa. Extraction of leaf parts by image analysis. In *Proceedings of the 9th international conference on Image Analysis and Recognition - Volume Part I*, ICIAR, 2012.

[85] O. Mzoughi, I. Yahiaoui, N. Boujemaa, and E. Zagrouba. Advanced tree species identification using multiple leaf parts image queries.

[86] O. Mzoughi, I. Yahiaoui, N. Boujemaa, and E. Zagrouba. Automated semantic leaf image categorization by geometric analysis. In *Multimedia and Expo (ICME), 2013 IEEE International Conference on*, pages 1–6, July 2013.

[87] Y. Nam, E. Hwang, and D. Kim. A similarity-based leaf image retrieval scheme: Joining shape and venation features. *Computer Vision and Image Understanding*, 110(2):245 – 259, 2008.

[88] K. Nandakumar, Y. Chen, S. Dass, and A. Jain. Likelihood ratio-based biometric score fusion. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(2):342 –347, feb. 2008.

[89] J. Neyman. Outline of a theory of statistical estimation based on the classical theory of probability. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 236(767):pp. 333–380, 1937.

[90] M.-E. Nilsback and A. Zisserman. A visual vocabulary for flower classification. In *CVPR*, volume 2, pages 1447–1454, 2006.

[91] N. Otsu. A Threshold Selection Method from Gray-level Histograms. *IEEE Transactions on Systems, Man and Cybernetics*, 9(1):62–66, 1979.

[92] S. Paris, X. Halkias, and H. Glotin. Participation of lsis/dyni to imageclef 2012 plant images classification task. In *CLEF (Online Working Notes/Labs/Workshop)*, 2012.

[93] J. Park, E. Hwang, and Y. Nam. Utilizing venation features for efficient leaf image retrieval. *Journal of Systems and Software*, 81(1):71 – 82, 2008.

[94] L. Paulevé, H. Jégou, and L. Amsaleg. Locality sensitive hashing: a comparison of hash function types and querying mechanisms. *Pattern Recognition Letters*, 31(11):1348–1358, Aug. 2010. QUAERO.

[95] J. C. Platt. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In *ADVANCES IN LARGE MARGIN CLASSIFIERS*, pages 61–74. MIT Press, 1999.

[96] S. Prasad, K. M. Kudiri, and R. Tripathi. Relative sub-image based features for leaf recognition using support vector machine. In *Proceedings of the 2011 International Conference on Communication, Computing & Security*, pages 343–346. ACM, 2011.

[97] P. Prasong and K. Chamnongthai. Face-recognition-based dog-breed classification using size and position of each local part, and pca. In *Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 2012 9th International Conference on*, pages 1–5, 2012.

[98] C. Rother, V. Kolmogorov, and A. Blake. Grabcut: Interactive foreground extraction using iterated graph cuts. In *ACM Transactions on Graphics (TOG)*, volume 23, pages 309–314. ACM, 2004.

[99] R. Scotland and A. Wortley. How many species of seed plants are there? *Taxon*, 2003.

[100] A. R. Sfar, N. Boujemaa, and D. Geman. Identification of plants from multiple images and botanical idkeys. In *ICMR*, pages 191–198, 2013.

[101] A. R. Sfar, N. Boujemaa, and D. Geman. Vantage feature frames for fine-grained categorization. In *CVPR*, 2013.

[102] J. Shawe-Taylor and N. Cristianini. *Kernel Methods for Pattern Analysis*. Cambridge University Press, New York, NY, USA, 2004.

[103] C. N. Silla, Jr. and A. A. Freitas. A survey of hierarchical classification across different application domains. *Data Min. Knowl. Discov.*, 2011.

[104] O. Söderkvist. Computer vision classification of leaves from swedish trees. Master's thesis, Linköping University, SE-581 83 Linköping, Sweden, September 2001. LiTH-ISY-EX-3132.

[105] C.-H. Teng, Y.-T. Kuo, and Y.-S. Chen. Leaf segmentation, its 3d position estimation and leaf classification from a few images with very close viewpoints. In *Image Analysis and Recognition*, pages 937–946. Springer, 2009.

[106] B. Tversky and K. Hemenway. Objects, parts, and categories. *Experimental Psychology: General*, 1984.

[107] M. Vatsa, R. Singh, A. Ross, and A. Noore. Likelihood ratio in a svm framework: Fusing linear and non-linear face classifiers. In *Computer Vision and Pattern Recognition Workshops, 2008. CVPRW '08. IEEE Computer Society Conference on*, pages 1 –6, june 2008.

[108] A. Vedaldi, V. Gulshan, M. Varma, and A. Zisserman. Multiple kernels for object detection. In *ICCV*, pages 606–613, 2009.

[109] C. Wah, S. Branson, P. Perona, and S. Belongie. Multiclass recognition and part localization with humans in the loop. In *ICCV*, Barcelona, 2011.

[110] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong. Locality-constrained linear coding for image classification. In *CVPR*, pages 3360–3367, 2010.

[111] X.-F. Wang, J.-X. Du, and G.-J. Zhang. Recognition of leaf images based on shape features using a hypersphere classifier. In *Advances in Intelligent Computing*, pages 87–96. Springer, 2005.

[112] X.-F. Wang, D.-S. Huang, J.-X. Du, H. Xu, and L. Heutte. Classification of plant leaf images with complicated background. *Applied mathematics and computation*, 205(2):916–926, 2008.

[113] X.-F. Wang, D.-S. Huang, J.-X. Du, H. Xu, and L. Heutte. Classification of plant leaf images with complicated background. *Applied Mathematics and Computation*, 205(2):916 – 926, 2008. ¡ce:title¿Special Issue on Advanced Intelligent Computing Theory and Methodology in Applied Mathematics and Computation¡/ce:title¿.

[114] Z. Wang, Z. Chi, and D. Feng. Shape based leaf image retrieval. *Vision, Image and Signal Processing, IEE Proceedings*, 2003.

[115] J. Ward Jr. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301):236–244, 1963.

[116] J. Wu and J. Rehg. Where am i: Place instance and category recognition using spatial pact. In *CVPR*, 2008.

[117] J. Wu and J. M. Rehg. Where am i: Place instance and category recognition using spatial pact. In *CVPR*, 2008.

[118] S. Wu, F. Bao, E. Xu, Y.-X. Wang, Y.-F. Chang, and Q.-L. Xiang. A leaf recognition algorithm for plant classification using probabilistic neural network. *Ratio*, abs/0707.4(2):6, 2007.

[119] S. G. Wu, F. S. Bao, E. Y. Xu, Y.-X. Wang, Y.-F. Chang, and Q.-L. Xiang. A leaf recognition algorithm for plant classification using probabilistic neural network. In *Signal Processing and Information Technology, 2007 IEEE International Symposium on*, pages 11–16. IEEE, 2007.

[120] T. Wu and S. Zhu. A numerical study of the bottom-up and top-down inference processes in and-or graphs. *International Journal of Computer Vision*, 93(2):226–252, 2011.

[121] I. Yahiaoui, O. Mzoughi, and N. Boujemaa. Leaf shape descriptor for tree species identification. In *ICME*, pages 254–259. IEEE, 2012.

[122] B. A. Yanikoglu, E. Aptoula, and C. Tirkaz. Sabanci-okan system at imageclef 2012: Combining features and classifiers for plant identification. In *CLEF (Online Working Notes/Labs/Workshop)*, 2012.

[123] G. Yao, G. Bradski, and L. Fei-Fei. A codebook-free and annotation-free approach for fine-grained image categorization. In *CVPR*, Providence, RI, USA, June 2012.

[124] M. Yuan and M. H. Wegkamp. Classification methods with reject option based on convex risk minimization. *Journal of Machine Learning Research*, 11:111–130, 2010.

[125] N. Zhang, R. Farrell, and T. Darrell. Pose pooling kernels for sub-category recognition. In *CVPR*, pages 3665–3672, 2012.

[126] W. Zhang, A. Surve, X. Fern, and T. Dietterich. Learning non-redundant codebooks for classifying complex objects. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 1241–1248. ACM, 2009.

[127] A. Zweig and D. Weinshall. Exploiting object hierarchy: Combining models from different category levels. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–8, 2007.

Fine-Grained Object Categorization: Plant Species Identification

Asma REJEB SFAR

RESUME : Nous étudions la problématique de classification dite *fine* en se concentrant sur la détermination des espèces botaniques à partir d'images de feuilles. Nous nous intéressons aussi bien à la description et la représentation de l'objet qu'aux algorithmes de classification et des scénarios d'identification utiles à l'utilisateur. Nous nous inspirons du processus manuel des botanistes pour introduire une nouvelle représentation hiérarchique des feuilles. Nous proposons aussi un nouveau mécanisme permettant d'attirer l'attention au tour de certains points caractéristiques de l'objet et d'apprendre des signatures spécifiques à chaque catégorie.
Nous adoptons une stratégie de classification hiérarchique utilisant une série de classifieurs locaux allant des plus grossiers vers les plus fins; la classification locale étant basée sur des rapports de vraisemblance. L'algorithme fournit une liste d'estimations ordonnées selon leurs rapports de vraisemblance. Motivés par les applications, nous introduisons un autre scénario proposant à l'utilisateur un ensemble de confiance contenant la bonne espèce avec une probabilité très élevée. Un nouveau critère de performance est donc considéré: la taille de l'ensemble retourné. Nous proposons un modèle probabiliste permettant de produire de tels ensembles de confiance. Toutes les méthodes sont illustrées sur plusieurs bases de feuilles ainsi que des comparaisons avec les méthodes existantes.

MOTS-CLEFS: Classification fine, représentation hiérarchique, ensemble de confiance, identification de plantes

ABSTRACT: We introduce models for fine-grained categorization, focusing on determining botanical species from leaf images. Images with both uniform and cluttered background are considered and several identification scenarios are presented, including different levels of human participation. Both feature extraction and classification algorithms are investigated.
We first leverage domain knowledge from botany to build a hierarchical representation of leaves based on *IdKeys*, which encode invariable characteristics, and refer to geometric properties (i.e., landmarks) and groups of species (e.g., taxonomic categories). The main idea is to sequentially refine the object description and thus narrow down the set of candidates during the identification task. We also introduce *vantage feature frames* as a more generic object representation and a mechanism for focusing attention around several vantage points (*where to look*) and learning dedicated features (*what to compute*).
Based on an underlying coarse-to-fine hierarchy, categorization then proceeds from coarse-grained to fine-grained using local classifiers which are based on likelihood ratios. Motivated by applications, we also introduce on a new approach and performance criterion: report a subset of species whose expected size is minimized subject to containing the true species with high probability. The approach is model-based and outputs a *confidence set* in analogy with confidence intervals in classical statistics. All methods are illustrated on multiple leaf datasets with comparisons to existing methods.

KEY-WORDS: Fine-grained categorization, hierarchical representation, confidence set, plant identification.

TELECOM ParisTech

EDiTE

ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY