POLITECNICO DI TORINO

Doctoral School

Department of Control and Computer Engineering – XXVII cycle

PhD Thesis

Feature Fusion for Pattern Recognition

Ihtesham Ul Islam

Tutor
Prof. Andrea Bottino

Coordinator of the PhD Program
Prof. Matteo Sonza Reorda

February 2015

to my Parents and Uncle(late)

Abstract

Information or data fusion is one of the solutions adopted for improving the performance of a pattern recognition system. Information can be gathered either from multiple data sources or through the use of multiple representations generated from a single data source. A single representation summarizes the information and provides a single cue on the data, and thus may not be able to fully reveal the inherent characteristics of the data. In visual recognition, image representations are generally categorized into global and local based types. A global representation captures features corresponds to some holistic characteristic in the image, and produces a coarse representation. Differently, a local representation reveals detail variations and traits inherent to the image. Psychological findings have shown that humans equally rely on both local and global visual information. Moreover, there is a large agreement in literature that the combination of different features, i.e. a multiview perspective, can have a positive effect on the performance of a pattern recognition system. In fact, different features can represent different and complementary characteristics of the data; in other words, each feature set represent a different view on the original dataset. Thus it is expected that a visual recognition system can benefit from different representations (both local and global) through the use of information fusion.

Information can generally be consolidated at three different levels: (i) decision level; (ii) match score level; and (iii) feature level. In the literature match level and decision level fusion (i.e. combining the output of different classifier, each of them working on different feature sets) have been extensively studied, whereas feature level fusion is a relatively understudied problem because of the difficulties inherent to its correct implementation. Feature level fusion may incorporate redundant, noisy or trivial information and the concatenated feature vectors may lead to the problem of curse of dimensionality. In addition, the feature sets may not be compatible and relationship between different feature spaces may not be known. Moreover, this integration comes at a cost, which may incur in units of time, computational resources or even money. Nevertheless, it is thought that fusing features at this level would still retain a richer source of discriminative information.

Motivated by the belief, this thesis investigates the use of feature level fusion and its correlation with feature selection and classification tasks for two recent pattern recognition problems. These include the classification of six types of HEp2 staining patterns and the automatic verification of kinship relations in a pair of face images. several image attributes are proposed, that are better capable of characterizing the

different kind of images associated with the two said classification tasks. Feature level fusion of the different attributes is performed followed by a careful reduction of features, through the use of pertinent feature selection and classification algorithms, that decide the most representative and discriminative feature sets for the patterns to classify. Results indicate that the proposed techniques working on the combination of features of different natures, which are capable of describing the data under different perspectives, is an effective strategy in achieving higher accuracy.

Acknowledgements

All praise and glory be to God, *the Most Gracious and the Most Merciful*, for giving me directions, opportunities and ease in life. I am thankful to Politecnico Di Torino for its good research facilities and working environment provided to me. This enabled me to enrich my knowledge and carry out my research. Similarly, I would like to express gratitude to Higher Education Commission (HEC), Government of Pakistan, for awarding me a scholarship that helped in expanding the breadth of my knowledge to the next level.

My appreciation and heartfelt gratitude goes to my professor and thesis adviser Prof. Andrea Bottino for his constant support, guidance and the time he devoted for my research work. I have learned so much from him and he will remain a source of inspiration to me in my career. I am thankful to Santa di Cataldo, Prof. Elisa Ficarra and Tiago Vieira for a valuable collaborative work. I am also thankful to Daniele Pianu for helping me on my work.

I would give big thanks to all my labmates Andrea, Christian, Oscar, Ali, Antonio, Luca, Federico, Najeeb and Iacopo. I always found them very respectful, friendly and helpful. In addition, I would like to say thanks to Omar shoaib, Haroon, Bilal, Gasser and Saroosh for their sincere friendship and encouragement.

It would be unfair if I don't mention the importance of my all time friends and flat mates in Italy. I am so grateful to my friends Javed, Anees, Affaq, Ali, Khalil, Omar Osman, Fahad, Adnan, Tariq and Ijaz, they made me feel at home. I have learnt so much from them and without their company this PhD would not be easy for me.

Finally I would mention my father, whom I see as my guide in life. He has always supported me with love, affection and advice, which keeps me going. From the bottom of my heart I appreciate my mother, father, sisters and all the loving family members for their support, patience and encouragement they gave me in pursuing my career.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Introduction and Motivation

Pattern recognition systems automatically recognize patterns in sensed data. A pattern recognition system design deals with the mathematical and technical aspects of determining facts from sensor data. Typical components of a pattern recognition system are data acquisition, preprocessing, feature extraction and selection and classification.

Pattern recognition applied to large-dimensional data, as those typical in computer vision and pattern recognition applications, requires the extraction of features capable of summarizing the information to be analyzed. The choice of the most suited feature extraction algorithms is very crucial for the design of a robust pattern recognition system. In general, feature extraction plays a key role in any pattern classification task. At this stage, a proper representation is chosen to make the subsequent processing not only computationally feasible but also robust to possible data variations. A single or unimodal representation utilizes information obtained from a single cue on the data. Every representation has an upper bound on performance (or lower bound on its error rates) due to its inherent characteristics even when operating in a constrained environment. The performance of the system can only be as good as that of this unimodal representation itself. Thus a single representation may not be able to fully capture all the characteristics of the data. A good representation should be well equipped to handle noise in real world data which may be caused during the data acquisition process. Furthermore, it should be capable of revealing significant inter-class variation while exhibiting minimal intra-class variation. However in reality, it is unlikely that a single image modality is capable of providing all these properties, due to its inherent limitations and real world constraints like

environmental conditions.

Broadly speaking, image representations can be categorized into two general types: global-based and local-based. A global based representation produces features that correspond to some holistic characteristic in the image, where each dimension of the feature vector contains the information embodied in every part (or pixel) of the image. It is proper to use global features for coarse representation. On the other hand, in a local based representation the features reveal and capture more detailed variations and traits within some local image regions. Local features are often used for finer representation. Global features can be revealed by several subspace based methods including the state of the art automatic methods like Principal component Analysis (PCA) and Linear Discriminant Anaylsis (LDA) for feature extraction. Similarly, local features for images can be acquired using some well known techniques like Gabor filters [125] and Local binary patterns (LBP) [70]. Many global and local representations have been proposed in the literature, but it still remains an open research problem that what is the most suitable representation for a particular pattern recognition task. However, psychological findings show that humans equally rely on both local and global visual information [32]. Put differently, global and local features play different roles in the process of visual perception and recognition; and therefore, it is natural to expect better performances by fusing together appropriate global and local information.

In fact different features can represent different and complementary characteristics of the data; in other words, each feature set represent a different view on the original dataset. Understanding which features can be more effective in tackling a specific computer vision or a pattern recognition problem is a very difficult task, but there is a large agreement in the literature that the combination of different features, i.e. a multiview perspective, can have a positive effect. Tanaka et al. [34] suggests that natural vision-based classification tasks are performed better when multiple visual cues can be combined to reduce ambiguity. Thus, we expect that by combining multiple cues (in the form of global or local information) through an integration scheme, we will achieve a better performance, i.e., higher classification performance and higher robustness.

In the computer vision and pattern recognition literature some authors have suggested different methods to combine or fuse information derived from different cues. Either way, fusing information may provide increased interpretation capabilities and more reliable results since data with different characteristics and complementary information is combined. In general, data fusion techniques can be classified into three different levels according to the stage of processing at which the fusion takes place; feature level, match score level and decision level. They are also respectively categorized as low-level, mid-level and high-level integration. It is believed that pattern

recognition systems that fuse data at an earlier stage (such as feature level) will outperform those which fuse data at latter stages. This is because the data at lower levels have richer proportions of information about the class or the identity [33].

Having said that, however, an approach with feature level fusion needs to face several issues. First, directly encoding the multiple features can result in incorporating redundant, noisy or trivial information, which can seriously affect the performances of the recognition process. Second, the integration of multiple feature sets increase the number of variables used to represent the original data, and the concatenated feature vectors may lead to the problem of curse of dimensionality. Third, different modalities can have varying (sometime unknown) confidence levels in accomplishing different tasks, multiple modalities may have incompatible feature sets and the relationship between different feature spaces may not be known. Finally, the integration of multiple features comes at a cost, which may incur in units of time, computational resources or even money.

In order to achieve gain in performances through fusing information at feature level, we have to be equally careful to deal with the problems it carries. Provided that we can increase the number of features by extracting and integrating the right combination of global and local features, thereby increasing the heterogeneity of the feature vectors, the performance of a pattern recognition system is likely to boost. However, in reality, it reaches a maximum and then starts to degrade, suggesting that the final vectors may suffer from the 'peaking phenomenon'or as the 'curse of dimensionality'. Primarily, the use of dimensionality reduction techniques can mitigate the 'curse of dimensionality'along with the reduction of computational time and measurement cost. Nevertheless, care should be taken in this practice as an excessive reduction in the number of features may result in the loss of discriminative capability of the feature set and thereby reducing the accuracy performance of the pattern recognition system. Ugly duckling theorem [31] emphasizes the importance of selecting the right set of uncorrelated and discriminative features, as it is possible to incorrectly classify two different patterns as belonging to the same class if their feature sets have a sufficiently large number of redundant features.

At this stage, dimensionality reduction can be performed in two different ways. It can be achieved through a feature extraction algorithm that transforms the existing fused features into a lower dimensional space, where the classification task is made easier. In other words, creating a subset of new features by combinations of the existing features. In this regard, several linear (e.g., PCA or LDA) or non-linear (e.g., Subclass Discriminant Analysis, SDA [54]) dimensionality reduction techniques can be put to use. Secondly, the dimensionality can be reduced using a feature selection scheme, where a subset of the existing features is selected without a transformation. Feature selection, also called feature subset selection (FSS), looks at the issue of

dimensionality reduction from a different perspective. It selects a minimal set of relevant features that contains maximum discriminatory information while eliminating the irrelevant and noisy ones. It may be necessary in a number of situations where features may be expensive to obtain or one may want to extract meaningful rules from a classifier. In addition, it may provide improved generalization capabilities and reduced complexity and run-time.

Feature selection algorithms fall into two general categories: wrappers and filters. While wrappers select features in combination with a learning process, filters are independent from a learning algorithm. Usually wrappers provide the best performing feature set for a particular type of model; nevertheless, they tend to be slow and thus may not scale to larger datasets with many features. To design an efficient and low cost feature selection algorithm, which is scalable to big datasets with larger feature sets, some works have suggested the combination of filter and wrapper based methods. Peng et al. [74], proposes a filter based feature selection method to select relevant features according to the maximal statistical dependency criterion based on mutual information. They further suggested the use of an efficient two-stage feature selection algorithm, by combining their filter based method with wrapper based feature selectors, to identify a compact set of features that leads to superior classification performance.

Compared to the abundance of research related to fusion at decision or match score level (i.e. combining the output of different classifier, each of them working on different feature sets), fusion at the feature level is relatively understudied problem, mainly due to its intrinsic difficulties. Nevertheless, it is believed that feature level fusion contains a rich amount of discriminative information. Given the said motivations, this thesis work exploits several single feature attributes, its fusion at feature level, and its correlation with suitable feature selection and classification algorithms to investigate two vision-based classification problems. These include a multi-classification task to categorize six types of cell images and a two-class problem to automatically identify kin or not-kin relation in pairs of face images.

Here, we first briefly introduce these two classification problems related to pattern recognition and computer vision domains and then provide key highlights to our research contributions, which are discussed in details in the relevant sections of the thesis.

Classification of HEp2-Cells: Human Epithelial (HEp-2) cells are commonly used in Indirect Immunofluorescent (IIF), a standard medical test, for the diagnosis of several autoimmune disorders. The specialist observes images, containing cultured cells of the HEp-2 cell line, acquired under a fluorescence microscope and makes a diagnosis based on two things i) on the perceived intensity of the fluorescence signal, that can either be negative, intermediate or positive, and ii) secondly on the type

of the staining pattern. The task of the expert is to decide whether there are some symptoms of disease in the patient's serum or not. This decision is made according to visual patterns, also called staining patterns, of all cells in the image. Literature describes six main fluorescent patterns that are here reported for both positive and intermediate levels of intensity. They are homogeneous, nucleolar, speckled (coarse and fine), centromere, and cytoplasmic respectively. Their correct description is fundamental for the differential diagnosis of the pathologies. The manual technique, based on human visual inspection, is time-consuming, subjective and dependent on the operator's experience. The idea here is to automate the process which may be a solution to these limitations, making the process faster and more reliable.

Kinship Verification: Automatic Kinship verification aims at recognizing the degree of kinship of two individuals from their facial images. In other words, when analyzing the facial images of two individuals, the computer must tell whether those subjects are related or not. Different degrees of kinship might be detected, such as grandparent-child, parent-child, siblings and so on. Unlike typical face verification, Kinship verification aims at identifying similar features between two different individuals. This is a challenging task since it has to deal with differences in age and gender between subjects. Additionally, it has to cope with the unpredictable amount of genetic information shared by relatives that reflect into individuals, showing different degrees of facial similarity. Many other difficulties arise when Kinship Verification must be applied to databases acquired under unconstrained environments. Kinship Verification has potential applications in image retrieval and annotation, forensics and historical studies.

Our work aimed at performing a comprehensive exploration of feature level fusion for the two recent and important image classification problems with several pattern recognition techniques. A brief discussion of the main contribution of our work is following. First, we suggest a set of features that are capable to correctly describe the images. For the cell classification problem, we propose a set of features to characterize cell images that are highly discriminative with respect to their fluorescence patterns. In particular, we investigated image attributes derived from morphological, global and local texture analysis. Similarly, for kinship verification, we identified several facial attributes more fit to tell kins from not kin-pairs; these attributes are derived from different geometric and both local and global textural data of the face images. For both the problems, feature sets are obtained through a detailed analysis of single image attributes, as well as of their integration at the feature level, and of the selection of the most relevant feature variables. For the cell classification problem, we propose a combination of a features selection process and of a discriminant analysis that allows effectively in reducing the high dimensionality of the data, and at the same time coping with the high intra-class variance of the samples. These include Minimum-Redundancy Maximum-Relevance methodology

(mRMR) [74] as the feature selection algorithm and Subclass Discriminant Analysis (SDA) [54], to overcome the limitations of the traditional LDA approach. The classification is based on k-Nearest Neighbour (k-NN) algorithm. Differently, for kinship verification task, a kin image pair classification is based on Support Vector Machines (SVM) [19] along with the integration of a robust two-step feature selection algorithm. The feature selection is a combination of a filter and wrapper based method i.e., respectively based on mRMR and Sequential Forward Selection (SFS) [24] algorithms. The results of our experiments show that the combination of features, at the feature level, of different nature provides higher performance than when the features are used individually. In addition, the higher the heterogeneity of the features used, the better the performance. More importantly, the use of a proper feature selection scheme or a feature transformation or both along with the choice of a suitable classifier is crucial in case of feature level fusion, which is likely to boost the classification performance of the pattern recognition system.

1.2 Thesis Organization

This thesis is structured as follows. Chapter 2 provides an overview of the work related to data fusion in general and feature level fusion in specific. Chapter 3 first introduces the HEp2 cells classification problem and provides related work. It then presents our proposed method conducting the detail analysis of several suitable features attributes and their combination capable of describing HEp2 staining patterns along with the use of pertinent feature selection and classification algorithms and their rationale. This chapter also includes a preliminary work on HEp2-cell segmentation. In Chapter 4 we first give an introduction to Kinship Verification and the relavant related works. This is followed by work that shows the significance of a proposed method to automatically discriminate Parent-Child pairs by providing a detailed discussion of several experimental results. Kinship verification is revisited in Chapter 5, with the introduction of a new set of features capable of performing well in an unconstrained environment. Finally, conclusions are drawn in Chapter 6. This structure of the thesis was adopted to maintain coherence with the chronology of the work developed.

Chapter 2

Background and Literature Review

Generally speaking, the phrase information fusion includes any area which deals with exploiting a combination of different sources of data and knowledge, either to generate a single representational format or to reach a decision. In biometrics or visual information processing, information fusion involves the synergistic combination of data from multiple sources (sensors) or multiple feature representations to provide more reliable and accurate information. In other words, inadequacies in the data can be complemented by data from other sources or through different representations of the same data. A fusion system should integrate data in such a way that it remove the influence of conflicting or irrelevant data, increase the interpretation of data, and thereby accurately and concisely represents the original information. Provided that a proper choice of the fusion method is made for the kind of data, and is performed correctly, the system should become less sensitive to noise and its performance is likely to boost.Pioneering work on the subject related to multi-sensor fusion can be traced back to the early 80s [100, 98, 99]. In literature, information fusion for a pattern recognition system can occur at three different levels; namely feature level, match or score level and decision level. They can also be respectively called as low level, mid level and high level fusion. Table 2.1 gives a brief comparsion of the fusion levels.

Decision level fusion is a high level fusion, which refers to the combination of accept/reject decisions already taken by the individual classifiers operating on multiple data or different representations of the single source data. In order words, separate feature extraction and selection would be performed for the original data sources and then passed to their respective classifiers to make accept/or reject decisions. Finally, the multiple decisions are combined to arrive at the final classification

Table 2.1: A comparsion of the three fusion levels

Characteristic	Decision Level	Score Level	Feature Level
Information content	low	moderate	rich
Efficiency	low	medium	high
Difficulty of realization	small	moderate	large

decision. This can also be termed as post-classification fusion or mixture of experts. There are a number of methods to combine these accept/reject decisions in order to converge on a final classification decision. These include Majority voting, behavior knowledge space, and weighted voting based on Dempster-Shafer theory of evidence. A more detailed description of the methods can be found in [102]. Decision level fusion has been utilized in a number of applications, a significant number of which are related to multi-modal biometrics [104, 105, 106, 107, 108, 109, 110]. Fusion at the decision level is considered to be too rigid since only a limited amount of information is available at this level; therefore, in most cases it does not provide a substantial increase in performance [101].

Match score level fusion, also termed as a measurement or confidence level fusion, refers to the combination of matching scores provided by different classifiers. These classifiers are trained separately using different feature sets; feature sets can be obtained either through multiple representations in a unimodal system or using mulitple source of data in a multi-modal system settings. The match scores can be merged in number of ways like the max rule, min rule, sum rule, mean rule, median rule, etc [103]. Ross et al. [101] use decision trees and linear discriminant classifiers for combining multiple match scores. Fusion at this level requires a moderate level of processing in order to achieve an appropriate level of system performance. In literature, score level fusion is utilized by a large number of applications including both uni-modal and multi-modal biometric applications [136, 114, 111, 112, 113, 115, 116, 117, 118].

Feature level fusion uses features sets generated separately from different feature extraction algorithms on the same or separate datasets, which are fused together to produce a single representation of the system. A common way to fuse the features sets is to simply concatenate them to form a single feature vector [119, 121, 123], although for some data representations a single resultant feature vector can be calculated as a weighted average of the individual feature vectors e.g., features obtained from data of identical sensors. Incompatible feature sets, where there are differences in the range of values and distribution of the individual feature vectors, needs to be normalized before concatenation. The combined feature sets can be used instead of the original data for further processing. The new feature vector now has

a higher dimensionality and represents the input data in a different, and hopefully more discriminating, hyperspace. Nevertheless, fusion at the feature level is hard to achieve in practice because of the following reasons: (i) the feature sets from multiple modalities/representations may not be compatible; (ii) the relationships between these feature sets in a hyperspace may not be known; (iii) the integration of multiple feature sets increase the number of variables used to represent the original data, resulting in a feature vector with very large dimensionality leading to the 'curse of dimensionality' problem; and (iv) the combination of multiple features comes at a cost, which may incur in units of time, computational resources or even money.

In order to mitigate the problems associated with feature level fusion, the proper choice of a feature selection and/or a transformation algorithm is crucial for the performance of a pattern recognition system. Thus subsequent steps may include feature reduction techniques, based on feature selection and transformation, before sending the feature vector to the decision or classification module. A feature transformation transforms the existing fused features into a lower dimensional space, where the classification task is made easier. This may include linear (e.g., PCA or LDA) or non-linear (e.g., Subclass Discriminant Analysis, SDA [54]) feature transformation techniques. On the other hand, a feature selection process selects a minimal set of relevant features that contains maximum discriminatory information while discarding many redundant or irrelevant features. It can be performed either prior or after the fusion to extract useful features from the larger set of features.

In general, feature selection is an important problem for many pattern classification systems e.g. bioinformatics and object recognition. Although, some classification algorithms have an inherent ability to select and use relevant variables [25], feature selection algorithms in general can be categorized in two broader types: wrappers and filters. Wrappers use a learning algorithm to evaluate the usefulness of features [126, 128, 127]. They employ cross validations scheme to estimate the accuracy of the feature subsets, and usually provide the best performing feature set for that particular type of model. A disadvantage is that they tend to very slow and do not scale to big datasets with larger feature sets. Differently, filters [130, 74, 129] evaluate features according to heuristics based on general characteristics of the data. They work independently from a learning process and provide a ranking on the feature variables. Thus computationally efficient they can be applied to huge datasets with larger feature sets. Peng et al. [74], proposes an efficient filter based feature selection method to select relevant features according to the maximal statistical dependency criterion based on mutual information. Further they suggest the use of a two-stage feature selection algorithm by combining their filter algorithm with other wrapper based methods, to identify a compact set of features that leads to better classification accuracies.

While feature fusion at high levels, match score and decision levels, has been extensively studied in the literature, fusion at the feature level is a relatively understudied problem. Although, integrating feature at the lower level is hard to achieve due to its intrinsic difficulties, information fusion at the feature level is believed to the best type of fusion. It is expected to provide better recognition results than other levels of fusion since the feature sets contain richer source of information about the input data than the fusion at the match score or decision level [113]. There is some work on fusing different biometric modalities at feature level (e.g. Hand and Face [119], ear and profile face [120]), but some studies concentrate on fusing different representations of a single underlying modality (e.g. Gender Classification [121], multi-feature fusion using subspace learning [122], 3D Face recognition [135], cell image classification [51]). Our work belongs to this category fusing different representations of a single underlying modality.

Selecting appropriate and complementary component features, carrying discriminant information about the input data and are resistant to all kind of variations, is crucial for good performance and is still an open research area. In image classification tasks, treating images the way humans do, image representations can be categorized into two general types. A global representation depicts the general and holistic characteristics of the image and is often used for a coarse representation. On the other hand, a local representation captures local features and encodes more detailed variations within some local image regions, resulting in a finer representation. Although many successful image representation methods based on global or local features have been proposed (PCA [124], Gabor features [125], Local binary patterns [70]), it remains an open problem that what is the most suitable representation for a particular pattern classification task. However, in the literature psychological findings have shown that both global and local features are crucial for visual information perception [32]. In other words, both play different roles in the visual perception and recognition; therefore it is essential to fuse them together in a smart way.

In this thesis we target image classification problems which involve the computer processing of images; simulating how humans would process and understand images. To this purpose, we utilize data fusion, since psychological studies (e.g., Tanaka et al. [34]) suggest that natural vision-based classification tasks are performed better when multiple visual cues can be combined to reduce ambiguity. More specifically, we are interested in exploiting feature level fusion together with feature selection and transformation procedure to elicit a minimal feature set from the high-dimensional feature vector, for two recent pattern recognition problems. This includes a multi-class problem of HEp2 staining patterns classification and a two-class problem to automatically recognize kin-relations from pairs of face images. In both studies, we choose the most relevant and complementary component features for the tasks.

For the staining pattern classification, we run feature selection based on minimum-Redundancy- Maximum-Relevance (mRMR) algorithm [74] and perform feature transformation based on Subclass Discriminant Analysis (SDA) [54]. The classification of the final vectors is performed by k-Nearest Neighbour (k-NN). Differently, for kinship verification problem, we perform a two-step feature selection algorithm based respectively on mRMR and Sequential Forward Selection (SFS) algorithm [24]. The classification is based on SVM [19]. In our work, we validate our claims to utilize feature level fusion for the two said applications by conducting several experiments and by providing comparisons to existing approaches.

Chapter 3

Classification of HEp2 Staining Patterns in Indirect Immunofluorescence Images

Classifying HEp-2 fluorescence patterns in Indirect Immunofluorescence (IIF) HEp-2 cell imaging is important for the differential diagnosis of autoimmune diseases. The current technique, based on human visual inspection, is time-consuming, subjective and dependent on the operator's experience. Automating this process may be a solution to these limitations, making IIF faster and more reliable. This work proposes a classification approach based on Subclass Discriminant Analysis (SDA), a dimensionality reduction technique that provides an effective representation of the cells in the feature space, suitably coping with the high within-class variance typical of HEp-2 cell patterns. In order to generate an adequate characterization of the fluorescence patterns, we investigate the individual and combined contributions of several image attributes, showing that the integration of morphological, global and local textural features is the most suited for this purpose. The proposed approach provides an accuracy of the staining pattern classification of about 90%.

A preliminary version of this work is presented in [55], where we first analysed the use of SDA to approach the staining pattern classification problem, using textural descriptors based on Gray-Level Co-occurrence Matrices (GLCM) and Discrete Cosine Transform (DCT). Here we present a deeper insight into the problem aimed at obtaining a better characterization of the staining patterns and a larger set of experiments supporting our findings.

The rest of the chapter is organized as follows. An introduction to the Problem

and related research work is presented in section 3.1. After providing a full characterization of the HEp-2 image dataset in Section 3.2, in Section 3.3 we present the main steps of our proposed solution. In Section 3.4 we describe the experimental protocol and in Section 3.5 we show and discuss our experimental results. Finally a preliminary work on cell segmentation is provided in Section 3.6.

3.1 Introduction and Related Work

Indirect Immunofluorescence (IIF) is a widespread microscopy imaging technique for the detection of antinuclear auto-antibodies (ANA), which can reveal the presence of important autoimmune pathologies such as systemic rheumatic diseases, multiple sclerosis and diabetes [35]. The ANA-screening is typically performed by visually inspecting cultured cells with a fluorescence microscope. The patient's serum is first dispensed on a HEp-2 cell substrate and then diluted and incubated. The auto-antibodies of the serum selectively bond with specific antigens on the substrate and are finally revealed by a fluorescence tag.

The specialists usually perform a three-step analysis. First, the slide is validated by checking the presence of at least one fluorescence mitotic cell. Second, the intensity of fluorescence signal is evaluated according to three levels: negative (*i.e.* absence of fluorescence), intermediate or positive. Finally, the cells of the intermediate and positive slides (Figure 3.1) are classified on the basis of the pattern of the fluorescence signal, which in turn reveals the auto-antibody type. According to literature, the different staining patterns can be classified into six main groups, namely *centromere, homogeneous, nucleolar, coarse speckled, fine speckled* and *cytoplasmatic* (see Figure 3.2).

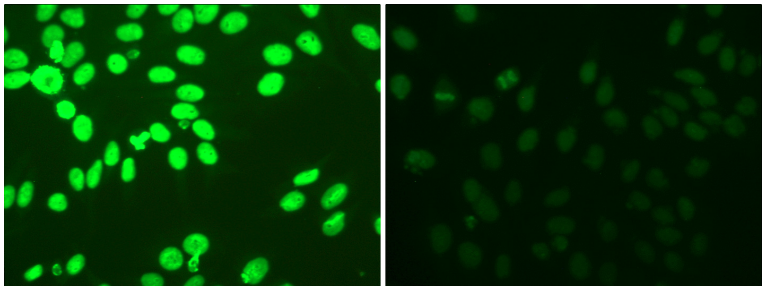


Figure 3.1: HEp-2 IIF images with positive (left) and intermediate (right) fluorescence intensity.

The accurate classification of the staining patterns is very important for differential diagnosis, since different patterns are associated with different types of

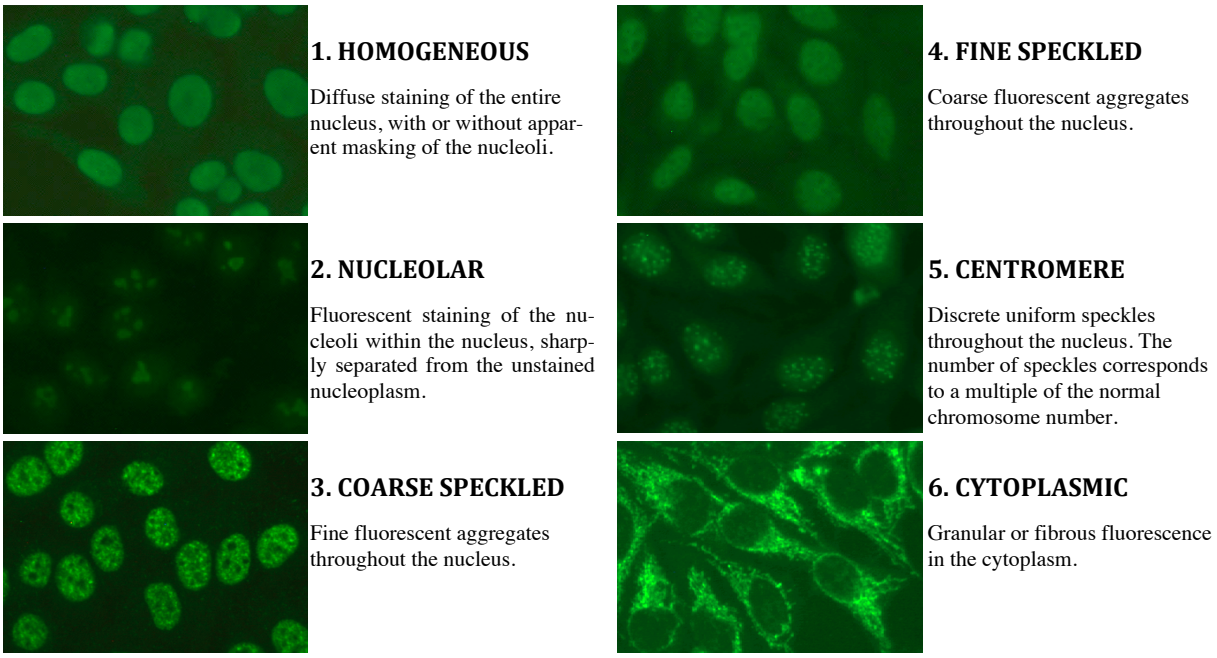


Figure 3.2: six types of staining patterns with positive and intermediate intensity levels.

autoimmune diseases. Nevertheless, in the standard practice this process suffers from intrinsic limitations related to the visual evaluation performed by human subjects. The analysis of large volume of images is a tedious and time-consuming task and requires highly trained and specialized personnel. Moreover, human evaluation for this type of screening is affected by very high inter laboratory variability (up to 24%, as reported in [36, 37]). This has tremendous impact on the reliability and reproducibility of the obtained results.

Computer-Aided Diagnosis (CAD) systems may overcome these limitations and effectively support the decision of the specialists. In particular, the automation of the staining pattern classification process may considerably reduce the time and effort required by the analysis and, at the same time, improve its repeatability. This would make IIF analysis easier, faster and more reliable.

In the recent years, the computer analysis of IIF images has attracted many researchers from the Pattern Recognition community proposing novel contributions to the different aspects of the analysis process. Yet, compared to the advancement in other medical imaging techniques, the automatic analysis of Indirect Immunofluorescence (IIF) images has not yet reached to a significant level of performance. Triggered by a growing demand for reliable CAD systems, recent research has proposed

solutions for all the major steps of IIF analysis, including methods for (i) the automated segmentation of the HEp-2 cells [38, 39, 40], (ii) the recognition of the mitotic cells in the slides [41, 42], (iii) the quantification of fluorescence intensity [43] and (iv) the automated classification of the staining patterns [44, 45, 46, 47, 48, 49, 50]. In particular, most of the recent efforts are focused on the latter task and the proposed classification schemes span the entire spectrum of machine learning (*e.g.*, learning vector quantization [44], decision tree induction algorithms [45, 46], support vector machines [49], random forests [50], self-organizing maps [48] and multi-expert systems [47]). The image attributes used to characterize the fluorescent patterns include several types of descriptors, with special focus on morphological and textural features [51, 52].

In spite of the recent extensive research, the accurate classification of the staining patterns still remains a challenge. Moreover, the comparison of the solutions presented in the literature is extremely difficult because they are based on different datasets and different experimental protocols. A growing interest to the computer analysis of Immunofluorescence (IIF) images has led to the organization of the first HEp-2 Cell Classification Contest by the International Conference on Pattern Recognition (ICPR2012), for a comparative evaluation on a common database. With the release of a public dataset of HEp-2 cell images [53], a Special Issue of the Pattern Recognition journal was dedicated on the Analysis and Recognition of Indirect Immunofluorescence Images. It was aimed at providing a more comprehensive assessment of the state of the art in the field of the staining pattern classification and simultaneously bounding the researchers to a standardized experimental procedure enabling the direct comparability of different approaches. Here we briefly overview the works published on the Special Issue.

Faraki et al.[86]: Each cell image is represented through the covariance descriptor (CovD), where it is assumed that data lie on a known Riemannian manifold of symmetric, positive definite matrices. The underlying principal here is that data is inherently non-Euclidean, and disregarding this geometrical structure can lead to suboptimal results. As compared to normal bag of words (BoW), a bag of Riemannian words (BoRW) is based on non-Euclidean Riemannian manifolds which encode the geometry of the signatures by exploiting the distribution of the signatures. A multi-class SVM is used for classification.

Kastaniotis et al.[95]: combine two descriptors based on gradient and textural characteristics. A SIFT based rotation invariant representation is utilized to encodes local gradients. A new descriptor, Gradient-Oriented Co-occurrence of LBP (GoC-LBP), is proposed to evaluate the orientation of the local gradient in the neighborhood of each pixel. Classification is performed in the dissimilarity space by using an approach based on discriminative sparse representations.

Kazanov et al.[91]: uses morphological features to capture the structural content of the stained regions inside the cell, and is achieved through a multi-threshold binarization technique. Here, two multi-class SVM classifiers are trained separately for the classification of the positive and the intermediate intensity levels.

Nosaka et al.[90]: proposes a new textural descriptor based on traditional LBP, Rotation Invariant Co-Occurrence among adjacent LBP (RIC-LBP). The training set is extended by generating synthetic images, rotating the original ones by various angles, with the purpose to deal with global rotations of the cells. Multi-class SVM remains the choice for the staining pattern classification.

Shen et al.[92]: proposes the integration of a recent descriptor, Multi support Region Order-Based Gradient Histogram (MROGH), within the bag of words framework. The classification is performed through a traditional multi-class SVM.

Snell et al.[93]: The authors compare four different texture-based descriptors: DCT, pixel difference statistics at different scales, morphological features and Gray-Level Co-occurrence Matrix and show how the performance of a cell-level classifier could be improved by analyzing the full distribution of cell parameters within a patient sample. The classification has been performed by employing a multi-class SVM with RBF kernel, whose parameters are determined by a grid search, and a comparison between the different features sets is performed.

Stoklasa et al.[94]: The method combines different sets of features namely Haralick features, Local Binary Patterns, SIFT, surface description and a granulometry-based descriptor. For each descriptor a separate distance is computed and the aggregated distance is computed as a linear combination of the basic descriptor distances. Within a framework of k-NN classification this aggregated distance is used to label the query object.

As compared to hand-coded descriptors Wang et al.[89] [30] proposes a method to learn the descriptors from the image data itself. The feature vectors are locally clustered and for each local region the most representative descriptors are selected by partial least squares (PLS). A BoF model is applied on each region and a codebook is extracted for each of them. The final representation of the cell image is obtained by the concatenation of the codewords obtained from these codebooks. The classification task is performed by using a multi-class SVM.

Wiliem et al.[96]: proposes a new framework for the classification of HEp-2 cell images, namely the Cell Pyramid Matching (CPM), based on the combination of regional histograms of visual words and of the Multiple Kernel Learning (MKL) framework.

Xiangfei et al.[88]: utilizes texture and gradient features based respectively on

response filter banks and the HOG descriptor. The dimensionality of the feature vector is reduced through PCA. Dictionary learning is performed according to a novel technique, and then it is used for the sparse representation of the cell images. Classification is performed by choosing the label of the class whose dictionary atoms minimize the reconstruction error.

Yang et al.[97]: The size of the training set is first increased by Spontaneous Activity Pattern (SAP). Using this augmented dataset image descriptors are learnt using Independent Component Analysis (ICA). The classification is performed by Kernel SVM, with a correlation kernel based on the Pearson correlation.

Finally our approach presented in [85], has two main contributions to the problem. First, an optimal set of features, based on the fusion of morphological and texture descriptors, is experimentally found. Second, the combination of a features selection process and of a discriminant analysis allows to effectively reduce the high dimensionality of the data and at the same time to cope with the high intra-class variance of the samples. These include Minimum-Redundancy Maximum-Relevance methodology (mrmr) algorithm as the feature selection algorithm and Subclass discriminant Analysis, to overcome the limitations of the traditional LDA approach, to encode the variability of the data. The classification is based on traditional (k-NN). Table 3.1 shows the average accuracy on HEp2 cell classification for the different methods, including our proposed approach, published on the Special Issue. A more detail comparision can be found in [87].

Table 3.1: Accuracy comparisons(%) for the different methods on the MIVIA [53] dataset

Method	Experimental Protocol 1		Experimental Protocol 2	
	Cell Level (%)	Image Level (%)	Cell Level (%)	Image Level (%)
Faraki	70	79	72	75
Kastaniotis	**75**	86	65	75
Kazanov	71	**100**	55	71
Nosaka	69	71	71	86
Shen	74	86	69	75
Snell	57	71	54	71
Stoklasa	64	79	64	71
Wang	67	79	59	64
Wiliem	67	71	57	64
Xiangfei	67	89	63	71
Yang	65	79	61	75
Our method	72	93	**90**	**96**

3.2 Image Database

The samples used in our experiments were obtained from the *MIVIA HEp-2*, an annotated database of Indirect Immunofluorescence (IIF) images that is publicly available at [53]. The database was firstly adopted and described in [41] and later used in the *2012 HEp-2 Cells Classification Contest* [37]. It contains 28 images of different patients (one image per patient), which were obtained with the following protocol. Each image was acquired with a fluorescence microscope coupled with a 50W mercury vapour lamp and a digital camera having a CCD with square pixels of size 6.45 μm. The microscope observed, with a 40-fold magnification, a slide of HEp-2 substrate prepared with a fixed dilution of 1:80, as recommended in [56]. The size of the IIF images is 1388x1038 pixels and their color bit depth is 24.

Specialists manually segmented and annotated each cell in the IIF images. All the segmentations were reviewed by medical doctors specialized in immunology, which also provided information related to fluorescence intensity, mitotic phase and staining pattern of the cells. Thus, the database provides for each image its fluorescence intensity (intermediate or positive) and, for each object in an image, (i) its seed point and bounding box, (ii) a binary mask describing its region of interest (ROI), (iii) the object type (cell, mitotic cell or artifact due to the slide's preparation process) and, if the object is a non-mitotic cell, (iv) the label of its staining pattern, *i.e.* the six classes reported in Figure 3.3. They are distinguished as follow:

Homogeneous: diffuse staining of the entire nucleus, with or without apparent masking of the nucleoli.

Nucleolar: fluorescent staining of the nucleoli within the nucleus, sharply separated from the unstained nucleoplasm.

Coarse/Fine Speckled: fluorescent aggregates throughout the nucleus which can be very fine to very coarse depending on the type of antibody present.

Centromere: discrete uniform speckles throughout the nucleus, the number corresponds to a multiple of the normal chromosome number.

Cytoplasmic fluorescence: granular or fibrous fluorescence in the cytoplasm.

The dataset used for training and testing our method contains all and only the normal cells (*i.e.*, neither artifacts nor mitotic cells) of the 28 IIF images included in the *MIVIA HEp-2* database. A full characterization of this dataset is reported in Table 3.2.

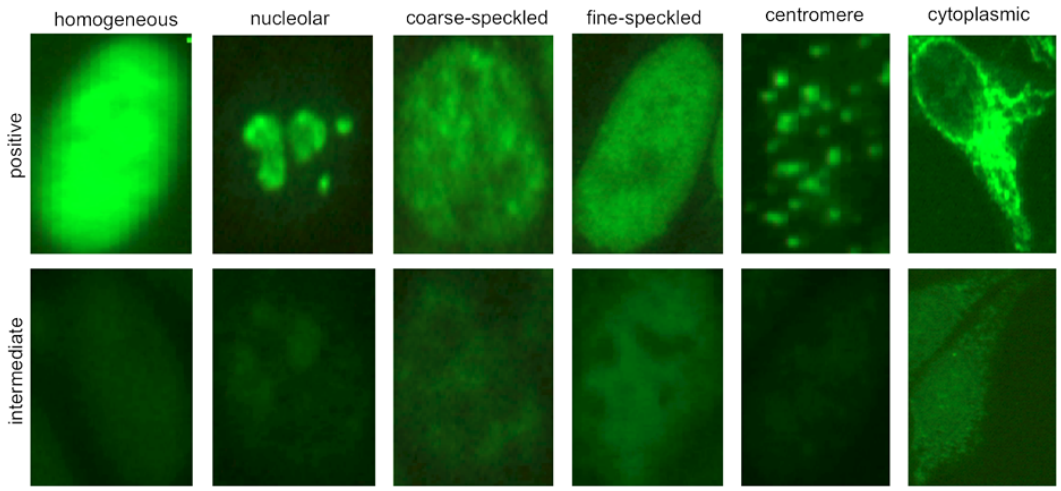


Figure 3.3: six types of staining patterns with positive and intermediate intensity levels.

Table 3.2: HEp-2 cell dataset characterization, grouped by image ID.

Image ID	**Pattern**	**Intensity**	**n. of cells**
1	Homogeneous	positive	61
2	Fine Speckled	intermediate	48
3	Centromere	positive	89
4	Nucleolar	intermediate	66
5	Homogeneous	intermediate	47
6	Coarse Speckled	positive	68
7	Centromere	intermediate	56
8	Nucleolar	positive	56
9	Fine Speckled	positive	46
10	Coarse Speckled	intermediate	33
11	Coarse Speckled	intermediate	41
12	Coarse Speckled	positive	49
13	Centromere	positive	46
14	Centromere	intermediate	63
15	Fine Speckled	intermediate	63
16	Centromere	positive	38
17	Coarse Speckled	positive	19
18	Homogeneous	positive	42
19	Centromere	intermediate	65
20	Nucleolar	intermediate	46
21	Homogeneous	intermediate	61
22	Homogeneous	positive	119
23	Fine Speckled	positive	51
24	Nucleolar	positive	73
25	Cytoplasmatic	intermediate	24
26	Cytoplasmatic	positive	34
27	Cytoplasmatic	intermediate	38
28	Cytoplasmatic	intermediate	13
		Tot.	**1455**

3.3 Proposed Method

In the recent literature several formulations of image descriptors have been applied to the automated classification of HEp-2 staining patterns. These descriptors are derived from either morphological or textural analysis. However, the relevance of these different types of attributes and their mutual contribution are still open issues.

In several works it has been found that the integration of information of different nature provides a substantial improvement of the classification accuracy [51, 52]. Furthermore, the combination of local and global descriptors has been claimed to generate a more robust mechanism of texture representation [57, 58].

Based on these considerations, in our work we performed a detailed analysis of different formulations of morphological, global and local texture characteristics and we investigated, through their integration at different levels, their reciprocal impact on the classification problem.

The outline of the proposed staining pattern classification algorithm is the following. For each cell: (i) we normalise its image and (ii) we extract different feature vectors, one for each considered attribute type. When combining different attributes to characterize a cell image, the corresponding feature vectors are concatenated. Then, for each individual attribute or concatenation of attributes, we select its most relevant feature variables and apply SDA to remap the cell representation into a novel feature space that provides a better class separation. Finally, we train a classifier on a sample set and apply it to a separate test set of samples.

3.3.1 Image Normalisation

Size and intensity normalisation of the samples is a necessary preprocessing step. In fact, small differences in the dimensions of the cell images are normal, and these differences are completely uncorrelated with their staining pattern. On the other hand, there are considerable variations of fluorescence intensity between intermediate and positive samples, and, in a smaller scale, even within cells of the same image (see Figure 3.1). Reducing such variability helps to decrease the noise in the classification process and avoids as well the necessity of training two separate classifiers for intermediate and positive samples.

In our work, all the cell images were resized to 128x128 pixels. As for the intensity, since the images are produced by green fluorescence, the samples were first converted to grayscale by taking into account only the green channel. Then, intensity normalisation was obtained by linearly remapping all the intensities so that

the bottom 1% and the top 1% of all pixel values are saturated at, respectively, the lowest and highest intensities.

3.3.2 Feature Extraction

In the following, we will use interchangeably the terms "feature set" and "attribute" to refer to the elements used to characterize a cell image according to a specific technique.

As already explained, we took into consideration different image attributes belonging to three main categories:

1. *morphological features*, focusing on shape attributes of the fluorescent signal of the cells or of specific regions within the cells;

2. *global texture descriptors*, which summarize the overall appearance or general distribution of the gray-levels in the cell image;

3. *local texture descriptors*, which summarize the isolated contribution of small regions or pixel neighbourhoods.

Table 3.3: Image attributes considered in this work.

Image Attributes			**size**
Morphological Features		MORPH	43
Global Texture Features	Gray-Level Co-occurrence Matrices [59, 60, 61]	GLCM	44
	Edge Orientation Histograms [62]	EOH	80
	Rotation-Invariant Gabor features [63]	RIGF	32
	Modified Zernike moments [64]	ZERN	55
Local Texture Features	Rotation-Invariant Uniform Local Binary Patterns [65]	LBP^{riu2}	42
	Completed Local Binary Patterns [66]	CLBP	42
	Co-occurrence of adjacent LBPs [67]	CoALBP	3072
	Rotation-invariant Co-occurrence of adjacent LBPs [68]	RIC-LBP	408

For each of these categories, several approaches have been described in the literature. However, performing an exhaustive analysis of all of them is extremely difficult and out of the scope of this work. Here we focused on the ones that, based on preliminary experiments that we do not report for the sake of brevity, appear to be the most promising for HEp-2 cell characterization. The list of these attributes can be seen in Table 3.3.

Morphological Features (MORPH)

Several morphological features have been used to characterize the cell images.

The first feature derives from the following consideration. As it is shown in Figure 3.2, for all the staining patterns except the cytoplasmatic, the fluorescence signal is localized into the nucleus of the cell. Therefore, in these cases, the cell ROI is round or elliptical. On the contrary, the shape of cytoplasmatic patterns is generally more irregular. These differences can be summarized by the ROI circularity, defined as $Area/Perimeter^2$.

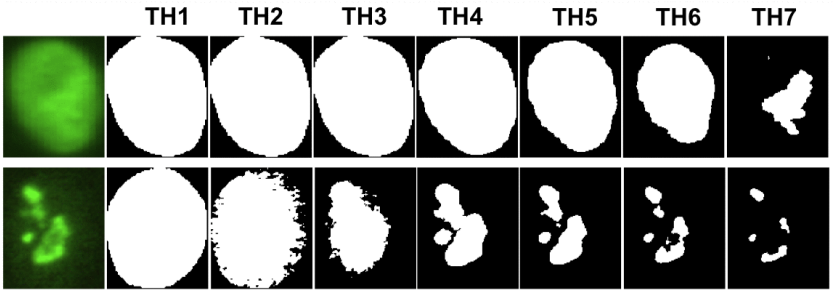


Figure 3.4: Examples of binary masks obtained from seven progressive thresholding levels (TH1-TH7) applied to images of cells with homogeneous (top) and nucleolar (bottom) staining patterns.

Other morphological measurements are extracted from the binarization of the cell image at increasing intensity thresholds. For each of these thresholds a different binary mask is obtained, which isolates the structuring elements of the cell pattern at that specific intensity level. As it can be seen in Figure 3.4, the shape changes of these masks show a very characteristic behaviour for distinct staining patterns. Therefore, the concatenation of morphological parameters measured at the increasing thresholding levels appears to be effective in characterizing the HEp-2 cell patterns [51].

As for the binarization, we used seven progressive thresholds. A first reference value (*TH4*, in Figure 3.5) is obtained applying the Otsu's method [69] to the intensity histogram of the cell. The two intervals of the intensity histogram defined by *TH4* are further divided intro three equal segments to obtain the other six thresholds, namely *TH1-3* and *TH5-7* for, respectively, the lower and upper parts of the intensity distribution (see Figure 3.5).

For each binary mask, we then extracted the following morphological properties:

1. proportion of the ROI area covered by foreground pixels;

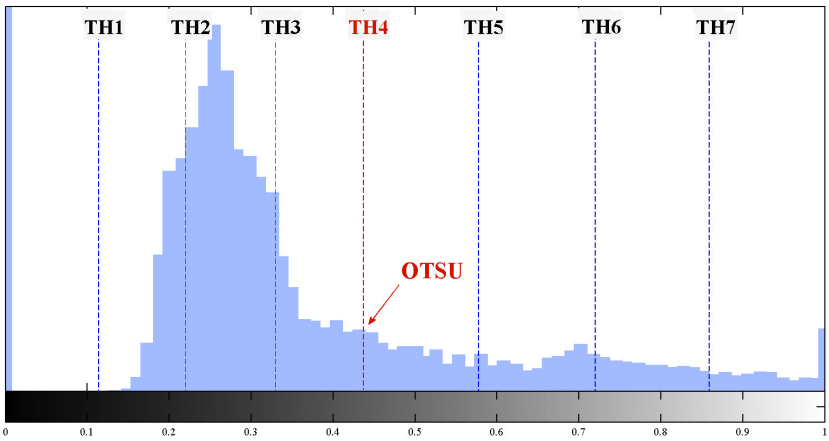


Figure 3.5: The seven progressive thresholds computed from a cell intensity histogram.

2. number of connected regions of the foreground;

3. proportion of the connected regions of the foreground which have circular shape (*i.e.* with $circularity >= 0.8$);

4. average area of the foreground's connected regions divided by the total area of the cell ROI;

5. average circularity of the connected regions of the foreground;

6. average solidity of the foreground, where the solidity of each connected region is computed as the percentage of the pixels in the convex hull that also belong to that region.

All the measurements (*i.e.* the circularity of the ROI, and the six properties extracted at the seven thresholding levels) are finally concatenated, obtaining a descriptor of size 43.

Some of the aforementioned properties can be more discriminative for some classes and less for others. For example, features 1 and 4 provide the same values when the foreground contains only one connected region (*e.g.* the homogeneous pattern in the first row of Figure 3.4) and very different ones in other cases (*e.g.* the nucleolar pattern in the second row of Figure 3.4). However, as we will show in the following, the application of a Feature Selection process helps to discard the less relevant variables.

Global Textural Features (GTF)

Gray-Level Co-occurrence Matrices (GLCM) Several types of global texture descriptors can be computed from Gray-Level Co-occurrence Matrices. These matrices report the distribution of co-occurring values between neighbouring pixels according to different distances and directions [59]. More specifically, each GLCM element $(i, j)_{d,\theta}$ contains the probability for a pair of pixels located at a distance d and direction θ to have gray levels i and j, respectively. In our work, we first computed four GLCMs for a fixed unitarian d and a varying $\theta = 0^o$,45^o,90^o,135^o after grouping intensity values into 16 levels. The use of 4 different directions makes the method less sensitive to image rotations. Then, we extracted from each GLCM a total of 22 statistical measures, whose full characterization can be found in [59, 60, 61]. Finally, we computed the mean M and the range value R (*i.e.* maximum-minimum) over the 4 GLCMs for each of these measures, obtaining a total of 44 GLCM features.

Modified Zernike moments (ZERN) Zernike radial polynomials can serve as basis functions to extract image moments, capable of describing the shape and textural characteristics of an object. However, they are strongly affected by changes in scale and position of the objects in the image. An effective descriptor overcoming this limitations and invariant to rotation, translation and scaling, has been presented in [64]. Its invariance properties are obtained computing modified Zernike moments on the normalised power spectrum of the image. In our case, for each cell we extracted a descriptor of 55 elements.

Edge Orientation Histograms (EOH) The general idea of Edge Orientation Histograms (EOH, [62]), as the name suggests, is to represent an image by a histogram obtained from the predominant gradient orientations of its edge pixels. More specifically, EOH applies a Canny operator to obtain contour pixels from the image, and then it constructs a histogram of directions, counting for each bin the number of contour pixels whose gradient falls within its specific orientation interval. In our work, we computed a 80-dimensional EOH feature vector for each cell image.

Rotation-Invariant Gabor features (RIGF) Gabor filters are widely used for edge detection and texture analysis. Although not rotation invariant, Gabor features can be modified to gain this property by computing their Discrete Fourier Transform (DFT). In this way, the circular shift caused in the Gabor feature vector by an image rotation is converted into a phase shift in the Fourier domain. This phase shift does not affect the magnitude of the DFT coefficients which, therefore, can be used as a rotation-independent representation of textural data. In our work we used the formulation of RIGF proposed in [63], obtaining for each cell a feature vector of dimension 32.

Local Textural Features (LTF)

In the context of local textural features, one of the most used descriptors is based on local binary patterns (LBPs) [70]. This approach has been found to be very promising also for the characterization of HEp-2 fluorescence patterns [51, 68].

The basic idea behind this descriptor is to represent the texture of the image as a histogram of LBPs, *i.e.* binary patterns representing the intensity relations between a pixel and its neighbours. For each image pixel, an LBP is obtained by binarizing its neighbouring region using the intensity of the pixel as threshold, and then by converting the resulting binary pattern to a decimal number. Finally, a histogram is generated by taking into account the occurrences of all the LBPs in the image. This is a very simple yet powerful textural descriptor, whose main advantage is the invariance to changes of illumination over the image.

Recent literature reports different descriptors that are supposed to extend and improve the descriptive capabilities of classical LBPs. In our work, we considered the following major formulations.

Rotation-Invariant Uniform Local Binary Patterns ($\mathbf{LBP^{riu2}}$) A binary pattern is called uniform if it contains not more than two bitwise transitions from 0 to 1 or vice versa when the bit pattern is transversed circularly. Uniform patterns were found to be predominant with respect to other patterns for texture description. Therefore, in [65] they are used to produce compact rotation-invariant LBP feature vectors. In our work we concatenated three histograms with different scales of LBP neighbourhood, as suggested in [65], obtaining a feature vector of dimension 42.

Completed Local Binary Patterns (CLBP) The classical formulation of LBP uses the intensity of the center pixel as threshold for its neighbours. Hence, only the sign of the difference between the center and the neighbour gray values is relevant. The key idea of Completed Local Binary Patterns (CLBP [66]) is to represent each neighbourhood by its center pixel and a local difference sign-magnitude transform (LDSMT) that computes both the sign and the magnitude of the difference between the central pixel and its neighbours. Using the same scale and neighbourhood of LBP^{riu2}, we obtained a 42-dimensional CLBP feature vector.

Co-occurrence of Adjacent LBPs (CoALBP) The original expression of LBPs lacks structural information among different binary patterns. In order to provide them, [67] introduced a novel formulation which measures the co-occurrence among multiple LBPs (and in particular, among adjacent LBPs). This formulation provides a high-dimensional feature vector that has been found to provide a better texture characterization compared to previous LBPs. Concatenating three CoALBP histograms with the parameters suggested in [68], the resulting feature vector has

size 3072.

Rotation-Invariant Co-occurrence of Adjacent LBPs (RIC-LBP) The CoALBP features can vary significantly depending on the orientation of the target object. In order to cope with this problem, [68] extends the concept of rotation equivalence class of LBP^{riu2} to the CoALBPs. This is achieved by attaching a rotation invariant label to each LBP pair, so that all CoALBPs corresponding to different rotations of the same LBPs have the same value. For each cell image, we extracted and concatenated three RIC-LBP histograms as suggested by [68], obtaining a feature vector of dimension 408.

3.3.3 Classification based on Subclass Discriminant Analysis

The categorization of the cell staining patterns is a multiclass classification problem. Our classifier is based on Subclass Discriminant Analysis (SDA), an algorithm that has been recently proposed in [54].

SDA belongs to the family of Discriminant analysis (DA) algorithms, which have been used for dimensionality reduction and feature extraction in many applications. Given a set of samples $\boldsymbol{X} = (x_1, x_2, \ldots, x_n)$, where each x_i is a vector in $\Re^D$ and has an associated class label $\in [1, C]$, DA algorithms compute a mapping of $\boldsymbol{X}$ to a subspace in $\Re^d$, with $d \ll D$, where the data can be more easily separated according to their class-labels. This mapping is defined by a projection matrix $V = (v_1, v_2, \ldots, v_d)$, with $v_i \in \Re^D$, which, in most DA algorithms, is found by maximizing the so-called Fisher-Rao's criterion:

$$J(V) = \frac{\left|V^T A V\right|}{|V^T B V|} \tag{3.1}$$

where A and B are symmetric and positive-defined matrices, so that they define a metric. The solution to this problem is given by the generalized eigenvalue decomposition:

$$AV = BV\Lambda \tag{3.2}$$

where Λ is the diagonal matrix of eigenvalues corresponding to the column-vectors of V.

Linear Discriminant Analysis (LDA) is probably the most well-known DA algorithm. It assumes that the C classes the data belong to are homoscedastic, *i.e.* that their underlying distributions are Gaussian with common variance and different

means. In equation (3.1), LDA uses $A = S_B$, the *between-class* scatter matrix, and $B = S_W$, the within-class scatter matrix, which are defined as follows:

$$S_B = \sum_{i=1}^{C}(\mu_i - \mu)(\mu_i - \mu)^T \tag{3.3}$$

$$S_W = \frac{1}{n}\sum_{i=1}^{C}\sum_{j=1}^{n_i}(x_{ij} - \mu_i)(x_{ij} - \mu_i)^T \tag{3.4}$$

where n is the number of the samples in $\boldsymbol{X}$, n_i the number of samples in class i, μ_i is the sample mean for class i, μ is the global mean and x_{ij} is the j^{th} sample of class i.

The mapping provided by LDA maximizes the between-class variance and minimizes the within-class variance in any particular data set. In other words, it guarantees maximal class separability of the training samples and, possibly, optimizes the accuracy in later classification. The main drawback of LDA is that the dimension of the reduced space is bounded by the rank of S_B, which is in turn equal or lower than $C - 1$. This means that each sample in the original space is represented by at most $C - 1$ features. As a consequence, LDA works well for problems that are linear in the origin space, *i.e.* those for which $C - 1$ features are sufficient to discriminate C classes, but fails to provide optimal subspaces for inherently non-linear structures in data.

To cope with this problem, several extensions of LDA have been introduced in literature [71] and SDA is one of the most effective. The main idea of SDA is to approximate the underlying distribution of each class using a mixture of Gaussians. This is achieved by dividing the classes into a set of subclasses, capable of describing for each class the variance of the data in a more subtle way. Therefore, the main problem to be solved in SDA is to find the optimal number of subclasses for each class.

Once the number of subclasses is known, the transformation matrix V is found by assigning to A in equation (3.1) the *between-subclass* scatter matrix defined as:

$$S_B = \sum_{i=1}^{C-1}\sum_{j=1}^{H_i}\sum_{k=i+1}^{C}\sum_{l=1}^{H_k} p_{ij}p_{kl}(\mu_{ij} - \mu_{kl})(\mu_{ij} - \mu_{kl})^T \tag{3.5}$$

where H_i is the number of subclasses of class i, μ_{ij} and p_{ij} are the mean and prior probability of the j^{th} subclass of class i, respectively. The priors are estimated as $p_{ij} = n_{ij}/n$, where n_{ij} is the number of samples in the j^{th} subclass of class i. With this formulation of S_B, equation (3.1) works towards the contemporary

maximization of the distance between the class means and of the one between the means of subclasses of the same class. The objective is, again, to maximize the classification accuracy in the reduced space.

In order to select the optimal number of subclasses, two different methods are proposed in [54] . The first is based on a stability criterion described in [72]. However, as pointed out in [73], the minimization of the metric used in this criterion is not guaranteed when data have a Gaussian homoscedastic subclass structure (and probably even for heteroscedastic). The second selection criterion is based on a *leave-one-object* test. In practice, a leave-one-out cross validation is applied for each subdivision, and the optimal subdivision is the one maximizing the recognition rate. The problem with this strategy is its very high computational costs, especially with large datasets and high number of classes as in our case. Therefore, in our work we use a slightly different formulation of this criterion, which is based on a stratified 5-fold cross validation of the training set and on the accuracies obtained with a k-Nearest Neighbour (k-NN) classifier.

Our implementation differs from the original SDA formulation for two other details. The first concerns the clustering method. In [54] data are assigned to subclasses by first sorting the class samples with a Nearest-Neighbour based algorithm and then by dividing the obtained list into a set of equal-sized clusters. However, this method does not allow to model effectively the non-linearity present in the data, as in the case under analysis of HEp-2 staining patterns. Therefore, we used the K-means algorithm, which partitions the samples into K clusters by minimizing iteratively the sum, over all clusters, of the within-cluster sums of sample-to-cluster-centroid distances.

The second difference is that, instead of increasing at each iteration the number of subclasses of all groups of the same amount, we first choose a maximal number of subclasses for each class, based heuristically on the number of samples per class. Then, all the possible permutations of class subdivisions are tested, halting the subdivision process of a specific class r if the minimal number of samples in any of its current H_r clusters drops below a predefined threshold. In order to reduce the computational burden, the clusters created for each class and each number of subclasses are computed only once and cached for further use.

Once the optimal number of subclasses have been found and the feature vectors of the samples have been obtained by projecting them on the sub-space defined by SDA, the classification is performed with k-NN. A value of 8 for k has been heuristically found to provide good classification results.

3.3.4 Feature Selection

Feature selection (FS) is a data preprocessing step that is frequently applied in machine learning. FS extracts the subset of features used to describe the data that improves the classification accuracy. The main purposes of FS are two. First, it reduces the dimensionality of the input data, removing irrelevant information and improving their comprehension by telling which are the most important features and how they are correlated. Second, it improves the chances of avoiding overfitting, whose probability increases with the dimension of the feature space.

In our work, a reduction of the dimensionality of the feature space is already provided by SDA. However, decreasing the initial number of parameters representing the data in order to include only the features that are the most relevant for the considered task improves the separability of the samples in the reduced space computed by SDA and, hence, the classification accuracy (as we will show in Section 3.5).

We applied a robust feature selection method based on the minimum-Redundancy-Maximum-Relevance (mRMR) algorithm, whose better performance over the conventional top-ranking methods has been widely demonstrated in literature [74]. mRMR sorts the features that are most relevant for the characterization of the classification variable, pointing at the contemporaneous minimization of their mutual similarity and maximization of their correlation with the classification label.

However, mRMR provides a mere ranking of the features with no information on the size of the optimal feature set. Therefore, heuristically, we looked for this optimal size by iteratively increasing the number of candidates in the feature set until the global optimum is found. The candidates are selected in order of relevance and, given the computational burden of the classification algorithm, at each iteration their number is increased by a factor 10. The initial number of candidates is chosen according to the following consideration. The rank of matrix S_B and, therefore, the dimensionality d of the reduced subspace obtained with SDA, is given by $min(H-1, rank(S_X))$, where H is the total number of subclasses, S_X is the matrix concatenating the representative vectors of the samples in the training set and $rank(S_X)$ is equal (or minor) to the number of features characterizing each sample. Since the data in our problem present high non-linearities, we set as reasonable lower bound for d the value 40.

3.4 Experimental Protocol

All experiments shown in this work were run on the HEp-2 cell dataset described in Section 3.2 using the following protocol.

The first experiment applies the *leave-one-out* technique over all the 28 IIF images composing the dataset. For each image, an instance of the classifier is trained with the cells of the remaining 27 images and then used to classify the cells of the image left out. The classification evaluates each cell individually, without using information on the other cells belonging to the same image. The classification accuracy is evaluated (i) on a *cell-level* basis, in terms of percentage of cells correctly classified for each class over all the 28 leave-one-out runs, and (ii) on an *image-level* basis, in terms of percentage of images correctly classified, using as the predicted image label the class most frequently assigned to its cells. In both cases, the accuracy is obtained by adding the classification counts of the 28 leave-one-out runs (i.e. the number of cells of class i assigned to class j are evaluated by summing the corresponding numbers in each run, and then percentages are computed), and not by averaging the percentages relative to each run.

The *leave-one-out* method over the images ensures a complete separability and independence of training and test set and it is the experimental protocol that maximizes the number of independent samples (i.e. cells belonging to different patients) used for training. Therefore, it is particularly suited to the HEp-2 cell dataset, that is characterized by a very low number of independent samples per class. Moreover, the execution of several training/test runs with different images allows to assess the robustness and generalization capabilities of the classifier.

A second additional experiment is performed by dividing the images into two *separate training and test sets* in exactly the same way as it was done in the *2012 HEp-2 Cells Classification Contest* [37]. The classifier is trained with the cells included in the training set, which contains only 14 of the images of the entire dataset, and the performance of the classifier is assessed on the test set (*i.e.* the cells belonging to the remaining 14 images). As for the previous experiment, cell-level and image-level accuracies are computed.

3.5 Results and Discussion

Our experiments were aimed at answering the following research questions:

- how do the different attributes described in Section 3.3.2 contribute to the classification of the HEp-2 staining patterns and how can these individual attributes be combined to improve the classification accuracy? (Section 3.5.1)
- how does SDA strategy cope with the high within-class variance that is typical of HEp-2 staining patterns? (Section 3.5.2)

3.5.1 Assessing the classification accuracy

For clarity, we briefly recall the main idea behind our feature extraction approach. Morphological, global and local textural descriptors have been found to be the most suitable features for characterizing HEp-2 fluorescence patterns. However, the relevance of each of this kind of attributes and their mutual contribution is still an open issue. In this work we postulate that the integration of information of different nature improves the capabilities of a classifier of recognizing different staining patterns. To this end, we defined a set of morphological features and we identified different textural descriptors that appeared promising for the task considered.

In the first tests, we computed the classification accuracy (on a *cell-level* basis) obtained with each individual feature set described in Section 3.3.2, using the *leave-one-out* experiment described in Section 3.4. Then we analysed the accuracies obtained with different groups of attributes.

The experimental results are summarized in Figure 3.6, where we report for different feature sets the classification accuracy and (within parentheses) the number of optimal features selected by the FS algorithm.

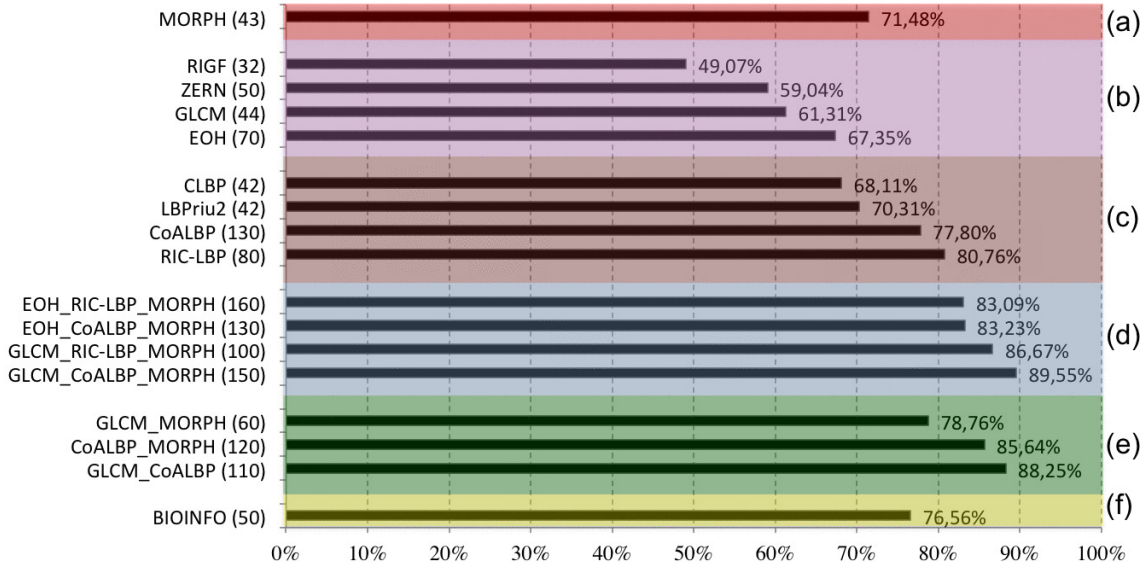


Figure 3.6: Cell-level accuracy obtained by different sets of features (*leave-one-out experiment*). The sections of the graph describe the results of: a) morphological features, b) global textural features, c) local textural features, d) best combinations of the three attribute categories, e) sub-groups of attributes of the best solution, f) our previous approach [55].

Figure 3.6 is divided into sections showing the accuracies of the individual attributes (a-c) and of different combinations of them (d-e). For the sake of brevity, we report only those combinations, between all the ones tested, providing the highest accuracies. For comparison, we also report, in section (f), the accuracy obtained by our previous approach to the problem [55] (feature set named BIOINFO). It should be noted that FS was not applied to attributes whose size was close to, or lower than, 40, which is the limit value for our FS process (Section 3.3.4).

The following initial remarks can be drawn from the results of Figure 3.6:

- the single attribute providing the highest accuracy was RIC-LBP (80.76%), followed by CoALBP (77.80%) and MORPH (71.48%);
- considering the categories of individual attributes, local textural features (LTF) performed much better than morphological features, which were in turn better than global textural features (GTF);
- in general, groups of feature sets consistently achieved better results than their single components;
- the integration of morphological and both local and global textural features provided the most comprehensive solution to characterize fluorescence patterns. That is, the more heterogeneous the information, the better the accuracies. The best result (89.55%) was obtained by the combination of GLCM, CoALBP and MORPH. The accuracy obtained by this optimal group was higher than any partial combination of its composing attributes (shown in figure 3.6(e));
- the proposed method outperformed our previous approach to the problem, which obtained a 76.56% accuracy.

These results require a more detailed discussion. First, while LTF are effective in characterizing the cells, as also found in [37], GTF do not perform equally well. In other words, GTF are less capable of capturing the differences between the various staining patterns, differences that, on the contrary, are more suitably described in terms of local variations of the intensity patterns. The relevance of local information is supported by the results obtained by morphological features, which, in a way, encode local characteristics into the description of the morphological properties of the cells and of the changes of these properties as a function of the intensity thresholds used to compute them.

Second, the combination of global and local texture descriptors consistently generates a more robust mechanism of texture representation (for example, this is shown

in Figure 3.6 by comparing the accuracy of GLCM-CoALBP, 88.25%, with the accuracy of the single attributes, 61.31% and 77.80% respectively). This result is consistent with the findings of previous works [57, 58].

Third, the combination of different attributes generates a completely new mechanism of image representation compared to the one provided by the same attributes considered individually. Thus, the feature sets best performing individually are not necessarily the ones that will give the best performance when combined together. Indeed, the best GTF and LTF attributes (respectively, EOH with 67.35% and RIC-LBP with 80.76%) are not included into the group that obtained the highest accuracy and, for instance, much better accuracies are obtained with groups based on GLCM as GTF and on CoALBP as LTF. This could be explained as follows:

- the integration with other features helps to soften the major limitations of specific attributes (*e.g.*, the dependence from texture rotations of CoALBP);
- Feature Selection contributes to this scenario by selecting a compact set of features that helps to reduce the noise and by strengthening the mutual contribution of specific attributes.

Feature Selection deserves another detailed discussion. As expected, FS always significantly improves the accuracy. As an example, in Figure 3.7 we show, for the group GLCM-RIC-LBP-MORPH, the behaviour of the accuracy against the number of features used by the classifier. As it can be seen, the accuracy reaches a global maximum for 100 features improving of more than 5% the accuracy obtained with all the 495 feature variables. Another interesting result is that the total accuracy around the optimal set size is rather stable (*i.e.* less than 2% in the range between 60 and 160), indicating that the performance of the proposed approach is not strictly tied to a precise identification of the optimal size. For other attributes and attribute groups, including the optimal GLCM-CoALBP-MORPH group, we obtained similar results that we omit for brevity. We simply underline the fact that with very large feature sets (*e.g.* those including CoALBP), the performance drop between the optimal set and the full feature set is significantly higher than the one shown in the example of Figure 3.7. As a matter of facts, in these cases, since the classifiers are based on a number of features largely greater than the number of observations, results with no FS are likely to be affected by overfitting. We believe that these results highlight as well the capabilities of FS to prevent this problem.

The analysis of the features surviving the FS pruning can provide some insights into the cell characteristics that are more relevant for the classification of the staining patterns. In general, in all groups all types of features are present, with a slight preference for local texture descriptors. If we consider, as an example, the optimal

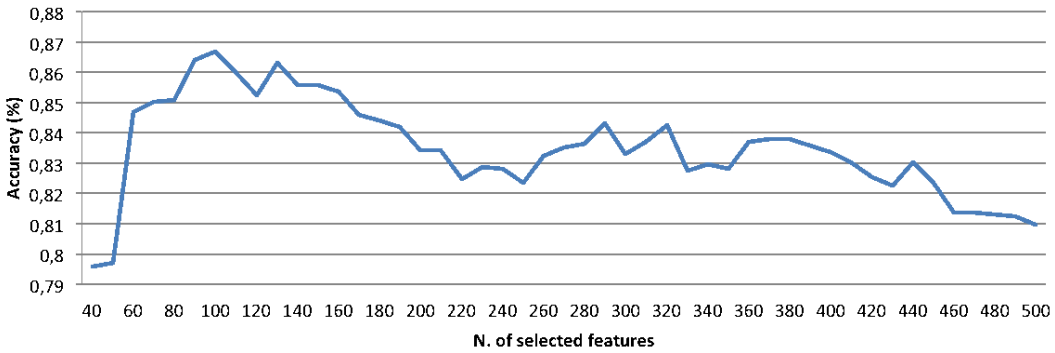


Figure 3.7: Cell classification accuracy vs. size of the feature set (features set GLCM-RIC-LBP-MORPH, *leave-one-out experiment*).

combination GLCM-CoALBP-MORPH, we have that, of the 150 features selected, 16 come from GLCM (out of a total of 44), 106 from CoALBP (3072) and the remaining 28 from MORPH (43). This is another evidence of the fact that features with different nature are effectively combined to provide a substantial improvement of the accuracy.

Concluding this section, we present the detailed results of the optimal solution found for this first experiment (GLCM-CoALBP-MORPH). In Table 3.4 we show for each of the 28 IIF images its true class and the number and the percentage of the image cells labelled with each of the six staining patterns. Table 3.5 shows on the left the confusion matrix and the total accuracy of cell classification (89.55%) and, on the right, the image-level confusion matrix and the total accuracy of image classification (96.43%).

3.5.2 Discriminative capabilities of SDA

Another interesting subject of analysis is the discriminative capability of our SDA-based classification approach. For the discussion of this issue, we will also make use of the results obtained with the second experiment described in Section 3.4, which is based on two separate training and test sets. For the sake of conciseness, we do not report the detailed results of this experiment, which basically confirmed the main findings of the *leave-one-out experiment*:

- grouped attributes perform better than single attributes;
- the combination of morphological, global and local textural information guarantees the highest accuracies;

Table 3.4: *Leave-one-out experiment*: number and percentage of cells assigned to each class. The values corresponding to the true class are highlighted in dark gray.

IMAGE		Cent.		Hom.		Nucl.		C. Sp.		F. Sp.		Cyt.	
ID	**True class**	n.	%	n.	%	n.	%	n.	%	n.	%	n.	%
1	Hom.	0	0	61	100	0	0	0	0	0	0	0	0
2	F. Sp.	1	2.08	4	8.33	0	0	1	2.08	42	87.5	0	0
3	Cent.	89	100	0	0	0	0	0	0	0	0	0	0
4	Nucl.	1	1.51	0	0	57	86.36	8	12.12	0	0	0	0
5	Hom.	0	0	46	97.87	0	0	0	0	1	2.13	0	0
6	C. Sp.	2	2.94	1	1.47	0	0	65	95.59	0	0	0	0
7	Cent.	49	87.5	0	0	5	8.93	2	3.57	0	0	0	0
8	Nucl.	1	1.79	0	0	55	98.21	0	0	0	0	0	0
9	F. Sp.	0	0	5	10.87	0	0	1	2.17	40	86.96	0	0
10	C. Sp.	3	9.09	0	0	0	0	29	87.88	1	3.03	0	0
11	C. Sp.	0	0	0	0	2	4.88	39	95.12	0	0	0	0
12	C. Sp.	0	0	0	0	0	0	49	100	0	0	0	0
13	Cent.	46	100	0	0	0	0	0	0	0	0	0	0
14	Cent.	14	22.22	20	31.75	0	0	9	14.29	20	31.75	0	0
15	F. Sp.	2	3.17	15	23.81	0	0	7	11.11	39	61.90	0	0
16	Cent.	37	97.37	0	0	1	2.63	0	0	0	0	0	0
17	C. Sp.	0	0	0	0	1	5.26	14	73.68	4	21.05	0	0
18	Hom.	0	0	41	97.62	0	0	1	2.38	0	0	0	0
19	Cent.	63	96.92	0	0	2	3.08	0	0	0	0	0	0
20	Nucl.	0	0	0	0	45	97.83	0	0	0	0	1	2.17
21	Hom.	3	4.92	44	72.13	0	0	2	3.28	12	19.67	0	0
22	Hom.	0	0	115	96.64	2	1.68	0	0	1	0.84	1	0.84
23	F. Sp.	0	0	2	3.92	0	0	0	0	49	96.08	0	0
24	Nucl.	1	1.37	3	4.11	69	94.52	0	0	0	0	0	0
25	Cyt.	0	0	0	0	0	0	2	8.33	0	0	22	91.67
26	Cyt.	0	0	0	0	0	0	0	0	0	0	34	100
27	Cyt.	0	0	0	0	0	0	1	2.63	0	0	37	97.37
28	Cyt.	0	0	0	0	0	0	0	0	0	0	13	100

Table 3.5: *Leave-one-out experiment*: confusion matrices (%).

CELL CLASSIFICATION

class	Cent.	Hom.	Nucl.	C. Sp.	F. Sp.	Cyt.
Cent.	83.47	5.60	2.24	3.08	5.60	0
Hom.	0.91	93.03	0.61	0.91	4.24	0.30
Nucl.	1.25	1.24	93.78	3.32	0	0.41
C. Sp.	2.38	0.48	1.43	93.33	2.38	0
F. Sp.	1.44	12.5	0	4.33	81.73	0
Cyt.	0	0	0	2.75	0	97.25

Total accuracy = 89.55%

IMAGE CLASSIFICATION

Cent.	Hom.	Nucl.	C. Sp.	F. Sp.	Cyt.
83.33	16.67	0	0	0	0
0	100	0	0	0	0
0	0	100	0	0	0
0	0	0	100	0	0
0	0	0	0	100	0
0	0	0	0	0	100

Total accuracy = 96.43%

- the optimal solution is provided, again, by the GLCM-CoALBP-MORPH group (with 100 features), although in this experiment other groups perform in a comparable way at both cell and image-level.
- the performance around the optimal set size is, again, rather stable, indicating that the accuracy is not strictly tied to a precise identification of the optimal number of features.

For completeness, we detail the best result achieved in Table 3.6, where as before we show the confusion matrix and classification accuracy at cell-level (on the left) and at image-level (on the right). As it can be seen, the cell-level accuracy drops

drastically from the 89.55% of the first experiment to the current 72.21%, while the accuracy at image level is similar (92.86% vs. 96.43% of the *leave-one-out experiment*, both corresponding to a single misclassified image).

This difference can be easily explained in terms of the characteristics of the training and test sets. In the second experiment, the availability of only 14 different images for training, coupled with the high within-class variance present in the HEp-2 cell patterns, translates into very few independent samples per class. As a consequence, the manifolds of the various staining patterns in the feature space are likely to be under-sampled, thus hampering the SDA capabilities to model them in a suitable manner and, consequently, reducing the cell-level classification accuracy of the whole method. The better performance at image-level can be simply explained analysing in details the results of the *leave-one-out experiment* (Table 3.4), where the per class correct accuracies are most of the times greater than 85%. Therefore, even if more cells are misclassified in the second experiment, the ones maintaining the correct label are still sufficient, in most of the cases, to assign the correct image label.

Table 3.6: *Separate training/test sets experiment*: confusion matrices (%).

	CELL CLASSIFICATION						**IMAGE CLASSIFICATION**					
class	Cent.	Hom.	Nucl.	C. Sp.	F. Sp.	Cyt.	Cent.	Hom.	Nucl.	C. Sp.	F. Sp.	Cyt.
Cent.	90.60	0	5.37	4.03	0	0	100	0	0	0	0	0
Hom.	7.78	74.44	1.11	0.56	16.11	0	0	100	0	0	0	0
Nucl.	11.51	2.88	65.47	20.14	0	0	0	0	100	0	0	0
C. Sp.	32.67	1.98	0	53.47	10.89	0.99	0	0	0	66.67	33.33	0
F. Sp.	13.16	21.93	2.63	0.88	61.40	0	0	0	0	0	100	0
Cyt.	5.88	0	0	3.92	0	90.20	0	0	0	0	0	100
	Total accuracy = 72.21%						**Total accuracy = 92.86%**					

The protocol of the second experiment let us also inspect in more details the strategy used by SDA in dividing class samples into subclasses. In fact, differently from the *leave-one-out* method, where the optimal class subdivision is likely to vary at each fold, the use of a single training set translates into a single optimal partition of subclasses. If we consider as an example the best solution shown in Table 3.6, the number of subclasses is two for nucleolar and centromere classes, three for cytoplasmatic class, four for coarse and fine speckled and five for homogeneous class.

For nucleolar and centromere classes, SDA performs a straightforward subdivision between cells with intermediate and positive fluorescence intensity. For other classes (*e.g.* the homogeneous pattern) the subdivision splits the cells in a more subtle way, grouping cells of similar appearance in the same subclass (Figure 3.8), which is obviously of major help for the classification process.

Finally, in order to further assess the discriminative capabilities of our SDA-based approach, we compared our method with Support Vector Machines (SVM), a

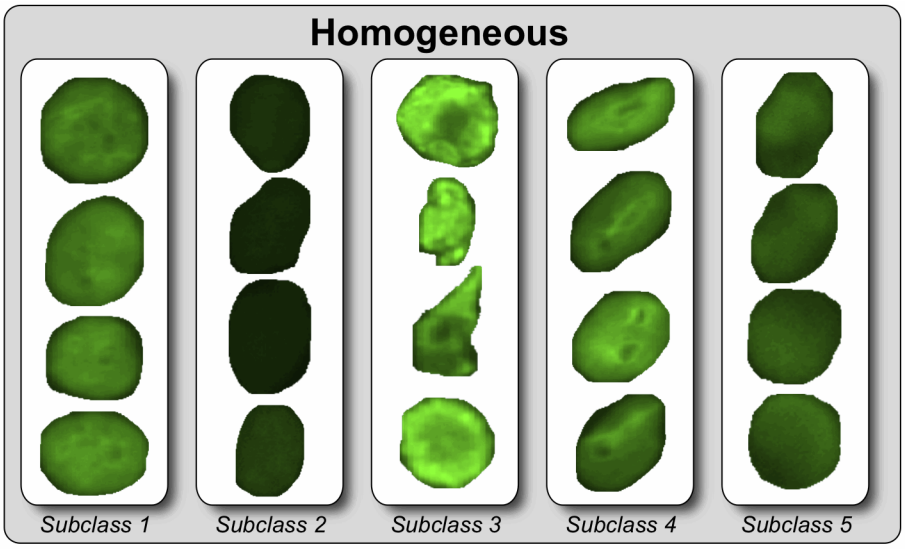


Figure 3.8: Sample images from the five homogeneous subclasses obtained in the optimal solution shown in Table 3.6.

state-of-the-art classifier that is widely recognized for its classification performance by computer scientists and machine learning researchers. For the classification, we used SVM with a radial basis kernel, optimizing its parameters by means of a grid search as suggested in [20]. Then, we applied the two-step FS process described in [49]: we first chose for each single/composite attribute a candidate feature set of size 200 with mRMR, or directly the whole feature set if its size was lower than 200, and then we reduced it to an optimal size applying a Sequential Forward Selection scheme (more details in [49]).

We repeated with SVM all the tests in the two experiments obtaining, on average, classification accuracies lower than SDA-based approach by about 10% in the *leave-one-out* experiment and about 3% in the *separate training-test set* experiment.

Concluding, we think that all these results further support our hypothesis that SDA, being particularly suited to deal with the high within-class heterogeneity of HEp-2 cell images, is a promising approach for staining pattern classification.

3.6 Segmentation of HEp2 Cells

The recognition of the HEP-2 cells in IIF images can be carried out in two ways:

- A cell level classification, where individual cells are classified to belong to one

of the HEp2 patterns. Here it is assumed that the masks for the individual cells in the IIF image are available. A pattern class label is assigned to an IIF image, based on majority voting of the individual cells in the image.

- Specimen level classification, where no information on the cell-segmentation is available, rather a segmentation mask containing the entire object present into the specimen, which not only include the cells in the interphase but also contains cells in mitosis and with possible artifacts (air bubbles, etc) .

Our study of the problem aims at classifying the HEp2 patterns both at cell and specimen level. We opt for a fully automatic solution, where an efficient segmentation step would be needed as preprocessing to the recognition system. The segmentation at the specimen-level is coarse, where cells in the interphase are segmented along with cells in the mitosis and artifacts (like air bubbles etc). On the other hand, a cell-level classification needs an isolated cell mask, provided by the segmentation step. Thus a prior segmentation step plays an important role in the automatic classification of the HEp2-cells. To this end, our first goal is to provide efficient segmentations of cells images, useful to both cell-level and specimen level classification approaches

Image segmentation aims at partitioning an image into significant details, making it easier to analyze certain regions, objects, shapes, contours or curves existed in the image. These details can be highlighted using the texture, color or intensity property of the image. Image segmentation has wide applications in computer vision not limited to medical imaging, object detection, video surveillance, and biometrics. Although image segmentation is a well studied problem, it is neither an easier task which highly depends on the application image characteristics nor it is robust to image quality. As discussed earlier, positive intensity IIF images are higher in contrast as compared to intermediate images; therefore, any segmentation algorithm for IIF images need to be designed carefully. A common segmentation algorithm may result in less accurate results for both types of images. Most existing techniques in the literature have utilized Otsu thresholding [69], watershed [131], level sets [133] and Laplacian of Gaussians filtering [132] for cell image segmentation. One drawback common to most of the proposed approaches based on these methods is that the optimal system parameters are not invariant to image quality, thereby making them less robust. Significant efforts are required to develop efficient techniques for HEp-2 cell segmentation from IIF images. We provide a preliminary work on cell segmentation using a recent filtering technique [134] which has been applied to a similar problem.

3.6.1 Local Convergence Filters

Local Convergence Filters (LCF) [134] estimates the degree of convergence of the gradient vectors within a local area (support region) toward a pixel of interest (area central location). The degree of convergence is related to the distribution of the directions of the gradient vectors and not to their magnitudes. This makes it easier to define a global threshold that is not affected by image illumination. The local convergence of a gradient vector at a given pixel can thus be defined as the cosine of its orientation with respect to the line connecting that pixel and the area's central pixel. Different LCF differ in the definition of the local area in which convergence is evaluated. In this work we use Sliding Band Filter (SBF), a member of the LCF family. The technqiue has valuable characteristics, which are suitable for the analysis of fluorescence images with inherent low contrast. It is straightforward that convergence filters are more complex and as such need more computation to perform detection than point thresholding operations. In our initial work on the problem it has shown reasonably good results on five of the classes with both intermediate and positive intensities. It has failed on some of the images of the cytoplasmic class. The reason is that cytoplasmic images have a very different pattern compared to all other types, since here the fluorescent signal comes from the cytoplasm rather than the nucleus of the cell.

3.6.2 Method

The main steps of our method are as follow:

- A Preprocessing step of K-means clustering for initial rough segmentation which is used as a mask in the subsequent step for speed up and to remove false positives. Morphological processing of the mask are performed such as filling, dilation, erosion and noise removal.

- Sliding Band Filter (SBF) applied to the original images at the locations of masks generated in step 1. It also includes Gaussian filtering of the original images and Shape regularization of the detected cell shapes to get a convex shape.

- Post-Processing retaining the best cell detection with the highest convergence, in-case of multiple detections.

3.6.3 Results

In the following figures we show the cell segmentations for images with both positive and intermediate level of intensity. It can be visually observed that the proper use of Local Convergence filters can produce reasonably good segmentation results.

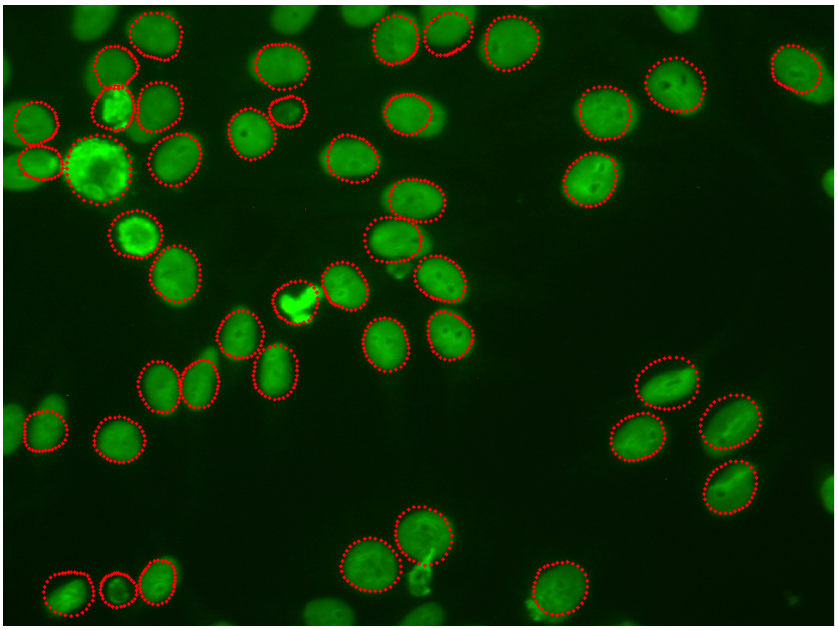


Figure 3.9: Segmentation results for Homogeneous class with positive intensity

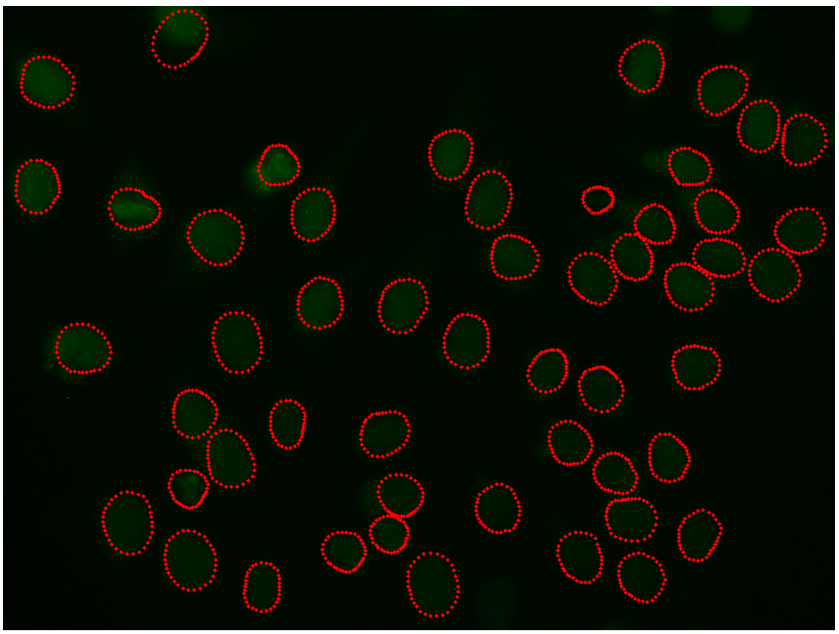


Figure 3.10: Segmentation results for Homogeneous class with intermediate intensity

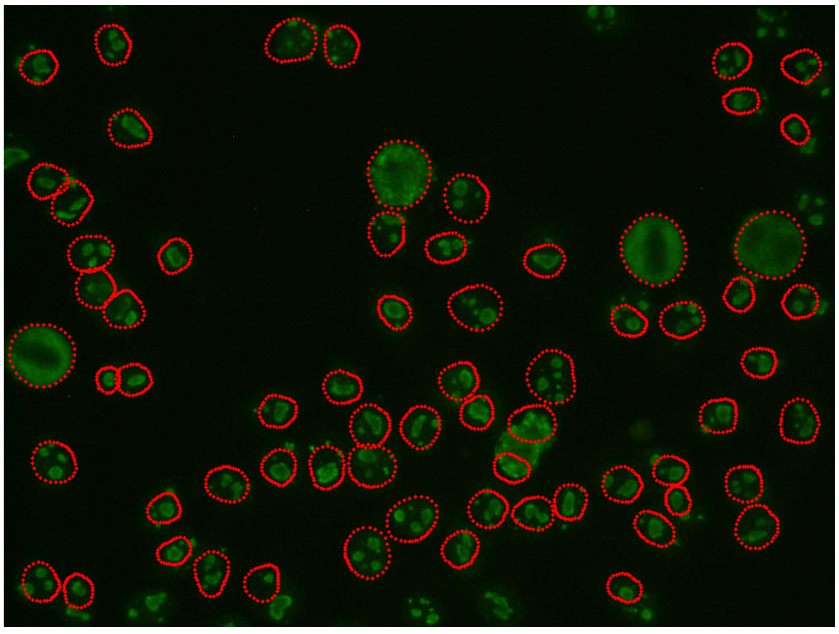


Figure 3.11: Segmentation results for Nucleor class with positive intensity

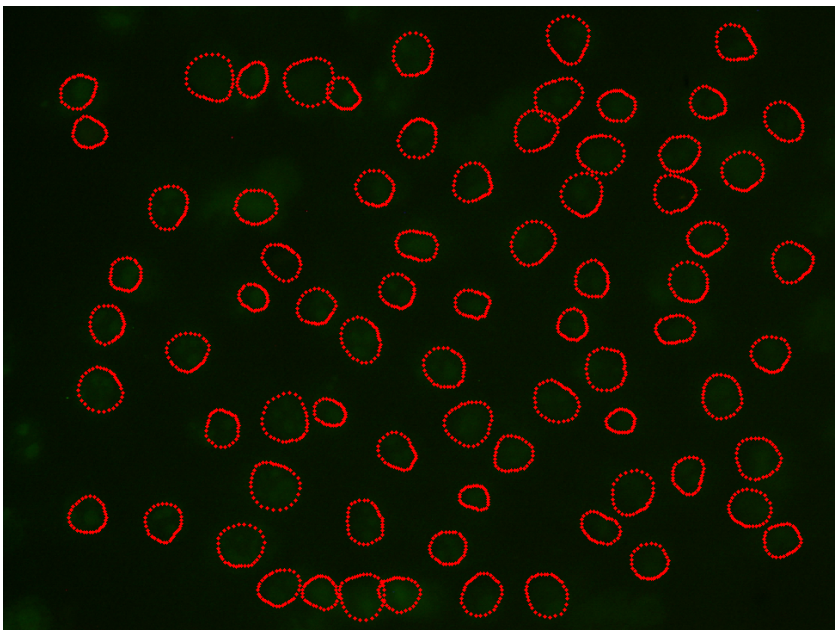


Figure 3.12: Segmentation results for Nucleor class with intermediate intensity

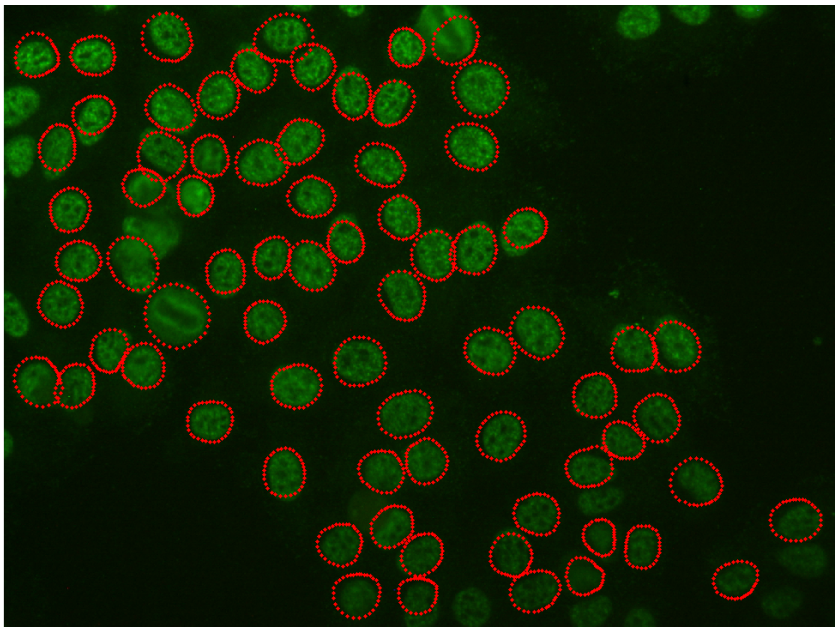


Figure 3.13: Segmentation results for Coarse Speckled class with positive intensity

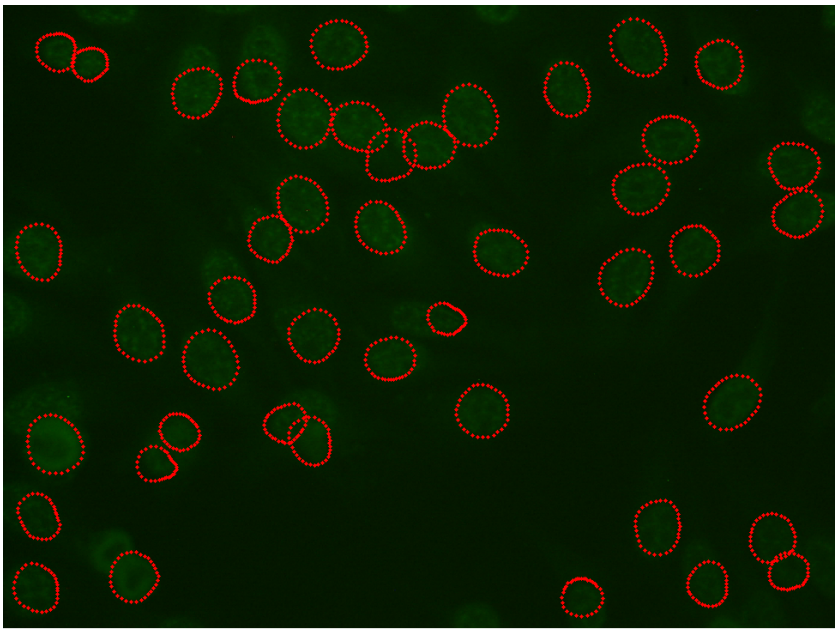


Figure 3.14: Segmentation results for Coarse Speckled class with intermediate intensity

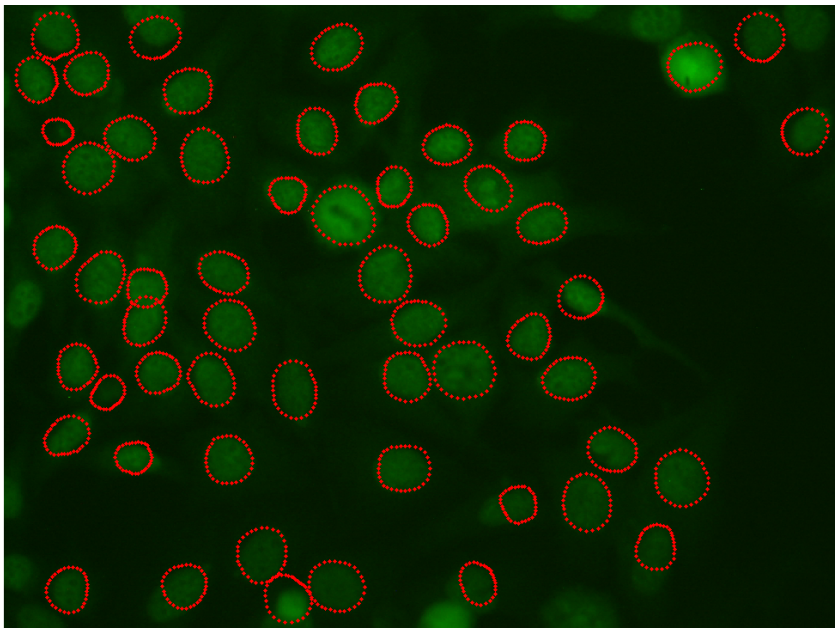


Figure 3.15: Segmentation results for Fine Speckled class with positive intensity

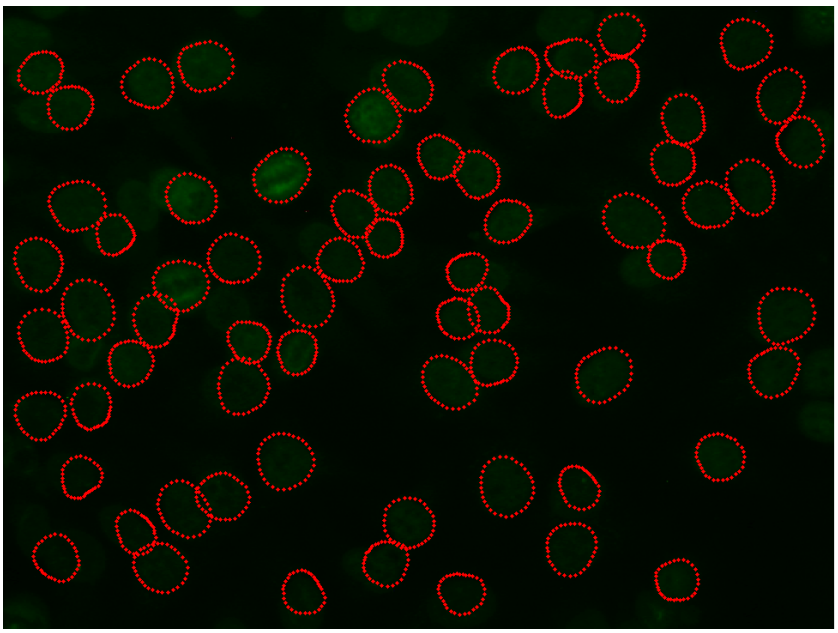


Figure 3.16: Segmentation results for Fine Speckled class with intermediate intensity

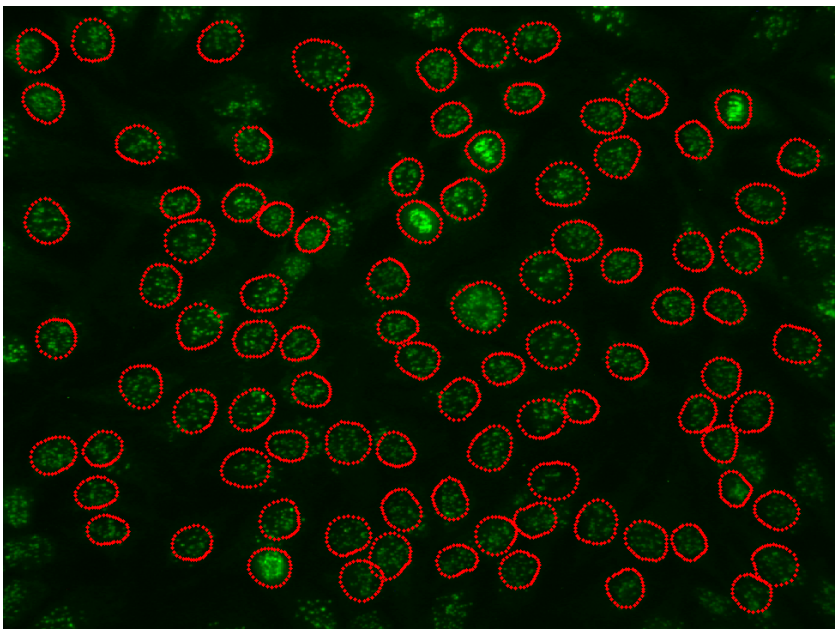


Figure 3.17: Segmentation results for Centromere class with positive intensity

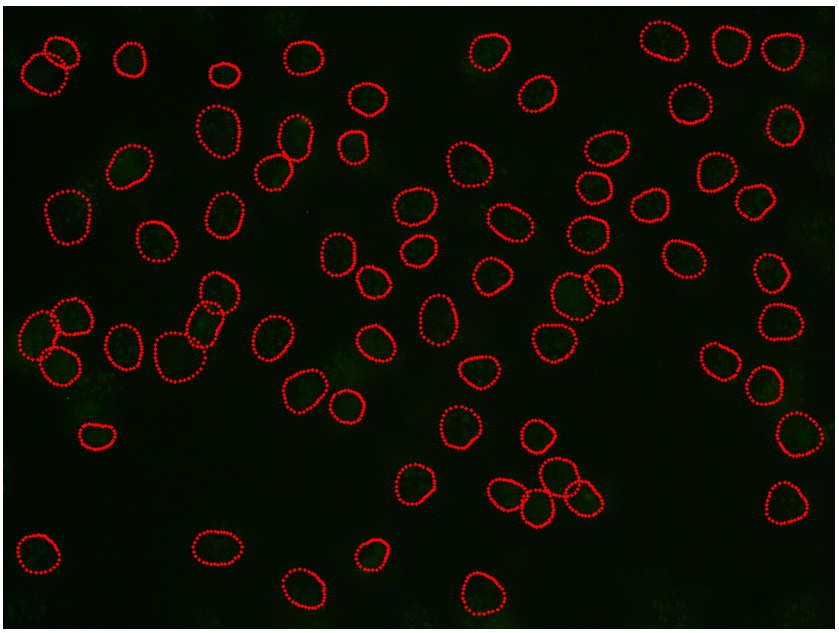


Figure 3.18: Segmentation results for Centromere class with intermediate intensity

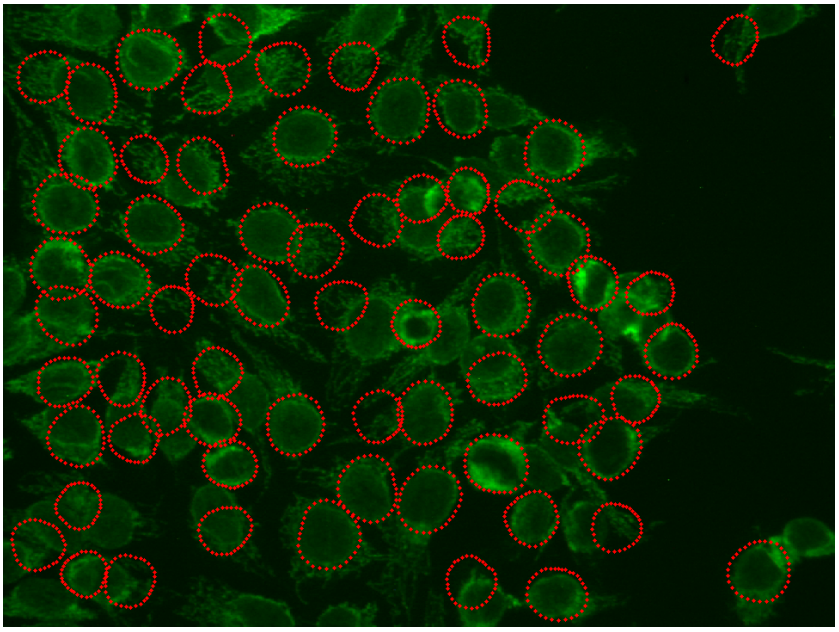


Figure 3.19: Segmentation results for Cytoplasmic class with positive intensity

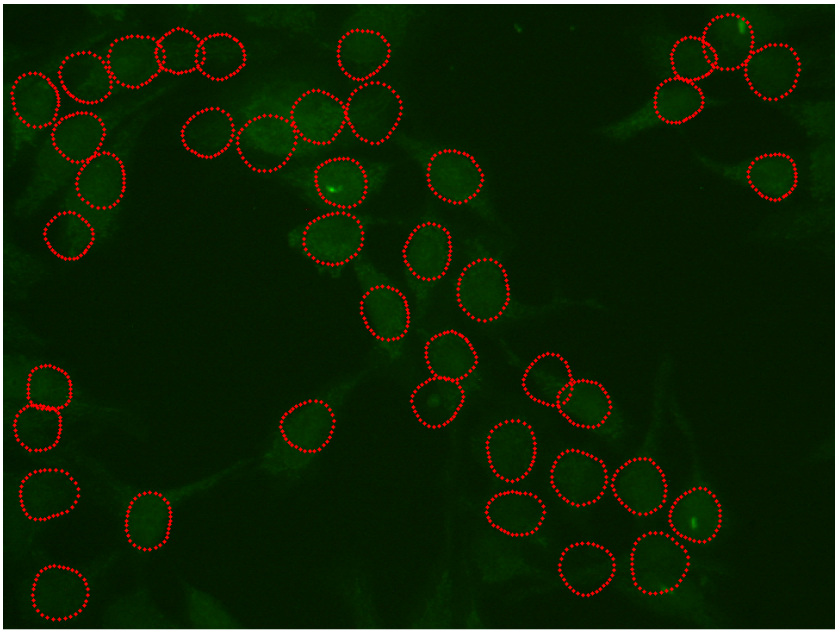


Figure 3.20: Segmentation results for Cytoplasmic class with intermediate intensity

Chapter 4

Automatic Kinship Verification of PARENT-CHILD Pairs from Face Images

Automatic Kinship verification aims at recognizing the degree of kinship of two individuals from their facial images and it has possible applications in image retrieval and annotation, forensics and historical studies. This is a recent and challenging problem, which must deal with different degrees of kinship and variations in age and gender. Our work explores the computer identification of parent-child pairs using a combination of (i) features of different natures, based on geometric and textural data, (ii) feature selection and (iii) state-of-the-art classifiers. Experiments show that the proposed approach provides a valuable solution to the kinship verification problem, as suggested by its comparison with different methods on the same data and the same experimental protocols. We further show the good generalization capabilities of our method in several cross-database experiments.

The remaining of the chapter is organized as follows. An introduction to the Problem and related research work is presented in section 4.1. In Section 4.2, we describe the database of parent-child images we used in our main experiments. Section 4.3 details the algorithm we used for tackling the kin verification problem. Finaly experimental results are presented in Section 4.4 and discussed in Section 4.5.

4.1 Introduction and Related Work

Human faces have been much studied in computer vision and machine learning. Investigations led to the development of a number of interesting applications, such as face recognition, expression analysis, age estimation, gender classification and ethnicity recognition. Among the new emerging areas there is automatic Kinship Verification. The automatic detection of kinship from facial images is a difficult problem that recently received a growing attention from the computer vision and pattern recognition research community. Kinship verification aims at recognizing the degree of kinship of two individuals from their facial appearance (See Figure 4.1). Kinship verification has several interesting applications, such as searching kin images into digital photo collections, looking for missing family members, and performing family relationship assessment in forensics.

Unlike typical face recognition, Kinship verification aims at identifying similar features between two different individuals. This is a challenging task since it has to deal with differences in age and gender between subjects; additionally, to the unpredictable amount of genetic information shared by relatives that reflects into individuals showing different degrees of facial similarity. Many other difficulties arise when Kinship Verification must be applied to databases acquired under unconstrained environments. Despite the potential interest into this topic, to the best of our knowledge, the very first work on kinship verification was published in 2010 by Fang et al.[3]. Since then, several papers on the subject have been presented, tackling different aspects of the problem.

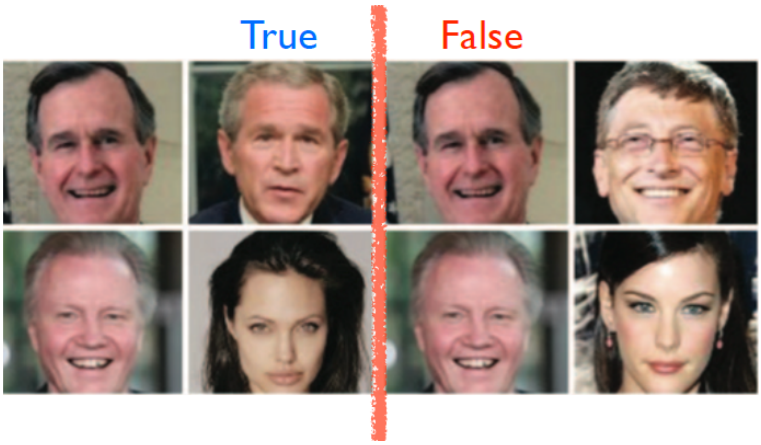


Figure 4.1: Examples of kin and not-kin samples of parent-child pairs from the work of Fanget al. [3]

Kinship has been largely discussed in human science areas such as psychology

and sociology. Researchers highlight that human beings have a natural capability to recognize kinship relationships between unknown individuals[1], and even attempted to identify the facial features providing kinship clues[1] [2].

Usually, various combinations of facial features, extracted with holistic or mixed geometric-textural techniques, were used to characterize individual's images. Geometric measurements, color of selected image regions and holistic textural features were analysed by Fang et al.[3], local image descriptors, Gabor wavelets and intensity related information in the work of Somanath et al.[4], DAISY descriptors of salient regions by Guo and Wang [5], various local textural features in Lu et al.[6], Gabor features from selected regions in Xia et al.[7], a combination of geometric, holistic and local and global textural descriptors in the work of Vieira et al.[8], and local self similarities in Weber normalized faces in Kholi et al.[9].

Most of the approaches relied on state of the art classifiers, such as k-Nearest Neighbours[3] [7] and Support Vector Machines (SVM)[3] [8] [9]. Metric learning methods were also introduced to improve the classification accuracy[4] [6]. Their basic idea is to learn, in the feature space, a distance metric aimed at pushing as close as possible the positive samples (kins) and as far as possible the negative ones (non kins), thus providing a better separation of classes. Another interesting approach was presented by Xia et al.[7], who proposed an extended Transfer Subspace Learning (TSL) to consider the influence of age factor in the identification of parent-child pairs. The aim of TSL is to simplify the recognition of children and parent pairs by transferring the knowledge learnt from the similar, but easier task, of recognizing the same parent-child pairs but using images of parents in youth.

A thorough comparison of these methods is difficult, since most of them analysed different data. However, in order to assess their results, several authors compared the capabilities of their approaches with that of a panel of human raters, showing that automatic approaches are superior, or at worse comparable, with the discriminative capabilities of human beings[3] [6] [7] [8].

The recent availability of a number of kinship datasets allows a comparative analysis of the different approaches (as in [6], [10]). Some of the publicly available datasets are the following: (*i*) KV [3]; (*ii*) UBKIN [7]; (*iii*) SiblingsDB [8]; (*iv*) VADANA [82] and (*v*) KinFaceW-I and II [6].

Among these databases, the one we found most suited to test our approach was the Kinship Verification dataset (KV) collected by Fang et al.[3]. The authors also provided results from human panels on the same dataset and thus providing a reliable ground for comparison. The photographs here were taken from several public celebrities of different age, gender and race in slight different poses (mostly frontal), illumination conditions and expression (some neutral but often smiling).

Examples of images in this DB are shown in Figure 4.2. A description of the dataset is provided in Section 4.2.

Figure 4.2: Database of image pairs collected from the internet by Fang et al.[3]

The general flow of our approach is following; i) normalizing all images in the database; ii) extracting features from the single images; iii) fusing the feature vectors obtained using different descriptors ; iv)combining the representative vectors of couples; v)scaling the feature vectors to the range [0 1] ;and, finally vi) classifying the datasets and selecting most descriptive features.

Our work explores the computer identification of parent-child pairs using a combination of geometric and textural features. These attributes, related to holistic and feature-based classification techniques, were first used individually in various classifiers, in order to investigate which data are more effective for our problem. Then, they were fused into a more effective classifier. Image pair classification is based on the Support Vector Machines [19] algorithm and on the integration of a two-step Feature Selection (FS) process, which led to improvements in the classification accuracy. We provide a comparison with different methods on the same dataset and the same experimental protocols. We further show the good generalization capabilities of our method in several cross-database experiments.

We also adopted a similar approach to address the problem of kinship verification in an unconstrained environment. For the purpose we utilized a challenging datasets Kinship Face in the Wild (KinFaceW) [6], containing large heterogeneity with unconstrained face orientations, expressions and occlusions. To this end, we employ two new textural descriptors, capable of describing the data under different

perspectives, resulting in significantly higher accuracies than two baseline methods from the literature. The details of the work are provided in Chapter 5.

4.2 Image Database

The recent interest in kinship verification led to availability of some databases composed of facial images of individuals related by different kinship degrees. Examples are those used in the works referenced in the Introduction. These databases are composed by a heterogeneous set of images, mostly collected through the Internet.

Among these databases, the one we found most suited to test our approach was the Kinship Verification dataset (KV) collected by Fang et al.[3]. It consists of 286 individuals' color images[1] (143 parent-child pairs). The image size is 100 × 100 pixels. The photographs depict several public celebrities with different age, gender and race in slightly different poses (mostly frontal), illumination conditions and expressions (some neutral but often smiling, see Fig. 4.4). The subjects are 50% Caucasians, 40% Asians, 7% African Americans, and 3% of other ethnicities; 40% of the samples are father-son, 22% are father-daughter, 13% are mother-son, and 26% are mother-daughter pairs.

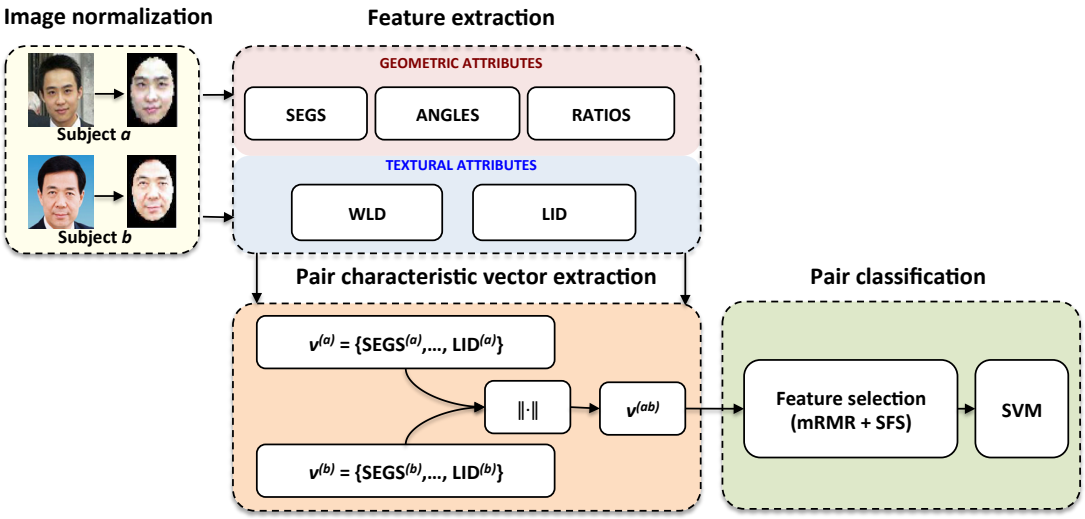


Figure 4.3: Scheme of the proposed approach.

[1]Although the authors reported the use of 150 pairs, 300 images, the online version contains only 143 pairs due to privacy constraints.

4.3 Proposed Method

The basic idea of our approach is to investigate: (i) which are the facial features most relevant for the automatic kinship verification problem and (ii) how to build a classifier on them. To this end, different types of features, or *attributes*, have been tested in our experiments, first individually and then combined, in order to define a feature set which best characterizes the images being analysed, *i.e.* which maximizes the classification accuracy.

The samples used in our experiments were always obtained from an *Individual Dataset*, which includes frontal images of subjects (one image per individual). For each *Individual Dataset*, we created a *Pair Dataset*, where each element represented a pair of subjects. The Pair Datasets were composed by an equal number of positive (kin) and negative (non-kin) samples, where the negative samples were chosen at random. Each individual appears exactly once in both positive and negative sets.

The outline of our kinship verification algorithm is the following (see also Figure 4.3).

1. All images have been normalized.
2. Each individual has been characterized with several *characteristic vectors*, each obtained extracting a different attribute from his/her image; the list of attributes will be described in the following.
3. For each attribute, we computed the *characteristic vector* $\mathbf{v}^{(ab)}$ for a pair of individuals a and b as the vector of Euclidean distances, in their respective n-dimensional space, of the corresponding elements of the characteristic vectors of a and b. With this formulation, the characteristic vectors of a pair are commutative, i.e. $\mathbf{v}^{(ab)} = \mathbf{v}^{(ba)}$.
4. For each *Pair Dataset* and each attribute, and for some combination of attributes (as explained in the following), the most relevant elements of the corresponding pair characteristic vectors were selected with a two step feature selection algorithm to construct an SVM classifier.

These steps are detailed in the following subsections.

4.3.1 Image Normalization

This pre-processing step was aimed at: (i) aligning images in the DBs with a reference frame and delimiting the same area to include the whole face region, and

(ii) subtracting the background, which could potentially affect some of the extracted features.

Both processes rely on different subsets of the facial landmarks used to compute relevant facial features (see Section 4.3.2). For each face, a total of 76 kepoints were automatically extracted with the Active Shape Models (ASM) technique[13] using an Open Source implementation available on the Internet[14].

Image alignment was first obtained by making coincident, in a reference rectangular area of size 100×100 pixels, the exterior eye corners with two predefined positions at, respectively, points (80,80) and (120,80). Then, the ellipse best fitting the 15 landmarks around the chin was used to segment the face. Examples of normalized images can be seen in Figure 4.4. As a further information, the average normalized face size is 60×60 pixels.

Figure 4.4: Examples of parent (1^{st} row) - child (2^{nd} row) pairs. For each individual, we show the original image, the detected landmarks and the normalized image

4.3.2 Face Representation

Different representations of the information conveyed by faces have been experimented for tackling the various face image processing problems. Previous studies have shown that the discriminative power of separate facial attributes is likely to be improved by that of the integration of pieces of information of different natures[6].

According to these findings, we also experimented the integration of different feature types. The choice of facial attributes used in our kinship verification approach took into account the lessons learned during our previous experience with sibling verification[8], where we analysed the contribution of different geometric, textural and holistic features. In that work, we found that the contribution of holistic attributes was negligible. We also identified, among many different options, a set of geometric and textural features particularly effective for siblings' classification.

Thus, we initially considered, for the parent-child verification problem, this optimal set of attributes, which was then integrated with some new attributes found highly discriminative in preliminary experiments[2].

In summary, the attributes we used to characterize images can be divided into two main categories (Table 4.1):

1. geometric attributes, focusing on shape features computed from the position of facial landmarks;
2. texture descriptors, summarizing texture features of the whole image or of small regions surrounding landmarks.

Table 4.1: Attribute dimensions.

Attribute	size of a feat. × feature nr.
Geometric attributes	
SEGS	1×184
ANGLES	1×342
RATIOS	1×862
Textural attributes	
WLD	1×2880
LID	384×76

Geometric Features

The geometric attributes analysed in our work have been computed from a dense net of segments connecting the 76 ASM landmarks extracted from each image.

In order to compare the same elements on all faces, we defined a reference net of 184 segments, corresponding to the edges of the Delaunay Triangulation (DT) of the average position of the normalized landmarks over all faces in the KV dataset. The reference triangle mesh (see Figure 4.5) is composed by 114 elements.

We considered the following geometric attributes.

SEGS – This attribute contains the (normalized) lengths of the 184 reference facial segments. These lengths are effective in capturing global face shape and in

[2]These results are not reported here for the sake of brevity.

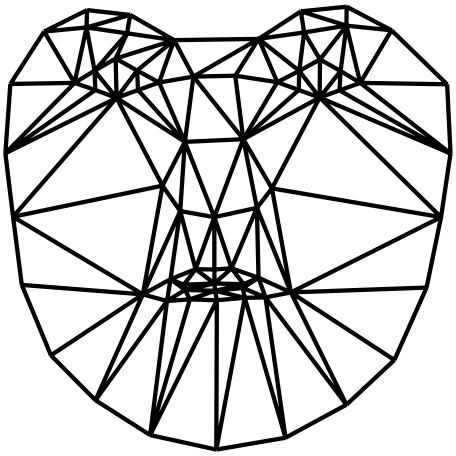


Figure 4.5: Reference Triangle mesh generated using Delaunay Triangulation of the average positions of the normalized landmarks over the KV dataset.

describing, directly or through segment chains, representative facial measurements (*eg.* distances between mouth and eyes, mouth width, etc.).

ANGLES – Angles of the triangles obtained from DT. Angles and segment lengths are strictly correlated, since the former can be computed from the latter. However, as we will show in the following, some variables can be discarded by a feature selection algorithm applied before classification, and the relationship between lengths and angles can be lost. The total number of angles is 342.

RATIOS – This attribute contains the ratios between lengths of DT segments having a vertex in common. Two segments are used only once to compute a ratio, *i.e.*, once a ratio is computed, its inverse is not considered. RATIOS can suitably summarize local shape similarities, which can be relevant to our problem. With this attribute, a face is represented with 862 values.

Textural Features

Two image descriptors were used to characterize each individual's normalized image.

WLD – The Weber local Descriptor (WLD) is a dense local image descriptor, which has shown good performances on texture classification and face detection [15] and, more recently, on face recognition [79]. This law states that a just-noticeable difference in a stimulus is proportional to the magnitude of the original stimulus. Translating this concept into image intensities, WLD first characterizes a pixel with two values: the *differential excitation* (the ratio between the sum of neighbouring pixel intensity and the intensity of the pixel itself) and the *orientation* of the pixel

gradient. Multi-scale analysis is then applied to characterize salient local patterns at different resolutions and, finally, all the WLD features of the image are encoded into a histogram containing 2.880 elements.

LID – Local Image Descriptors. In this attribute, SIFT-based descriptors have been used to characterize the regions surrounding facial landmarks. SIFT[16] works on monochromatic images and provides invariance to several image changes, such as uniform scaling and rotation, and partial invariance to affine distortion and illumination changes.

Recently, color local image descriptors have been introduced. Their general idea is to encode, for a reference point and for a specific color model, the Scale-Invariant Feature Transform (SIFT) descriptors computed separately on each image channel. This allows to improve the invariance properties of the descriptor (e.g. introducing invariance to light intensity change and shift and light color change and shift[17]). Among the color descriptors surveyed in Ref. [17], we choose C-SIFT[18] since in our preliminary experiments it performed slightly but consistently better. To characterize a sample with this attribute, we computed a C-SIFT descriptor (a vector of 384 components) on each of the 76 facial landmarks.

It should be noted that C-SIFT can only be used when all images to be analysed are colour images. When dealing with datasets including grayscale images, a different characterization of the LID attribute must be considered. We will discuss this point in more details in Section 4.4.2.

4.3.3 Building the Classifiers

Kinship recognition can be modelled as a binary classification problem. Our classifiers were based on SVM[19], a well established machine learning technique that has been proven successful in many applications[20].

In our work, we used SVM with a radial basis kernel, optimizing the kernel parameters by means of five-fold cross-validation technique and a grid search over the parameters space, as suggested in Ref. [20]. Before applying SVM, each element of the characteristic vectors was linearly scaled to the range [0,1]. This avoids the variables in larger scales to dominate those in smaller ranges and reduces numerical problems in the computation of the SVM kernels[21].

Feature Selection

Feature selection (FS) is a dimensionality reduction pre-processing step vital in many machine learning and pattern recognition tasks. It aims at discarding redundant and/or irrelevant information by selecting a subset of variables that best describes the data according to a specific criterion. In the classification context, this criterion is the improvement of the classification accuracy. Another advantage of FS is that, reducing the data dimension, it improves the comprehension of data and results by telling which are the most important features and how they are correlated.

While SVMs are generally acknowledged for their generalization capabilities, since they implicitly map data in a transformed space that emphasizes the features more crucial to the classification purpose, their integration with FS schemes provides several advantages[22, 23]. In particular, it helps reducing the probability of overfitting, which is likely to affect our classifiers since in most of the experiments the number of features is much greater than the number of available samples.

In our method, we applied a robust two-step FS method (see Figure 4.6).

First step (mRMR). Features are ranked according to their relevance in characterizing the classification variable with the minimum-Redundancy-Maximum-Relevance (mRMR) algorithm[74]. The sorting criterion of mRMR points at the contemporaneous minimization of the mutual similarity and maximization of the correlation between the features and the classification variable. Since mRMR provides a mere ranking of the features with no information on the size of the optimal feature set, we heuristically looked for this size by iteratively increasing the number of candidates in the feature set until the global optimum of the classification accuracy was found.

Second step (SFS). The set of candidate features is further refined through a Sequential Forward Selection (SFS) scheme[24]. The classical implementation of SFS starts from an empty feature set S and adds at each iteration the variable that provides the highest improvement of the classification accuracy. The process is repeated until no more improvements can be obtained. However, as we experienced, this stopping criterion tends to trap the algorithm in a local minima. Therefore, we proceeded until at most 70 features[3] were added to S, and then we selected the feature set corresponding to the iteration providing the best accuracy.

[3]This choice aimed at maintaining reasonable computational times.

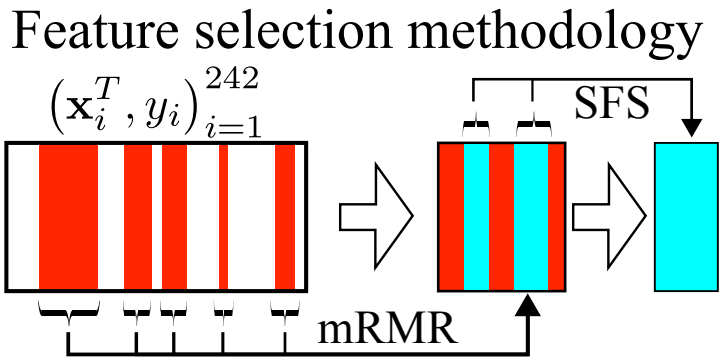


Figure 4.6: Two-stage Feature Selection process.

4.4 Results

Our experiments were organized as follows. We first analysed the contribution of different facial attributes to the parent-child classification and how they can be combined to improve the classification accuracy, *i.e.* the percentage of samples correctly classified (Section 4.4.1). This thorough analysis was performed on the KV dataset (*c.f.*, Section 4.2). Then, in order to understand the generalization properties of our classifier, we first conducted within-database experiments on different datasets (Section 4.4.2) and finally we performed more challenging cross-database experiments (Section 4.4.3).

4.4.1 Classification Accuracies

For the sake of clarity, we briefly recall the main idea behind our approach. Different representations of the information conveyed by faces have been found effective in various problems related to facial analysis. However, the relevance of these pieces of information and their mutual contribution to the kinship verification problem is still an open issue.

In this work we postulate that the integration of facial attributes of different natures improves the capabilities of a classifier in recognizing kin relationships. To this end, we defined a set of geometric features and we identified different textural descriptors that appeared promising for the task considered.

In our experiments, we first computed the accuracies based on the classification of the individual attributes described in Section 4.3.2. Then, we evaluated the accuracies obtained by characterizing each facial image with three different groups of

attributes: GEOMETRIC, grouping the geometric attributes, TEXTURE, combining textural information, and ALL, concatenating all the described attributes. FS for grouped attributes was applied to the sets of all the merged features.

Results were assessed using stratified five-fold cross-validation (CV) and, hence, we report the average classification rates of each classifier over the different CV rounds. Results are summarized in Table 4.2 and organized by attribute, or attribute group.

Table 4.2: Accuracy results for different attributes/compound attributes on the KV dataset. We also show the number of variables selected by the FS process and the calssification accuracy obtained without applying FS.

	Accuracy	Nr. features selected	Accuracy without FS
SEGS	73.1	21	59.4
ANGLES	69.6	6	56.6
RATIOS	75.2	35	55.6
LID	76.9	14	70.4
WLD	79.3	30	50.0
GEOMETRIC	75.8	26	55.9
TEXTURE	81.1	25	47.9
ALL	83.2	55	48.6

The following initial remarks can be drawn:

- when considered individually, textural features have higher discriminative power than the geometric ones, with WLD obtaining the best performances (79.3%);
- as for geometric attributes, the better performances of RATIOS (75.1%) suggest that their capability of capturing local similarities of facial shapes provides stronger kinship clues;
- the more heterogeneous the information, the better the accuracies. Grouped attributes performed consistently better than their single components, and the best accuracies were obtained considering all attributes together (83.2%).

The quality of our results can be assessed by comparing the classification accuracies of our experiments with that obtained on the same dataset by Fang et al.[3] and Lu et al.[6]. The performance of their approaches (respectively, 70.69%[3] and 71.6%[6]) and that of a panel of human raters on the same data (67.19%[3]) are already improved by that obtained in our work with several individual attributes, and outperformed by our best result (83.2%), achieved with the integration of all attributes.

Analysis of the Feature Selection process

As it can be seen in Table 4.2, Feature Selection (FS) always provides a significant accuracy improvement. Larger gaps can be observed when the number of features is much greater than the number of samples, thus highlighting the FS capabilities to prevent overfitting.

Concerning the selection process, it is also interesting to analyse the distribution of features surviving the FS pruning for composite attributes (Figure 4.7), which could provide some insights about the more relevant facial characteristics to recognize kins.

We can observe the following: (i) RATIOS is the most relevant geometric attribute, suggesting its good descriptive capabilities; (ii) as for textural features, WLD is more relevant than CLID in the TEXTURE group, and their relative contribution is similar when geometric features are added (ALL group); (iii) when attributes are grouped, elements from all attributes are chosen to compose the final vector; (iv) when geometric and textural features are combined, the latter are preferably selected to compose the final dataset.

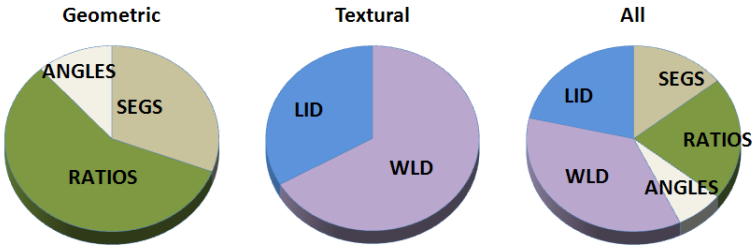


Figure 4.7: Feature selection: distribution of features per type for different attribute groups.

Comparison with Different Classifiers

We compared the performances of our SVM-based classifiers with other state-of-the-art classification algorithms (namely, Random Decision Forests, k-Nearest Neighbors and Bagging Trees). Different FS approaches were applied. The k-Nearest Neighbors classifier was integrated with the same two step FS method described in Section 4.3.3. For Random Decision Forests, we selected the most relevant features by first sorting them according to their Variable Importance[25] and then choosing heuristically the optimal feature set size. A similar approach was applied to Bagging

Trees. Figure 4.8 shows the Out-Of-Bag (OOB) error rate for the Random Decision Forests.

The results of these experiments are summarized in Table 4.3, where, for the sake of brevity, we simply report the best result, obtained by each algorithm with the ALL group. Overall, SVM provided the best performance. However, the relatively small differences with other classifiers highlight that the major contribution to the classification accuracy is provided by the design of our features rather than by the classification algorithm itself.

Table 4.3: Accuracies of different classifiers.

	SVM	RDF	k-NN	BTrees
Accuracy	83.2	77.5	79.1	78.5

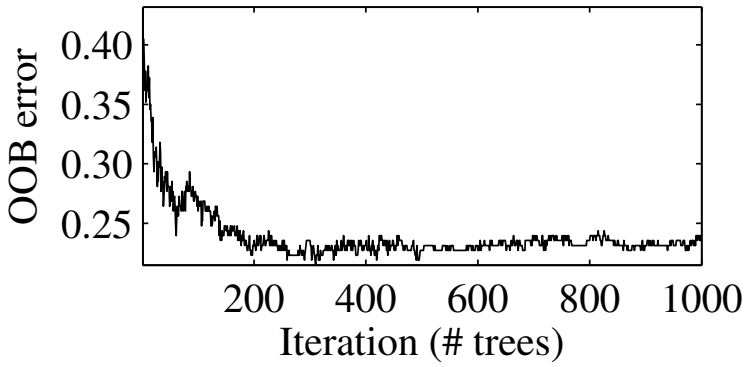


Figure 4.8: Out-Of-Bag (OOB) error rate for Random Decision Forest.

4.4.2 Experiments on Different Datasets

We further explored the potentialities of our approach analysing two other publicly available datasets, the UB Kinface[7] and the KinFaceW-II[6]. In both datasets, images have different resolutions, orientations, illuminations and expressions. Most of them are grayscale and in KinFaceW-II some contain partial occlusions. The variability of the imaging conditions makes these datasets much more challenging than the KV dataset. In particular, the not perfectly frontal alignment of the faces can affect some attributes, reducing their discriminative capabilities.

The analysis of these datasets required to face two other issues. First, the presence of grayscale images prevents the extraction of color descriptors for all samples. To address this problem, we applied a different definition of the LID attribute,

building it from the SIFT descriptors computed on each facial landmark. Second, the automatic identification of landmarks provided large mismatches or false face identification for several images. To this end, we manually assigned the position of the external eye corners, an information which was available for both datasets, to initialize the ASMs.

Experiments on UB Kinface

The UB Kinface dataset includes 600 images of 400 individuals, divided into 200 kin pairs. Each pair is characterized by a single frontal image of a child and two images of his/her young and old parent. The average image size is 422×336 pixels and the average face size, obtained with ASM, is 113×114 pixels.

The dataset is divided into two individual subsets: Set 1, including the 200 children and their old parents, and Set 2, where the 200 children were paired with their young parents. As in the experiments of Section 4.4.1, for each Individual dataset a Pair dataset was created and classification results were assessed using stratified five-fold cross-validation.

The results are summarized in Table 4.4. The classification of Set 1 and 2 achieved, respectively, 70.3% and 70.5%. These results can be assessed by comparing them with two other approaches, namely, the Transfer Subspace Learning[7] (TSL), and the Multiview Neighborhood Repulsed Metric Learning[6] (MNRML). TSL used the images of the young parents in Set 2 as an intermediate class to learn an optimal discriminative subspace between target classes in Set 1 (children and their old parents). Thus, TSL accuracies are available only for Set 1.

The results show that our approach outperforms TSL (+10.3%) and is comparable with MNRML (+3.0% and +3.7% on, respectively, Sets 1 and 2). A further comparison can be made with the discriminative capabilities of human raters. Human ratings were collected for Set 1 only[7] and our results largely improves them (+14.3%).

A possible question arises from these results: how did the choice of using SIFT, instead of colour descriptors, in the LID attribute affect the performances? To answer this question, we selected the subset of 169 positive pairs from Set 1 having color images for both individuals and we built a Pair dataset on them. We applied the same classification procedure described in Section 4.3 but with two different formulations of the LID attribute (namely, using SIFT and C-SIFT). The same procedure was repeated for the KV dataset. Results show that, despite the lower discriminative power of SIFT, the overall accuracy remains substantially unaffected (-2.6% for UB Kinface and -2.4% for KV dataset). A possible explanation of this

behaviour is that the integration with other features, backed by the FS process, helps to soften the major limitations of specific attributes.

Table 4.4: UB Kin results.

	Set 1	Set 2
TSL	60.0	N.A.
MNRML	67.3	66.8
Human Panel	56.0	N.A.
Our method	70.3	70.5
Nr. features selected	*16*	*18*

Experiments on KinFaceW-II

The KinFaceW-II includes 2000 images of individuals related by four kinship relations (Father-Son, Father-Daughter, Mother-Son and Mother-Daughter). Each category contains 250 pairs. The online version of KinFaceW-II provides the images after the first preprocessing step of the method in Ref. [6], *i.e.*, facial identification and normalization into a standard area of 64×64 pixels. Since these images were cropped discarding in most cases relevant facial features exploited by our method (*eg.* the chin line), we processed the original raw images used to create the datasets, which were kindly granted by the authors. Their average size is 200×146 pixels and the average face size is 103×103 pixels.

Each subset of the KinFaceW-II was first analysed separately. A Pair dataset was built for each subset and sequentially divided into five folds for cross validation according to the experimental protocol described in Ref. [6]. The accuracy of our approach can be compared with those of a panel of human raters and of the MNRML method[6]. As can be seen in Table 4.5 our method improves both of them.

Finally, consistent with previous experiments, we gathered together all the KinFaceW-II dataset to analyse the more general two-class kinship verification problem, *i.e.* deciding if two individuals are kin or not. The accuracy obtained was 77.8%. The loss in the performances, with respect to the average accuracy obtained by the classification of the separate kin subclasses, suggests that the global kin manifold in face space has a more complex shape than the sub-manifolds of specific kinship relations. This observation highlights as well the fact that methods capable of better coping with the high within-variance of kin samples should be designed.

Table 4.5: KinFaceW-II results.

	F-S	F-D	M-S	M-D	Mean
MNRML	76.9	74.3	77.4	77.6	76.5
Human Panel	70.0	68.0	78.0	80.0	74.0
Our method	82.7	76.9	82.3	83.9	81.5
Nr. features selected	*67*	*43*	*59*	*68*	-

4.4.3 Cross-database Experiments

In order to further assess the generalization capabilities of our approach we performed the following cross-database experiments:

- we built a classifier using ALL attributes and SIFT as LID;
- we trained the classifier with all the samples in the KV Pair dataset used in the experiments in Section 4.4.1;
- we used three different test Pair datasets: the two used in Section 4.4.2 for the experiments on the UB Kinset (namely, Set 1 and Set 2) and the dataset used in Section 4.4.2 gathering all KinFaceW-II samples;
- we classified the test sets.

Such experimental settings are challenging, due to the large differences between the characteristics of the samples in the training and test sets, and their rationale is to evaluate the performance of the proposed approach in realistic conditions, when unseen data have to be classified.

The cross-database accuracies can be assessed by comparing them with the within-database accuracies of the corresponding test dataset. The results listed in Table 4.6 show the good behaviour of the classifier in the cross-database experiments.

For completeness, we performed more cross-database experiments using the KV dataset as test set and various datasets for training (namely, UB Kinface Set 1 and Set 2 and KinFace-II). The baseline for these experiments is 79.8%, the within-database accuracy computed on the KV dataset using SIFT descriptors in the LID attribute. As can be seen in Table 4.7, despite the lower quality of the train sets, the drop of accuracy is comparable to that obtained in the previous cross-database experiments, highlighting another time the good generalization properties of our approach.

Table 4.6: Cross and within database accuracies will ALL attributes and using KV as train dataset.

Test dataset	Cross-DB acc. (%)	Within-DB acc. (%)
UB Kinface Set 1	70.0	70.3
UB Kinface Set 2	68.8	70.5
KinFaceW-II	76.6	77.8

Table 4.7: Cross database accuracies on KV dataset will ALL attributes.

Train dataset	Cross-DB acc. (%)
UB Kinface Set 1	77.3
UB Kinface Set 2	74.9
KinFaceW-II	79.7

4.4.4 Processing and Simulation Times

Here we briefly discuss the processing times for the different steps that comprise our proposed algorithm. The very first step is the detection of facial landmarks and the normalization of the images. In our work we detect 76 facial landmarks followed by image normalization, which indeed depend on the size of the image, take time in multiple of seconds. However, this time is not crucial as compared to the time taken by the entire algorithm. In the feature extractions step, each feature attribute would be extracted within the range of approximately 1-6 seconds. The color-sift for example took 4-6 seconds, while WLD for instance would complete in 1-2 seconds.

The training of the classifier is the most expensive stage of our pattern recognition system. In the two-step feature selection algorithm the first step mRMR ranking, which works independently from the classifier, categorize the features using the mean $\pm$ deviation scheme. An mRMR implementation is available in Matlab executable file (.mex), which performs better than script files. It needs only a few seconds to rank the features in a single iteration. Since to heuristically find the best-mRMR size as discussed in Section 4.3.3, we need multiple such iterations on the full feature set. However, the second FS step Sequential Forward Selection (SFS) is wrapper based algorithm which involves the classifier or the learning process, is the most expensive one due to its iterative nature. Depending on the size of a single or fused feature set, it would last from tens of minutes to hours. In our experiments, simulations would eventually last for days, due to iterative evaluation of different parameters and testing different combinations of single image attributes.

The testing of an already trained classifier, SVM, is quickly executed. Here, we

first need to normalize and extract features for two images in the pair the same way as in the training. The classification consists in mapping these features to a higher dimensional space using the optimized kernel and finding its distance to the separating hyperplane. This is quickly performed (in a few seconds) by the libSVM package [20], available in Matlab executable file (.mex).

4.5 Discussions

Experimental results showed that the proposed approach is indeed an effective solution to the automatic kinship verification problem. Our method showed good generalization capabilities and outperformed the recognition capabilities of human raters and of previous approaches in the literature.

Despite that, as we discussed in the Introduction, there are several factors affecting the accuracy of an automatic kinship predictor and fully understanding their influence requires further work. A problem to face with such studies is the lack of large annotated databases of kins, similar to those used in face recognition studies, spanning several dimensions including ethnicity, age, expressions and image quality. Given the recent attention of the research community towards automatic kinship recognition, we expect this issue to be addressed soon.

Preliminary insights on some of these elements can be obtained from our experiments.

Ethnicity. Except KinFace-II, which includes a majority of Caucasian individuals (about 90%), the other datasets used in our experiments are more balanced. Both within and cross-database experiments seem to suggest that ethnicity is not largely affecting the results.

Age. UB kinface included, for each child, images of his/her young and old parent. Although age information was not available, which makes it impossible to evaluate the correlation between age and classification variable, the results in Table 4.4 seem to indicate that our approach is unaffected by age.

Similarity among individuals. Unrelated people showing high facial resemblance could affect similarity learning-based methods[7] [6] and, to some extent, also descriptor based methods[3] [5] [4] [9], as the one presented in this article. This challenge is similar to that faced by computer vision based biometric systems trying to discriminate between monozygotic or identical twins. We tried to build a dataset of similar but unrelated pairs from our data, using as similarity score the output of FACEVacs[26], a face recognition software which provides for an image pair a

score value in [0,1]. The higher the score, the higher the probability that the images come from the same subject or, in our case, the higher the similarity between two individuals. Results on our datasets showed that, in general, absolute values are quite low (*i.e.*, both positive and negative pairs show a low degree of similarity) and no negative sample scores higher than positive samples in the same dataset. Summarizing, the data used in our work were unsuited to investigate this problem.

Image quality. Image resolution and quality are the most important factors affecting the extracted features and thus the accuracy of the results. A related problem is the accuracy of the facial landmark identification, a necessary preliminary step for several descriptor based methods. This could be an issue when datasets with heterogeneous imaging conditions have to be analysed. Despite that, recent advances in automatic facial landmark identification demonstrate impressive performances in dealing with extremely challenging images[27] [28] [29].

Concluding, we have shown that our proposed approach provides a valuable solution to the kinship verification problem. We recognize child-parent relationships with reasonable accuracies as suggested by its comparison with different methods on the same data and the same experimental protocols. We further show the good generalization capabilities of our method in several cross-database experiments. In particular, a proposed approach that utilizes feature level fusion by combining features of different natures has been found effective in boosting the discriminative power of individual attributes by gathering, after a proper selection, different and complementary cues that helps better characterizing the analyzed kin samples.

Chapter 5

Kinship Verification in Unconstrained Environment

In this chapter we provide a further investigation to the verification of parent-child pairs using the extensive and challenging KinFaceW dataset, where images have no restrictions of pose, lighting, background, expression, age, ethnicity, and partial occlusion. In our approach, a set of textural descriptors, which characterize the faces from different and complementary perspectives, are first used to identify kinship clues on parent-child pairs and then combined together to obtain a more discriminative sample characterization. Then, we applied a robust feature selection algorithm to improve the classification accuracy of an SVM classifier.

The remainder of the chapter is organized as follows. Section 5.1 describes the KinFaceW database used in the experiments. Section 5.2 outlines our approach. Finally Section 5.3 presents and discusses the experimental results.

5.1 Image Database

In our experiment we used a public dataset, proposed as benchmark for the *The Kinship Verification in the Wild Evaluation*, held in conjunction with the FG2015 conference. This dataset is divided into two subsets, KinFaceW-I and KinFaceW-II. Their difference is that face images with a kin relationship were acquired from different photos in KinFaceW-I and from the same photo in KinFaceW-II.

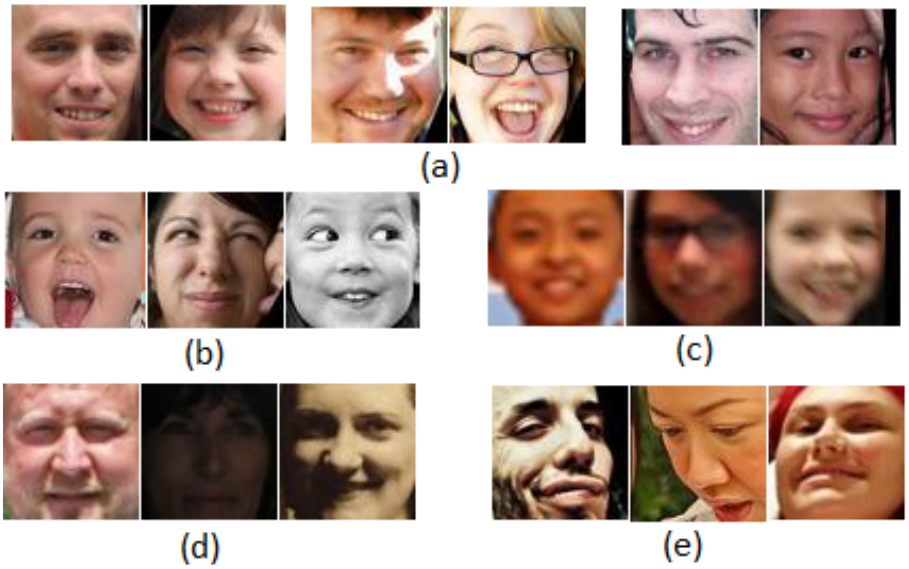


Figure 5.1: Examples of images in the KinFaceW database: father-daughter pairs (a), samples with varying expressions (b), defocused images (c), heterogeneity in illumination (d) and pose (e).

Each dataset consists of pairs of individuals characterized by parent-child relationships: Father-Son (F-S), Father-Daughter (F-D), Mother-Son (M-S), and Mother-Daughter (M-D). Specifically, there are 156, 134, 116, and 127 pairs of facial kinship images for these four relations in KinFaceW-I and 250 for each category in KinFaceW-II.

The faces have been aligned by manually labeling the eye corners. Then they were cropped and re-scaled to a resolution of 64×64 pixels. The dataset includes color and grayscale images showing different illumination conditions, poses and expressions. As can be seen in Figure 5.1, some of the images contain partial occlusions and artifacts such as blurring and defocusing. Furthermore, the majority of images are cropped such that relevant facial traits are discarded, such as chin and ears. These image characteristics make the analysis of this dataset a challenging task.

5.2 Proposed Method

Different representations of the information conveyed by faces have been experimented for tackling the various face image processing problems. Common facial

attributes for such tasks rely on geometric distances, local image descriptors or textural features. Previous studies have also shown that the discriminative power of facial attributes of different nature is likely to be improved by that of their integration [6].

As for the facial attributes, a pre-requisite for computing geometric attributes and local image descriptors is the proper identification of facial landmarks. Despite the good performances in dealing with extremely challenging images demonstrated by recent advances in automatic facial landmark identification ([27, 28, 29]), the accuracies of these methods applied to the KinFaceW images were in several cases far from satisfactory. A possible explanation is the inability to exploit the whole face information, due to the specific image cropping of most images in the dataset. Finally, it should be underlined that even if accurate landmarks were available, the very low image resolution would likely reduce the discriminative power of geometrical features.

Taking these observations into account, in our solution we focused on several textural descriptors which are capable of learning similarity between images and, thus, particularly suited to the problem tackled. Since these descriptors capture different and complementary information from the facial images, we designed a multi-perspective approach where the descriptors were first analyzed individually and then integrated in order to provide a more robust solution to the kinship recognition problem.

In the following, we will describe (i) the different type of features used to characterize each facial image, (ii) how they were used (individually or grouped) to characterize an image pair and (iii) our classification approach.

5.2.1 New set of features

Along with the use of Weber Local Descriptor (WLD) [15], see Section 4.3.2, the individual images have been characterized with the following attributes.

Local Phase Quantization. As described in Section 5.1, several images in KinFaceW appears blurred due to either subject motion during exposure or to out of focus camera lens. Blur can seriously affect the recognition and addressing this problem is sorely needed into this context.

Local Phase Quantization (LPQ) is an operator originally proposed for texture classification and later applied to face recognition tasks. LPQ achieved complementary performance to other descriptors on challenging face recognition datasets ([75, 78, 76]). This operator is insensitive to centrally symmetric blur (such as the

one caused by linear motion and out of focus) and is invariant to changes in the mean level of illumination and in illumination contrast.

LPQ exploits the blur invariance property of the phase spectrum and encodes phase information in a way similar to the coding mechanism of Local Binary Patterns (LBPs). LPQ codes are obtained by computing, in a $w \times w$ neighborhood of each image pixel, the phase of the 2D Short Term Fourier Transform (STFT). To maintain a compact representation, only the quantized phase of selected frequency components are extracted. Each quantized LPQ code is then encoded as an 8-bit digit.

In [76] the trade off between the discrimination power of the LPQ descriptor and its blur-tolerance is discussed. Decreasing the window size w helps capturing more detailed local information, but at the same time it reduces the descriptor tolerance to blur. To optimize this trade off and increase the effectiveness of LPQ based representation, a multi-scale approach can be applied by varying the local window size w. The LPQ image descriptor is finally obtained by dividing the image into non-overlapping blocks, aggregating for each block histograms of LPQ codes at different sizes and, finally, concatenating the obtained histograms. The resulting vector has size 768.

Patch based LBPs. Patch based LBPs is a family of novel image descriptors, recently introduced in [84, 83] to deal with face image pair-matching problems. These descriptors encode in a compact from the similarity between neighboring image patches and capture useful information complementary to that of pixel-based descriptors. The basic idea behind patch-based approaches is to detect local texture properties by analyzing the cross-correlation between a central patch (a window of known size) and its neighbouring patches. These properties are then encoded, as in standard LBP approaches, as a bit string. Patch-based LBPs histograms are capable of summarizing regions, lines, edges and complex textures in a unified way.

Two types of such descriptors were proposed in [83]: Three-Patch LBP (TPLBP) and Four-Patch LBP (FPLBP).

TPLBPs codes are computed per pixel and each bit of a code is obtained by comparing three different patches. First, S patches of size $w \times w$ are distributed uniformly along a ring of radius r centred on each pixel x. Then, given a parameter α, the value of a single code bit is set according to which of the two α-patches apart is more similar to the patch centred on x. Patch similarity is computed using the L_2-norm of their gray level differences. The resulting code is an S-bit string. The image signature is then computed by (i) dividing the image into a set of B non-overlapping blocks of equal size, (ii) computing a normalized histogram of code frequencies for each block and then (iii) concatenating the block histograms into a

single vector.

FBLBPs are similar to TBLBPs. In this case, S patches are uniformly distributed on each of two rings with radius r_1 and r_2, both centred at pixel x. A patch positioned on ring 1 and another one on ring 2 being α-patches apart, form a patch pair. Then, the value of a code bit is set according to which among two patch-pairs, symmetric with respect to the central pixel x, includes more similar patches. Thus, for S patches on each circle the number of symmetric patch-pairs (*i.e.* the code length) is $S/2$. The FBLBP image signature is then obtained using the same procedure described for TBLBP.

From these descriptions, it is clear that the strength of these descriptors may decrease increasing the ring radius, due to minimal correlation of the sampling points with the center pixel.

TPLBP results in a 3.072-size vector and a more compact 192-size vector is obtained with FPLBP.

5.2.2 Classification Approach

The outline of our classification approach is summarized in Figure 5.2. We follow a very similar approach as presented in Section 4.3. Differently, here we utilized only textural descriptors and their combinations to characterize the images in the kin-pairs. In addition, as describe in Section 5.1, the images are already normalized using eye-positions and are re-scaled to a fixed-size.

5.3 Results and discussion

The described classification approach has been applied to the dataset of facial image pairs introduced in Section 5.1 according to the *Image restricted* experimental protocol defined by the evaluation organizers.

In this protocol, for each data subset (KinFaceW-I and KinFaceW-II) and each kin category (F-D, F-S, M-D and M-S), the set of all positive (kin) pairs and a randomly generated set of negative (non-kin) pairs was provided. Positive and negative sets have the same size in all the cases. The pair datasets were then divided into 5 equally sized sets having a uniform distribution of positive and negative pairs. The classification is repeated 5 times and, for each round, one set is used for testing and the others for training. Again, the composition of the five sets was provided by the organizers in order to allow for a fair comparison between different approaches.

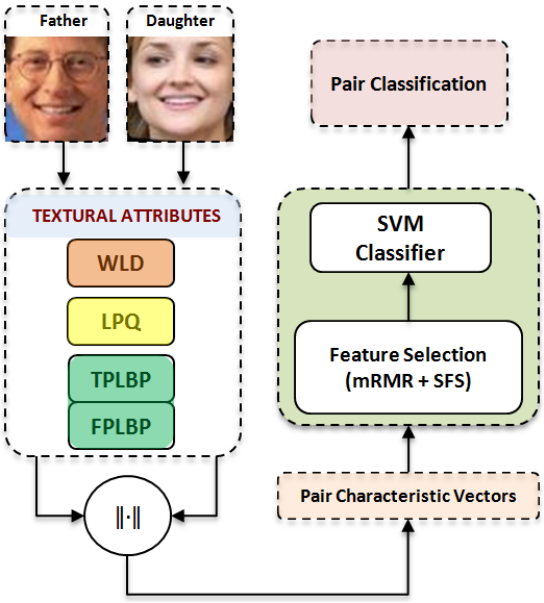


Figure 5.2: Outline of the classification algorithm.

Our experiments were aimed at validating the proposed multi-perspective approach. To this end, we first evaluated the classification based on the individual attributes and then that based on different combinations of them. When composing attribute groups, we considered patch-based descriptor as a unique attribute. That is, we analyzed the individual contribution of TPLBP and FPLBP to different groups, but we never included both of them into the same group.

In order to assess our results, we compared them with:

- two baseline methods (two variations of the Side-Information based Linear Discriminant Analysis, SILD, described in [80]) provided by the contest organizers;
- the recognition capabilities of a panel of human raters[1], available in [6]
- four methods submitted to the Kinship Verification in The Wild evaluation ([12]), namely:

[1]although these data were obtained with a slightly different experimental protocol, they are interesting since, as we will show in the following, they provide an indication of the superior capabilities of automatic approaches w.r.t. that of human observers

- LIRIS, a Metric Learning approach combining different features (LBP, HOG, Over-Complete LBP and Fisher Vector faces) and reducing the feature vector dimensionality to 100 exploiting whitened PCA;
- ULPGC, where multiple descriptors (HOG, LBP, Local Salient Patterns, Local Directional Patterns, LPQ and Local Oriented Statistics Information Booster) were computed at different resolutions, encoded into histograms and reduced with PCA to compute the final characteristic vectors; SVM was used as classifier;
- NUAA, exploiting a similarity learning approach relying on SIFT features computed on a regular grid; feature selection and SVM provided the final judgement;
- BIU, which proposed an asymmetric metric learning scheme with respect to kin similarity, where the age-related appearance differences between parents and children were handled with two different projection matrices for *old* (parents) and *young* (children) samples; samples were characterized with either LBP or HOG features and SVM was used for classification;

Experimental results are summarized in Table 5.1-5.2 and organized by attribute, or attribute group. The last lines of the tables show the accuracies of the human panel and of the other methods used for comparison. ROC curves are shown in 5.3 where, for our method and for the sake of clarity, we only report the combination of features providing the best results.

The following remarks can be drawn.

- In most of the cases, the TPLBP-LPQ-WLD group performs better on different kin-pairs and achieves the highest mean accuracies on both datasets (86.3% and 83.1% on KinFaceW-I and II, respectively).
- Grouped attributes perform generally better than their individual components. This indicates that samples are better characterized by the combination of different features.
- TPLBP performs slightly better than FPLBP, either individually or in combination with other attributes; this result can be explained considering that FPLBP is a more compact descriptor while TPLBP has a higher dimensional feature vector and, thus, it is likely to convey more subtle information;
- The performance of two baseline methods (71.1% and 74.9% on KinFaceW-I and 71.8% and 73.5% on KinFaceW-II) and of the human panel (63.8% and

Table 5.1: KinFace-I accuracies (%)

	F-D	F-S	M-D	M-S	Mean
FPLBP	75.1	75.0	74.6	75.1	75.0
TPLBP	76.9	76.9	75.7	78.9	77.1
LPQ	79.5	77.9	83.8	81.1	80.6
WLD	83.6	79.2	81.5	81.0	81.3
FPLBP-LPQ	80.6	78.6	83.1	84.6	81.7
TPLBP-LPQ	80.6	79.2	84.0	84.5	82.1
FPLBP-WLD	83.6	82.1	83.0	85.0	83.4
TPLBP-WLD	81.0	83.0	87.9	84.1	84.0
LPQ-WLD	84.3	78.9	88.2	85.4	84.2
FPLBP-LPQ-WLD	85.4	82.1	89.7	84.5	85.4
TPLBP-LPQ-WLD	85.8	85.3	86.7	87.5	86.3
SILD (LBP)	69.4	78.2	70.1	66.8	71.1
SILD (HOG)	72.4	80.5	77.1	69.8	74.9
LIRIS	83.0	80.6	82.3	85.0	82.7
ULPGC	71.3	70.9	58.5	80.9	70.0
NUAA	86.3	80.6	81.0	83.9	83.0
BIU	86.9	76.5	73.9	79.8	79.3
Human Panel	58.0	61.0	70.0	66.0	63.8

66.8%) are already improved by that obtained in our work with several individual attributes, and outperformed by our best results, achieved integrating the three different attribute types[2];

- As for the participants to the evaluation contest, our approach was the best performer in KinFaceW-I and the second in KinFaceW-II dataset, where LIRIS was the optimal one; in general our method showed global performances similar to the LIRIS methods, achieving the best average result on the two datasets (84.7% versus 84.3% of LIRIS).

Another interesting result, not shown in details for the sake of brevity, is that the various steps of our FS approach provide an effective improvement of the classification accuracies. For instance, in the optimal case (TPLBP-LPQ-WLD) the final accuracies are about 6% higher than that obtained with the first FS step (mRMR) and about 18% higher that that obtained without applying FS. Another result is

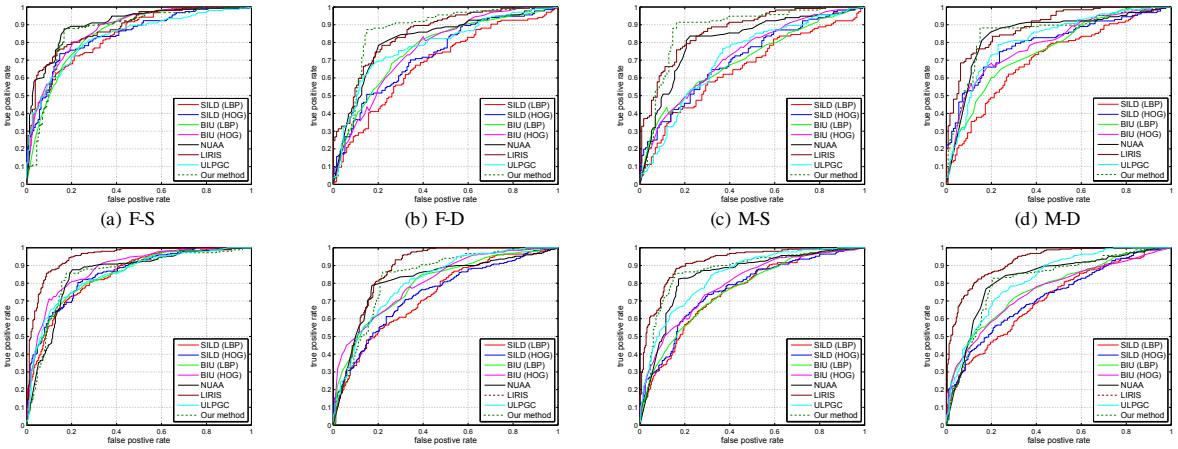
[2]It should be noted that all participants to the evaluation contest achieved the same result

Table 5.2: KinFace-II accuracies (%)

	F-D	F-S	M-D	M-S	Mean
FPLBP	73.8	67.0	69.4	70.0	70.1
TPLBP	70.0	73.6	68.8	77.2	72.4
LPQ	80.8	81.8	78.0	77.6	79.6
WLD	78.8	77.4	78.6	78.4	78.3
FPLBP-LPQ	80.8	79.6	78.2	78.6	79.3
TPLBP-LPQ	78.4	80.8	78.6	80.6	79.6
FPLBP-WLD	78.6	78.4	79.4	80.6	79.3
TPLBP-WLD	77.6	79.2	80.0	80.8	79.4
LPQ-WLD	82.6	81.0	80.0	82.0	81.4
FPLBP-LPQ-WLD	80.6	81.8	81.0	81.2	81.2
TPLBP-LPQ-WLD	82.2	84.0	81.2	84.8	83.1
SILD (LBP)	70.0	78.2	67.8	71.2	71.8
SILD (HOG)	71.6	79.6	69.6	73.2	73.5
LIRIS	89.4	83.6	86.2	85.0	86.0
ULPGC	85.4	75.6	75.6	81.6	80.0
NUAA	84.4	81.6	82.8	81.6	82.5
BIU	87.5	80.8	79.7	75.6	80.9
Human Panel	61.0	61.0	73.0	69.0	66.8

that, in attribute groups, the feature variable chosen by FS belongs, in different percentage, to all of the group components. For instance, the average percentage, over all dataset and kin groups, of variables per type in the optimal attribute combination is about 50% for WLD, 20% for TPLBP and 30% for LPQ.

Concluding, we have shown that our approach is capable of recognizing different type of child-parent relationships with reasonable accuracies and providing optimal performances with respect to other methods. In particular, the multi-perspective approach proposed has been found effective in strengthening the discriminative power of individual attributes by gathering, after a proper selection, different and complementary information that helps better characterizing the analyzed samples.



(a) F-S (b) F-D (c) M-S (d) M-D

(e) F-S (f) F-D (g) M-S (h) M-D

Figure 5.3: ROC curves for TBLPB-LPQ-WLD (our method), for different approaches and for the baselines provided by the event organizers (SILD). The top (a-d) and bottom (e-h) rows present the results obtained from the KinFaceW-I and KinFaceW-II, respectively. Each column shows the results on a given dataset, according to the kin relationship of the samples.

Chapter 6

Conclusion and Future Work

This thesis has presented the use of feature level fusion in two recent research applications aimed at improving the system classification accuracy together with the use of several pattern analysis and image processing techniques. In the literature match level and decision level fusion have been extensively studied, whereas feature level fusion is a relatively understudied problem because of the difficulties inherent to its correct implementation. Nevertheless, it is believed that fusing features at this stage would retain a richer source of discriminative information. If employed, subsequently, feature level fusion would need a careful reduction of features that would decide the most representative and discriminative feature sets for the patterns to classify. With this motivation we utilized feature level fusion for two recent pattern recognition problems, achieving significant level of performance as compared to other methods in the literature.

For a multi-class problem, HEp-2 staining pattern classification in IIF images, we devised an automated solution that benefit from feature level fusion and is based on Subclass Discriminant Analysis approach. The main findings of our work are two. First, after investigating the contribution of several types of attributes, individually and then grouped at feature level, we obtained an optimal set of features that combines morphological descriptors with global and local textural analysis based on, respectively, GLCM and CoALBP features. With this set of attributes we achieved classification accuracy close to 90%, showing that the integration of information of different nature provides an effective representation of the fluorescence patterns. Furthermore, the classification accuracy was stable in a neighborhood of the optimal set size, indicating a good robustness of the proposed approach. Second, we verified that the proposed SDA-based strategy can effectively cope with the high within-class variance of HEp-2 samples and, coupled with a Feature Selection process, can provide a compact set of image attributes that is highly discriminative and effective

for the classification task.

Similarly in a two class problem, we presented an approach for the computer identification of parent-child pairs from the facial images of two individuals. Individual images were characterized by different facial attributes, related to geometric, local and global textural data. These attributes were first analyzed individually and then fused together at feature level to build, with the contribution of a robust two step feature selection algorithm, different SVM classifiers. Simulation results show that the integration of features of different natures is indeed an effective solution to the automatic kinship verification problem. Our approach achieved, on the same data, significantly higher accuracies than those obtained by panel of human raters and previous approaches in the literature. We have also shown the good generalization capabilities of our approach through challenging cross-database experimental settings.

In another work, we addressed the problem of kinship verification in an unconstrained environment. We adopt a similar approach and use a dataset where the photographs contain a large heterogeneity, with unconstrained face orientations, expressions and occlusions. Here, our multi-perspective strategy consists in considering different textural features and extracting the variables that are more relevant to the problem using feature selection. The results of our experiments suggests that the combinations of different types of attributes, capable of describing the data under different perspectives, is a strategy to better characterize the samples and, ultimately, to improve the classification accuracy. We have also shown, through the comparison with other methods on the same data, that our approach is optimal and, furthermore, capable of largely outperforming the human capabilities of recognizing kin.

All relevant results presented in this thesis are published in [55, 85, 10, 11, 12]

As future works, we would like to improve our proposed solutions for the current problems. We are planning to implement a complete solution tackling all the steps of an automated IIF analysis, including the quantification of the staining intensity, the cell segmentation and the recognition of the mitotic cells. Further improvements of the SDA-based solution will be also investigated, *e.g.* by experimenting different types of classifiers other than k-NN. We are also planning to address the multi-class problem of simultaneously identifying different degrees of kinship (e.g., parent-child, siblings and grandparent-grandchild), a problem which has not yet been tackled in the literature. As suggested in Section 4.4.2, we will also investigate methods capable of effectively coping with the high within-class variance of the kin samples. Another point of interest is to fully analyze how factors such as ethnicity, expressions, gender and age affect a kinship predictor, and possible approaches to alleviate such influences. Recently a work [30] introduced the problem of kinship verification from

dynamic expressions, which opens a new dimension worth to be explored.

Despite the possibility for a range of future extensions to our work, we think that a multi-perspective approach through feature level fusion is relevant and should be deeply analyzed when addressing these novel areas of research. Having, said that similar approaches based on feature level fusion can also be developed for other related and recent applications, including, but not limited to Fingerprint Liveness detection [138] and 2D face Anti-Spoofing [137]. More interestingly, data fusion at feature level for computer vision and pattern recognition applications can also be integrated within other recent frameworks e.g. multi-task learning, sparse coding and deep learning and will be a focus of our future research work.

Bibliography

[1] Gwenaël Kaminski, Slimane Dridi, Christian Graff, and Edouard Gentaz. Human ability to detect kinship in strangers' faces: effects of the degree of relatedness. *Proc Biol Sci*, 276(1670):3193–3200, Sep 2009.

[2] Maria F Dal Martello and Laurence T Maloney. Where are kin recognition signals in the human face? *Journal of Vision*, 6(12):1356–66, 2006.

[3] Ruogu Fang, Kevin D. Tang, Noah Snavely, and Tsuhan Chen. Towards computational models of kinship verification. In *Image Processing (ICIP), 2010 17th IEEE International Conference on*, pages 1577–1580. IEEE, September 2010.

[4] Gowri Somanath and Chandra Kambhamettu. Can faces verify blood-relations? In *IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, 2012.

[5] Guodong Guo and Xiaolong Wang. Kinship measurement on salient facial features. *Instrumentation and Measurement, IEEE Transactions on*, 61(8):2322–2325, Aug 2012.

[6] Jiwen Lu, Xiuzhuang Zhou, Yap-Pen Tan, Yuanyuan Shang, and Jie Zhou. Neighborhood repulsed metric learning for kinship verification. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 36(2):331–345, Feb 2014.

[7] Siyu Xia, Ming Shao, Jiebo Luo, and Yun Fu. Understanding kin relationships in a photo. *Multimedia, IEEE Transactions on*, 14(4):1046–1056, Aug 2012.

[8] Tiago Figueiredo Vieira, Andrea Bottino, Aldo Laurentini, and Matteo De Simone. Detecting Siblings in Image Pairs. *The Visual Computer*, pages 1–13, 2013.

[9] N. Kohli, R. Singh, and M. Vatsa. Self-similarity representation of weber faces for kinship classification. In *Biometrics: Theory, Applications and Systems (BTAS), 2012 IEEE Fifth International Conference on*, pages 245–250, Sept 2012.

[10] TiagoF. Vieira, Andrea Bottino, and IhteshamUl Islam. Automatic verification of parent-child pairs from face images. In JosÃĺ Ruiz-Shulcloper and Gabriella

Sanniti di Baja, editors, *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, volume 8259 of *Lecture Notes in Computer Science*, pages 326–333. Springer Berlin Heidelberg, 2013.

[11] Andrea Bottino, TiagoF. Vieira and Ihtesham Ul Islam. Geometric and Textural Cues for Automatic Kinship Verification. *International JournaL of Pattern Recognition and Artificial Intelligence*, (IN PRESS).

[12] Jiwen Lu et al. The fg 2015 kinship verification in the wild evaluation. In *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition*, 2015.

[13] T F Cootes, C J Taylor, D H Cooper, and J Graham. Active shape models\âĂŤtheir training and application. *Comput. Vis. Image Underst.*, 61(1):38–59, January 1995.

[14] Stephen Milborrow. Active Shape Models with Stasm. http://www.milbo.users.sonic.net/stasm/.

[15] Jie Chen, Shiguang Shan, Chu He, Guoying Zhao, M. PietikaÂĺinen, Xilin Chen, and Wen Gao. Wld: A robust local image descriptor. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(9):1705 –1720, sept. 2010.

[16] David G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2):91–110, November 2004.

[17] Koen E a van de Sande, Theo Gevers, and Cees G M Snoek. Evaluating color descriptors for object and scene recognition. *IEEE transactions on pattern analysis and machine intelligence*, 32(9):1582–96, September 2010.

[18] a.E. Abdel-Hakim and a.a. Farag. CSIFT: A SIFT Descriptor with Color Invariant Characteristics. *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2 (CVPR'06)*, 2:1978–1983, 2006.

[19] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995.

[20] Chih-Chung Chang and Chih-Jen Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011.

[21] Chih-Wei Hsu, Chih-Chung Chang, and Chih-Jen Lin. A practical guide to support vector classification. Technical report, Department of Computer Science, National Taiwan University, 2003.

[22] L Wolf and S Bileschi. Combining variable selection with dimensionality reduction. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 2, pages 801 – 806 vol. 2, June 2005.

[23] Andrew Y. Ng. Feature selection, L1 vs. L2 regularization, and rotational invariance. In *Proceedings of the twenty-first international conference on Machine learning*, ICML '04, pages 78+, New York, NY, USA, 2004. ACM.

[24] A W Whitney. A Direct Method of Nonparametric Measurement Selection. *Computers, IEEE Transactions on*, C-20(9):1100–1103, 1971.

[25] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, Oct. 2001.

[26] Cognitec. FACEVacs SDK. http://www.cognitec.com/facevacs-sdk.html.

[27] A. Asthana, S. Zafeiriou, Shiyang Cheng, and M. Pantic. Robust discriminative response map fitting with constrained local models. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 3444–3451, June 2013.

[28] B. Martinez, M.F. Valstar, X. Binefa, and M. Pantic. Local evidence aggregation for regression-based facial point detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(5):1149–1163, May 2013.

[29] TimF. Cootes, MirceaC. Ionita, Claudia Lindner, and Patrick Sauer. Robust and accurate shape model fitting using random forest regression voting. In Andrew Fitzgibbon, Svetlana Lazebnik, Pietro Perona, Yoichi Sato, and Cordelia Schmid, editors, *Computer Vision âĂŞ ECCV 2012*, volume 7578 of *Lecture Notes in Computer Science*, pages 278–291. Springer Berlin Heidelberg, 2012.

[30] H. Dibeklioglu, A. A. Salah, and T. Gevers. Like father, like son: Facial expression dynamics for kinship verification. In *IEEE International Conference on Computer Vision*, pages 1497–1504, 2013.

[31] Satosi Watanabe. *Pattern recognition: human and mechanical*. John Wiley & Sons, Inc., 1985.

[32] Julia Vogel, Adrian Schwaninger, Christian Wallraven, and Heinrich H Bülthoff. Categorization of natural scenes: local vs. global information. In *Proceedings of the 3rd symposium on Applied perception in graphics and visualization*, pages 33–40. ACM, 2006.

[33] Arun Ross and Anil Jain. *Multimodal biometrics: An overview*. In *Proceedings of 12th Signal Processing Conference (EUSIPCO)*, pages 1221–1224. 2004.

[34] Keiji Tanaka, Hide-aki Saito, Yoshiro Fukada, and Madoka Moriya. Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *J Neurophysiol*, 66(1):170–189, 1991.

[35] K. Egerer, D. Roggenbuck, R. Hiemann, M. Weyer, T. Buettner, B. Radau, R. Krause, B. Lehmann, E. Feist, G. Burmester, Automated evaluation of autoantibodies on human epithelial-2 cells as an approach to standardize cell-based immunofluorescence tests., Arthritis Res Ther 12 (2) (2010) R40.

[36] N. Bizzaro, R. Tozzoli, E. Tonutti, A. Piazza, F. Manoni, A. Ghirardello, D. Bassetti, D. Villalta, M. Pradella, P. Rizzotti, Variability between methods to determine ana, anti-dsdna and anti-ena autoantibodies: a collaborative study with the biomedical industry, Journal of Immunological Methods 219 (1ÃćâĆňâĂIJ2) (1998) 99 – 107.

[37] P. Foggia, G. Percannella, P. Soda, M. Vento, Benchmarking HEp-2 Cells Classification Methods, Medical Imaging, IEEE Transactions on, (2013).

[38] Y.-L. Huang, Y.-L. Jao, T.-Y. Hsieh, C.-W. Chung, Adaptive automatic segmentation of hep-2 cells in indirect immunofluorescence images, in: Proceedings

of the 2008 IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing (sutc 2008), SUTC '08, 2008, pp. 418–422.

[39] C. Creemers, K. Guerti, S. Geerts, K. Van Cotthem, A. Ledda, V. Spruyt, Hep-2 cell pattern segmentation for the support of autoimmune disease diagnosis, in: Proceedings of the 4th International Symposium on Applied Sciences in Biomedical and Communication Technologies, ISABEL '11, 2011, pp. 28:1–28:5.

[40] G. Percannella, P. Soda, M. Vento, A classification-based approach to segment hep-2 cells, in: Computer-Based Medical Systems (CBMS), 2012 25th International Symposium on, 2012, pp. 1–5.

[41] P. Foggia, G. Percannella, P. Soda, M. Vento, Early experiences in mitotic cells recognition on hep-2 slides, in: Computer-Based Medical Systems (CBMS), 2010 IEEE 23rd International Symposium on, 2010, pp. 38–43.

[42] G. Percannella, P. Soda, M. Vento, Mitotic hep-2 cells recognition under class skew, in: Image Analysis and Processing âĂŞ ICIAP 2011, Vol. 6979 of Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2011, pp. 353–362.

[43] P. Soda, G. Iannello, M. Vento, A multiple expert system for classifying fluorescent intensity in antinuclear autoantibodies analysis, Pattern Anal. Appl. 12 (3) (2009) 215–226.

[44] T.-Y. Hsieh, Y.-C. Huang, C.-W. Chung, Y.-L. Huang, Hep-2 cell classification in indirect immunofluorescence image, in: Proceedings of the 7th international conference on Information, communications and signal processing, ICICS'09, 2009, pp. 211–214.

[45] P. Perner, H. Perner, B. MÃijller, Mining knowledge for hep-2 cell image classification, Artificial Intelligence in Medicine 26 (2002) 161–173.

[46] U. Sack, S. Knoechner, H. Warschkau, U. Pigla, F. Emmrich, M. Kamprad, Computer-assisted classification of hep-2 immunofluorescence patterns in autoimmune diagnostics, Autoimmunity Reviews 2 (5) (2003) 298 – 304.

[47] P. Soda, G. Iannello, A hybrid multi-expert systems for hep-2 staining pattern classification, in: Proceedings of the 14th International Conference on Image Analysis and Processing, 2007, pp. 685–690.

[48] Y.-C. Huang, T.-Y. Hsieh, C.-Y. Chang, W.-T. Cheng, Y.-C. Lin, Y.-L. Huang, Hep-2 cell images classification based on textural and statistic features using self-organizing map, in: Proceedings of the 4th Asian conference on Intelligent Information and Database Systems - Volume Part II, ACIIDS'12, 2012, pp. 529–538.

[49] S. Di Cataldo, A. Bottino, E. Ficarra, E. Macii, Applying textural features to the classification of hep-2 cell patterns in iif images, in: Proceedings of the 21st International Conference on Pattern Recognition, ICPR 2012, 2012.

[50] P. Strandmark, J. Ulen, F. Kahl, Hep-2 staining pattern classification, in: Pattern Recognition (ICPR), 2012 21st International Conference on, Nov., pp.

33–36.

[51] I. Theodorakopoulos, D. Kastaniotis, G. Economou, S. Fotopoulos, Hep-2 cells classification via fusion of morphological and textural features., in: IEEE 12th International Conference on Bioinformatics and Bioengineering (BIBE), 2012, pp. 689–694.

[52] G. Thibault, J. Angulo, Efficient statistical/morphological cell texture characterization and classification, in: Pattern Recognition (ICPR), 2012 21st International Conference on, 2012, pp. 2440 –2443.

[53] http://mivia.unisa.it/datasets/biomedical-image-datasets/hep2-image-dataset/(Last accessed: January 2015).

[54] M. Zhu, A. M. Martinez, Subclass discriminant analysis, Pattern Analysis and Machine Intelligence, IEEE Transactions on 28.8 (2006): 1274-1286.

[55] I. Islam, S. Di Cataldo, A. Bottino, E. Ficarra, E. Macii, Classification of hep-2 staining patterns in immunofluorescence images - comparison of support vector machines and subclass discriminant analysis strategies, in: Proceedings of the International Conference on Bioinformatics Models, Methods and Algorithms, 2013, pp. 1–9.

[56] R. Tozzoli, N. Bizzaro, E. Tonutti, D. Villalta, D. Bassetti, F. Manoni, A. Piazza, M. Pradella, P. Rizzotti, Guidelines for the laboratory use of autoantibody tests in the diagnosis and monitoring of autoimmune rheumatic diseases., Am J Clin Pathol 117 (2) (2002) 316–24.

[57] Y. Xu, S. B. Huang, H. Ji, Integrating local feature and global statistics for texture analysis, in: Image Processing (ICIP), 2009 16th IEEE International Conference on, 2009, pp. 1377 –1380.

[58] J. A. Montoya-Zegarra, J. Beeck, N. Leite, R. Torres, A. Falcão, Combining global with local texture information for image retrieval applications, in: Proceedings of the 2008 Tenth IEEE International Symposium on Multimedia, ISM '08, 2008, pp. 148–153.

[59] R. M. Haralick, K. Shanmugam, I. Dinstein, Textural Features for Image Classification, Systems, Man and Cybernetics, IEEE Transactions on SMC-3 (6) (1973) 610–621.

[60] L. kiat Soh, C. Tsatsoulis, Texture analysis of sar sea ice imagery using gray level co-occurrence matrices, IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (1999) 780–795.

[61] D. Clausi, An analysis of co-occurrence texture statistics as a function of grey level quantization, Can. J. Remote Sensing 1 (28) (2002) 45–62.

[62] A. Pinheiro, Image descriptors based on the edge orientation, in: Semantic Media Adaptation and Personalization, 2009. SMAP '09. 4th International Workshop on, 2009, pp. 73–78.

[63] F. Bianconi, A. Fernández, A. Mancini, Assessment of rotation-invariant texture classification through Gabor filters and discrete Fourier transform, in:

Proceedings of the 20th International Congress on Graphical Engineering (XX INGEGRAF), Valencia, Spain, 2008.

[64] D.-G. Sim, H.-K. Kim, R.-H. Park, Invariant texture retrieval using modified Zernike moments, Image and Vision Computing 22 (4) (2004) 331–342.

[65] T. Ojala, M. Pietikäinen, T. Mäenpää, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Trans. Pattern Anal. Mach. Intell. 24 (7) (2002) 971–987.

[66] Z. Guo, D. Zhang, D. Zhang, A completed modeling of local binary pattern operator for texture classification, Image Processing, IEEE Transactions on 19 (6) (June) 1657–1663.

[67] R. Nosaka, Y. Ohkawa, K. Fukui, Feature extraction based on co-occurrence of adjacent local binary patterns., in: Lecture Notes in Computer Science, Vol. 7088, Springer, 2011, pp. 82–91.

[68] R. Nosaka, C. H. Suryanto, K. Fukui, Rotation invariant co-occurrence among adjacent lbps, in: Proceedings of the ACCV2012 Workshop LBP2012, 2012, pp. 1–11.

[69] N. Otsu, A threshold selection method from gray-level histograms, IEEE Transactions on Systems, Man and Cybernetics 9 (1) (1979) 62–66.

[70] T. Ojala, M. Pietikäinen, D. Harwood, A comparative study of texture measures with classification based on featured distributions, Pattern Recognition 29 (1) (1996) 51–59.

[71] N. Boulgouris, K. N. Plataniotis, E. Micheli-Tzanakou, Discriminant analysis for dimensionality reduction: An overview of recent developments, Biometrics: Theory, Methods, and Applications, Wiley.

[72] A. M. Martinez, M. Zhu, Where are linear feature extraction methods applicable?, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (12) (2005) 1934–1944.

[73] N. Gkalelis, V. Mezaris, I. Kompatsiaris, Mixture subclass discriminant analysis, IEEE Signal Process. Lett. 18 (5) (2011) 319–322.

[74] H. Peng, F. Long, C. Ding, Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (2005) 1226–1238.

[75] T. Ahonen, E. Rahtu, V. Ojansivu, and J. Heikkila. Recognition of blurred faces using local phase quantization. In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, pages 1–4. IEEE, 2008.

[76] C. H. Chan, M. A. Tahir, J. Kittler, and M. Pietikainen. Multiscale local phase quantization for robust component-based face recognition using kernel fusion of multiple descriptors. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(5):1164–1177, 2013.

[77] J. Chen, S. Shan, C. He, G. Zhao, M. Pietikainen, X. Chen, and W. Gao. Wld: A robust local image descriptor. *IEEE Transactions on Pattern Analysis and*

Machine Intelligence, 32(9):1705–1720, 2010.

[78] J. Heikkila and V. Ojansivu. Methods for local phase quantization in blur-insensitive image analysis. In *Local and Non-Local Approximation in Image Processing, 2009. LNLA 2009. International Workshop on*, pages 104–111. IEEE, 2009.

[79] S. Li, D. Gong, and Y. Yuan. Face recognition using weber local descriptors. *Neurocomputing*, 122(0):272–283, 2013.

[80] D. X. Meina Kan, Shiguang Shan and X. Chen. Side-information based linear discriminant analysis for face recognition. In *Proceedings of the British Machine Vision Conference*, pages 125.1–125.0. BMVA Press, 2011. http://dx.doi.org/10.5244/C.25.125.

[81] P. Pudil, J. NovoviÄĎovÃą, and J. Kittler. Floating search methods in feature selection. *Pattern Recognition Letters*, 15(11):1119 – 1125, 1994.

[82] G. Somanath, M. V. Rohith, and C. Kambhamettu. Vadana: A dense dataset for facial image analysis. In *ICCV Workshops*, pages 2175–2182. IEEE, 2011.

[83] L. Wolf, T. Hassner, and Y. Taigman. Effective unconstrained face recognition by combining multiple descriptors and learned background statistics. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(10):1978–1990, 2011.

[84] L. Wolf, T. Hassner, Y. Taigman, et al. Descriptor based methods in the wild. In *Workshop on Faces in'Real-Life'Images: Detection, Alignment, and Recognition*, 2008.

[85] Santa Di Cataldo, Andrea Bottino, Ihtesham Ul Islam, Tiago Figueiredo Vieira, and Elisa Ficarra. Subclass discriminant analysis of morphological and textural features for hep-2 staining pattern classification. *Pattern Recognition*, 47(7):2389–2399, 2014.

[86] Masoud Faraki, Mehrtash T Harandi, Arnold Wiliem, and Brian C Lovell. Fisher tensors for classifying human epithelial cells. *Pattern Recognition*, 47(7):2348–2359, 2014.

[87] Pasquale Foggia, Gennaro Percannella, Alessia Saggese, and Mario Vento. Pattern recognition in stained hep-2 cells: Where are we now? *Pattern Recognition*, 47(7):2305–2314, 2014.

[88] Xiangfei Kong, Kuan Li, Jingjing Cao, Qingxiong Yang, and Liu Wenyin. Hep-2 cell pattern classification with discriminative dictionary learning. *Pattern Recognition*, 47(7):2379–2388, 2014.

[89] Lingqiao Liu and Lei Wang. Hep-2 cell image classification with multiple linear descriptors. *Pattern Recognition*, 47(7):2400–2408, 2014.

[90] Ryusuke Nosaka and Kazuhiro Fukui. Hep-2 cell classification using rotation invariant co-occurrence among local binary patterns. *Pattern Recognition*, 47(7):2428–2436, 2014.

[91] Gennady V Ponomarev, Vladimir L Arlazarov, Mikhail S Gelfand, and Marat D Kazanov. Ana hep-2 cells image classification using number, size, shape and localization of targeted cell regions. *Pattern Recognition*, 47(7):2360–2366, 2014.

[92] Linlin Shen, Jiaming Lin, Shengyin Wu, and Shiqi Yu. Hep-2 image classification using intensity order pooling based features and bag of words. *Pattern Recognition*, 47(7):2419–2427, 2014.

[93] V Snell, W Christmas, and Josef Kittler. Hep-2 fluorescence pattern classification. *Pattern Recognition*, 47(7):2338–2347, 2014.

[94] Roman Stoklasa, Tomáš Majtner, and David Svoboda. Efficient< i> k</i>-nn based hep-2 cells classifier. *Pattern Recognition*, 47(7):2409–2418, 2014.

[95] Ilias Theodorakopoulos, Dimitris Kastaniotis, George Economou, and Spiros Fotopoulos. Hep-2 cells classification via sparse representation of textural features fused into dissimilarity space. *Pattern Recognition*, 47(7):2367–2378, 2014.

[96] Arnold Wiliem, Conrad Sanderson, Yongkang Wong, Peter Hobson, Rodney F Minchin, and Brian C Lovell. Automatic classification of human epithelial type 2 cell indirect immunofluorescence images using cell pyramid matching. *Pattern Recognition*, 47(7):2315–2324, 2014.

[97] Yan Yang, Arnold Wiliem, Azadeh Alavi, Brian C Lovell, and Peter Hobson. Visual learning and classification of human epithelial type 2 cell images through spontaneous activity patterns. *Pattern Recognition*, 47(7):2325–2337, 2014.

[98] Yair Barniv and David Casasent. Multisensor image registration: Experimental verification. In *25th Annual Technical Symposium*, pages 160–171. International Society for Optics and Photonics, 1981.

[99] LF Pau. Fusion of multisensor data in pattern recognition. In *Pattern recognition theory and applications*, pages 189–201. Springer, 1982.

[100] Robert R Tenney and Nils R Sandell Jr. Detection with distributed sensors. *IEEE Transactions on Aerospace Electronic Systems*, 17:501–510, 1981.

[101] Arun Ross and Anil Jain. Information fusion in biometrics. *Pattern recognition letters*, 24(13):2115–2125, 2003.

[102] David Lee Hall and Sonya AH McMullen. *Mathematical techniques in multisensor data fusion*. Artech House, 2004.

[103] Josef Kittler, Jiri Matas, Kenneth Jonsson, and MU Ramos Sánchez. Combining evidence in personal identity verification systems. *Pattern Recognition Letters*, 18(9):845–852, 1997.

[104] Dirk Borghys, Michal Shimoni, Gregory Degueldre, and Christiaan Perneel. Improved object recognition by fusion of hyperspectral and sar data. In *5th EARSeL SIG IS workshop on Imaging Spectroscopy:" Innovation in environmental research*, 2007.

[105] Vassilios Chatzis, Adrian G Bors, and Ioannis Pitas. Multimodal decision-level fusion for person authentication. *Systems, Man and Cybernetics, Part A:*

Systems and Humans, IEEE Transactions on, 29(6):674–680, 1999.

[106] Samuel Foucher, Mickaël Germain, Jean-Marc Boucher, and Goze B Benie. Multisource classification using icm and dempster-shafer theory. *IEEE Transactions on Instrumentation and Measurement*, 51(2):277–281, 2002.

[107] Anil K Jain, Lin Hong, and Yatin Kulkarni. A multimodal biometric system using fingerprint, face and speech. In *Proceedings of 2nd Int'l Conference on Audio-and Video-based Biometric Person Authentication, Washington DC*, pages 182–187, 1999.

[108] Zhiming Liu and Chengjun Liu. Fusion of color, local spatial and global frequency information for face recognition. *Pattern Recognition*, 43(8):2882–2890, 2010.

[109] Xiaoguang Lu, Yunhong Wang, and Anil K Jain. Combining classifiers for face recognition. In *Multimedia and Expo, 2003. ICME'03. Proceedings. 2003 International Conference on*, volume 3, pages III–13. IEEE, 2003.

[110] Yu Su, Shiguang Shan, Xilin Chen, and Wen Gao. Hierarchical ensemble of global and local classifiers for face recognition. *Image Processing, IEEE Transactions on*, 18(8):1885–1896, 2009.

[111] Kyong I Chang, Kevin W Bowyer, Patrick J Flynn, and Xin Chen. Multibiometrics using facial appearance, shape and temperature. In *Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on*, pages 43–48. IEEE, 2004.

[112] Syed MS Islam, Mohammed Bennamoun, Ajmal S Mian, and Rowan Davies. Score level fusion of ear and face local 3d features for fast and expression-invariant human recognition. In *Image Analysis and Recognition*, pages 387–396. Springer, 2009.

[113] Anil Jain, Karthik Nandakumar, and Arun Ross. Score normalization in multimodal biometric systems. *Pattern recognition*, 38(12):2270–2285, 2005.

[114] Zhiming Liu and Chengjun Liu. Fusion of the complementary discrete cosine features in the *YIQ* color space for face recognition. *Computer Vision and Image Understanding*, 111(3):249–262, 2008.

[115] Ajmal Mian, Mohammed Bennamoun, and Robyn Owens. 2d and 3d multimodal hybrid face recognition. In *Computer Vision–ECCV 2006*, pages 344–355. Springer, 2006.

[116] Damon L Woodard, Timothy C Faltemier, Ping Yan, Patrick J Flynn, and Kevin W Bowyer. A comparison of 3d biometric modalities. In *Computer Vision and Pattern Recognition Workshop, 2006. CVPRW'06. Conference on*, pages 57–57. IEEE, 2006.

[117] Xiaona Xu and Zhichun Mu. Multimodal recognition based on fusion of ear and profile face. In *Image and Graphics, 2007. ICIG 2007. Fourth International Conference on*, pages 598–603. IEEE, 2007.

[118] Xiaoli Zhou and Bir Bhanu. Integrating face and gait for human recognition. In *Computer Vision and Pattern Recognition Workshop, 2006. CVPRW'06. Conference on*, pages 55–55. IEEE, 2006.

[119] Arun A Ross and Rohin Govindarajan. Feature level fusion of hand and face biometrics. In *Defense and Security*, pages 196–204. International Society for Optics and Photonics, 2005.

[120] Xiao-Na Xu, Zhi-Chun Mu, and Li Yuan. Feature-level fusion method based on kfda for multimodal recognition fusing ear and profile face. In *Wavelet Analysis and Pattern Recognition, 2007. ICWAPR'07. International Conference on*, volume 3, pages 1306–1310. IEEE, 2007.

[121] Juan E Tapia and Claudio Andrés Pérez Flores. Gender classification based on fusion of different spatial scale features selected by mutual information from histogram of lbp, intensity, and shape. 2013.

[122] Yun Fu, Liangliang Cao, Guodong Guo, and Thomas S Huang. Multiple feature fusion by subspace learning. In *Proceedings of the 2008 international conference on Content-based image and video retrieval*, pages 127–134. ACM, 2008.

[123] Karthik Nandakumar. *Multibiometric Systems: Fusion Strategies and Template Security*. PhD thesis, Michigan State University, 2008.

[124] Lawrence Sirovich and Michael Kirby. Low-dimensional procedure for the characterization of human faces. *JOSA A*, 4(3):519–524, 1987.

[125] Martin Lades, Jan C Vorbruggen, Joachim Buhmann, Jörg Lange, Christoph von der Malsburg, Rolf P Wurtz, and Wolfgang Konen. Distortion invariant object recognition in the dynamic link architecture. *Computers, IEEE Transactions on*, 42(3):300–311, 1993.

[126] Ron Kohavi and George H John. Wrappers for feature subset selection. *Artificial intelligence*, 97(1):273–324, 1997.

[127] Nojun Kwak and Chong-Ho Choi. Input feature selection by mutual information based on parzen window. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(12):1667–1671, 2002.

[128] Momiao Xiong, Xiangzhong Fang, and Jinying Zhao. Biomarker identification by feature wrappers. *Genome Research*, 11(11):1878–1887, 2001.

[129] Włodzisław Duch, Tadeusz Wieczorek, Jacek Biesiada, and Marcin Blachnik. Comparison of feature ranking methods based on information entropy. In *Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on*, volume 2, pages 1415–1419. IEEE, 2004.

[130] Włodzisław Duch, Tomasz Winiarski, Jacek Biesiada, and Adam Kachel. Feature ranking, selection and discretization. In *Proceedings of Int. Conf. on Artificial Neural Networks (ICANN)*, pages 251–254, 2003.

[131] André Bleau and L Joshua Leon. Watershed-based segmentation and region merging. *Computer Vision and Image Understanding*, 77(3):317–370, 2000.

[132] M Usaj, D Torkar, M Kanduser, and D Miklavcic. Cell counting tool parameters optimization approach for electroporation efficiency determination of attached cells in phase contrast images. *Journal of microscopy*, 241(3):303–314, 2011.

[133] Guanglei Xiong, Xiaobo Zhou, and Liang Ji. Automated segmentation of drosophila rnai fluorescence cellular images using deformable models. *Circuits and Systems I: Regular Papers, IEEE Transactions on*, 53(11):2415–2424, 2006.

[134] Tiago Esteves, Pedro Quelhas, Ana Maria Mendonça, and Aurélio Campilho. Gradient convergence filters and a phase congruency approach for in vivo cell nuclei detection. *Machine Vision and Applications*, 23(4):623–638, 2012.

[135] Faisal R Al-Osaimi, Mohammed Bennamoun, and A Mian. Integration of local and global geometrical cues for 3d face recognition. *Pattern Recognition*, 41(3):1030–1040, 2008.

[136] Shufu Xie, Shiguang Shan, Xilin Chen, and Jie Chen. Fusing local patterns of gabor magnitude and phase for face recognition. *Image Processing, IEEE Transactions on*, 19(5):1349–1361, 2010.

[137] Ivana Chingovska, Jinwei Yang, Zhen Lei, Dong Yi, Stan Z Li, O Kahm, Christian Glaser, N Damer, Arjan Kuijper, Alexander Nouak, et al. The 2nd competition on counter measures to 2d face spoofing attacks. In *Biometrics (ICB), 2013 International Conference on*, pages 1–6. IEEE, 2013.

[138] Diego Gragnaniello, Giovanni Poggi, Carlo Sansone, and Luisa Verdoliva. Local contrast phase descriptor for fingerprint liveness detection. *Pattern Recognition*, 48(4):1050–1058, 2015.