Summer 8-31-2011

# Investigation on advanced image search techniques

Abhishek Verma

*New Jersey Institute of Technology*

## Recommended Citation

Verma, Abhishek, "Investigation on advanced image search techniques" (2011). *Dissertations*. 283.
https://digitalcommons.njit.edu/dissertations/283

# ABSTRACT

# INVESTIGATION ON ADVANCED IMAGE SEARCH TECHNIQUES

by
**Abhishek Verma**

Content-based image search for retrieval of images based on the similarity in their visual contents, such as color, texture, and shape, to a query image is an active research area due to its broad applications. Color, for example, provides powerful information for image search and classification. This dissertation investigates advanced image search techniques and presents new color descriptors for image search and classification and robust image enhancement and segmentation methods for iris recognition.

First, several new color descriptors have been developed for color image search. Specifically, a new oRGB-SIFT descriptor, which integrates the oRGB color space and the Scale-Invariant Feature Transform (SIFT), is proposed for image search and classification. The oRGB-SIFT descriptor is further integrated with other color SIFT features to produce the novel Color SIFT Fusion (CSF), the Color Grayscale SIFT Fusion (CGSF), and the CGSF+PHOG descriptors for image category search with applications to biometrics. Image classification is implemented using a novel EFM-KNN classifier, which combines the Enhanced Fisher Model (EFM) and the K Nearest Neighbor (KNN) decision rule. Experimental results on four large scale, grand challenge datasets have shown that the proposed oRGB-SIFT descriptor improves recognition performance upon other color SIFT descriptors, and the CSF, the CGSF, and the CGSF+PHOG descriptors perform better than the other color SIFT descriptors. The fusion of both Color SIFT descriptors (CSF) and Color Grayscale SIFT descriptor (CGSF) shows significant improvement in the classification performance, which indicates that various color-SIFT descriptors and grayscale-SIFT descriptor are not redundant for image search.

Second, four novel color Local Binary Pattern (LBP) descriptors are presented for scene image and image texture classification. Specifically, the oRGB-LBP descriptor is

derived in the oRGB color space. The other three color LBP descriptors, namely, the Color LBP Fusion (CLF), the Color Grayscale LBP Fusion (CGLF), and the CGLF+PHOG descriptors, are obtained by integrating the oRGB-LBP descriptor with some additional image features. Experimental results on three large scale, grand challenge datasets have shown that the proposed descriptors can improve scene image and image texture classification performance.

Finally, a new iris recognition method based on a robust iris segmentation approach is presented for improving iris recognition performance. The proposed robust iris segmentation approach applies power-law transformations for more accurate detection of the pupil region, which significantly reduces the candidate limbic boundary search space for increasing detection accuracy and efficiency. As the limbic circle, which has a center within a close range of the pupil center, is selectively detected, the eyelid detection approach leads to improved iris recognition performance. Experiments using the Iris Challenge Evaluation (ICE) database show the effectiveness of the proposed method.

# INVESTIGATION ON ADVANCED IMAGE SEARCH TECHNIQUES

by
**Abhishek Verma**

**A Dissertation**
**Submitted to the Faculty of**
**New Jersey Institute of Technology**
**in Partial Fulfillment of the Requirements for the Degree of**
**Doctor of Philosophy in Computer Science**

**Department of Computer Science**

**August 2011**

# APPROVAL PAGE

## INVESTIGATION ON ADVANCED IMAGE SEARCH TECHNIQUES

**Abhishek Verma**

---

Dr. Chengjun Liu, Dissertation Advisor Date
Associate Professor of Computer Science, NJIT

---

Dr. James McHugh, Committee Member Date
Professor of Computer Science, NJIT

---

Dr. David Nassimi, Committee Member Date
Associate Professor of Computer Science, NJIT

---

Dr. Usman Roshan, Committee Member Date
Associate Professor of Computer Science, NJIT

---

Dr. Hongya Ge, Committee Member Date
Associate Professor of Electrical and Computer Engineering, NJIT

# BIOGRAPHICAL SKETCH

**Author:** Abhishek Verma

**Degree:** Doctor of Philosophy

**Date:** August 2011

**Undergraduate and Graduate Education:**

- Doctor of Philosophy in Computer Science,
  New Jersey Institute of Technology, Newark, New Jersey, 2011
- Master of Science in Computer Science,
  New Jersey Institute of Technology, Newark, New Jersey, 2006
- Master of Computer Applications,
  Bangalore University, Bangalore, India, 2003
- Bachelor in Commerce (Honors),
  University of Delhi, New Delhi, India, 2000

**Major:** Computer Science

**Presentations and Publications:**

Abhishek Verma, Chengjun Liu, and Jiancheng Jia, "Iris Recognition based on Robust Iris Segmentation and Image Enhancement," *International Journal of Biometrics*, 2011 (to appear).

Abhishek Verma, Chengjun Liu, and Jiancheng Jia, "New Color SIFT Descriptors for Image Classification with Applications to Biometrics," *International Journal of Biometrics*, vol. 3, no. 1, pp. 56-75, 2011.

Sugata Banerji, Abhishek Verma, and Chengjun Liu, "Novel Color LBP Descriptors for Scene and Image Texture Classification," *Proceedings of the 15th International Conference on Image Processing, Computer Vision, and Pattern Recognition*, Las Vegas, Nevada, July 18-21, 2011.

Abhishek Verma and Chengjun Liu, "Fusion of Color SIFT Features for Image Classification with Applications to Biometrics," *Proceedings of the 11th IAPR International Conference on Pattern Recognition and Information Processing*, Minsk, Belarus, May 18-20, 2011.

Abhishek Verma and Chengjun Liu, "Novel EFM-KNN Classifier and a New Color Descriptor for Image Classification," *Proceedings of the 20th IEEE Wireless and Optical Communications Conference*, Newark, New Jersey, USA, April 15-16, 2011.

Abhishek Verma, Sugata Banerji, and Chengjun Liu, "A New Color SIFT Descriptor and Methods for Image Category Classification," *Proceedings of the 2010 International Congress on Computer Applications and Computational Science*, Singapore, December 4-6, 2010.

Abhishek Verma and Chengjun Liu, "Efficient Iris Identification with Improved Segmentation Techniques," *Cross-Disciplinary Applications of Artificial Intelligence and Pattern Recognition: Advancing Technologies*, V.K. Mago and N. Bhatia Eds., IGI Global, Hershey, PA (to appear).

Abhishek Verma, "Novel EFM-KNN Classifier and a New Color Descriptor for Image Classification," *Oral Presentation at the 20th IEEE Wireless and Optical Communications Conference*, Newark, New Jersey, USA, April 15-16, 2011.

Abhishek Verma, "New SIFT Descriptors for Image Category Classification," *Poster Presentation at the 6th Annual Graduate Student Research Day*, New Jersey Institute of Technology, Nov 4, 2010, Newark, New Jersey, USA.

To my beloved parents.

# ACKNOWLEDGMENT

Foremost, I would like to express my sincere appreciation to my dissertation advisor, Dr. Chengjun Liu, for his technical guidance, kindness, and confidence in me to bring this dissertation to fruition. Over the past several years, Dr. Liu has been a valued mentor to me. I will always be indebted to Dr. Liu for encouraging and supporting me to present our research achievements to international conferences, which significantly improved my confidence and gave me great opportunities to exchange insights with leading researchers in the world.

Secondly, I am extremely grateful to Dr. James McHugh, Dr. David Nassimi, Dr. Usman Roshan, and Hongya Ge for serving on my dissertation committee. In addition, I would like to extend special thanks to Dr. Kevin Jia from International Game Technology, Reno, NV. They have provided me with academic advices inside and outside my research field. Their friendship and wit have been a tremendous source of emotional support. This dissertation would not have been possible without their invaluable advice and generous help. I would also like to thank my fellow graduate students for their assistance and support, in particular to Sugata Banerji, Shuo Chen, Zhiming Liu, and Venkata Edupuganti.

Most importantly, I heartily appreciate my family for being my emotional anchor through not only the roughness of Ph.D. study, but my entire life. I thank my parents, Dr. Bindra Prasad and Dr. Pushpa Verma, for their faith in me. It was under their watchful eye that I gained so much strength and ability to tackle head-on challenges. Along the way, I am thankful to my sister, Sonika Sharma, for sharing her caring thoughts and providing endless encouragement.

# TABLE OF CONTENTS

# LIST OF TABLES

## LIST OF FIGURES

# CHAPTER 1

# INTRODUCTION

Content-based image search for retrieval of images based on similarities in their visual contents such as features from color, texture, shapes, etc. to a user-supplied query image or user-specified image features has been a focus of interest for the last several years. Color features provide powerful information for image search, indexing, and classification (Liu and Yang 2009), (Yang and Liu 2008), (Shih and Liu 2005), in particular for identification of biometric images (Verma et al. 2011b; Verma and Liu 2011b), objects, natural scene, image texture and flower categories (Verma et al. 2010; Verma and Liu 2011c), (Banerji et al. 2011) and geographical features from images. The choice of a color space is important for many computer vision algorithms. Different color spaces display different color properties. With the large variety of available color spaces, the inevitable question that arises is how to select a color space that produces best results for a particular computer vision task. Two important criteria for color feature detectors are that they should be stable under varying viewing conditions, such as changes in illumination, shading, highlights, and they should have high discriminative power. Color features such as the color histogram, color texture and local invariant features provide varying degrees of success against image variations such as viewpoint and lighting changes, clutter and occlusions (Datta et al. 2008), (Burghouts and Geusebroek 2009), (Stokman and Gevers 2007).

## 1.1 SIFT Feature Representation

Lately, there has been much emphasis on the detection and recognition of locally affine invariant regions (Lowe 2004), (Mikolajczyk et al. 2005). Successful methods are based on representing a salient region of an image by way of an elliptical affine region, which describes local orientation and scale. After normalizing the local region to its canonical form, image descriptors are able to capture the invariant region appearance. Interest point detection methods and region descriptors can robustly detect regions, which are invariant to translation, rotation and scaling (Lowe 2004), (Mikolajczyk et al. 2005). Affine region detectors when combined with the intensity Scale-Invariant Feature Transform (SIFT) descriptor (Lowe 2004) has been shown to outperform many alternatives (Mikolajczyk et al. 2005).

In this dissertation, the SIFT descriptor is extended to different color spaces, including the recently proposed oRGB color space (Bratkova et al. 2009), a new oRGB-SIFT feature representation is proposed, furthermore it is integrated with other color SIFT features to produce the Color SIFT Fusion (CSF), and the Color Grayscale SIFT Fusion (CGSF) descriptors. Additionally, the CGSF is combined with the Pyramid of Histograms of Orientation Gradients (PHOG) to obtain a new CGSF+PHOG descriptor for image category classification with special applications to biometrics. Classification is implemented using a novel EFM-KNN classifier (Liu and Wechsler 2002), (Liu and Wechsler 2000b), which combines the Enhanced Fisher Model (EFM) and the K Nearest Neighbor (KNN) decision rule (Fukunaga 1990). The effectiveness of the proposed descriptors and classification method will be evaluated on four large scale, grand challenge datasets: the Caltech 256 database, MIT scene database, Oxford flower database, and the UPOL Iris database.

## 1.2 LBP Feature Representation

In recent years, the recognition and classification of textures using the Local Binary Pattern (LBP) features has been shown to be promising (Ojala et al. 1994), (Ojala et al. 1996), (Zhu et al. 2010), (Chen et al. 2010), (Crosier and Griffin 2008). Color features when combined with the intensity based texture descriptors are able to outperform many alternatives. In this dissertation, a variable mask size is employed in order to generate a multi-scale LBP feature vector that is more robust to changes of scale and orientation. Further, the multi-scale LBP descriptor is extended to different color spaces including the oRGB color space

(Bratkova et al. 2009). A new multi-scale oRGB-LBP feature representation is proposed, which is further integrated with other color LBP features to produce the novel multi-scale Color LBP Fusion (CLF) and the multi-scale Color Grayscale LBP Fusion (CGLF) descriptors. The CGLF is further combined with PHOG to obtain the novel CGLF+PHOG descriptor. Feature extraction applies the Enhanced Fisher Model (EFM) (Liu and Wechsler 2000b), (Liu and Wechsler 2002) and image classification is based on the nearest neighbor classification rule (EFM-NN). The effectiveness of the proposed descriptors and classification methodology will be evaluated using three grand challenge datasets: the MIT scene database, the KTH-TIPS2-b and the KTH-TIPS materials databases.

## 1.3 Iris Recognition and Biometric Authentication

Over the past decade biometric authentication has become a very active area of research due to the increasing demands in automated personal identification. More recently several new notable techniques and methods with applications to face recognition (Shih and Liu 2011; Liu and Yang 2009; Liu 2007; Yang et al. 2010), eye detection (Shuo and Liu 2010) and iris (Verma et al. 2011a; Verma and Liu 2011a; Verma et al. 2011b) biometrics have been proposed. Among many biometric techniques, iris recognition is one of the most promising approaches due to its high reliability for person identification (Ma et al. 2004).

The iris is a thin circular diaphragm, which lies between the lens and cornea of the human eye. The formation of the unique patterns of the iris is random and not related to any genetic factors (Wildes 1997), and the iris patterns remain stable throughout the adult life. Thus, the patterns within the iris are unique to each person and two eyes of an individual have independent iris patterns. Some research shows that when compared with other biometric features such as face and fingerprint, iris patterns are more stable and reliable (Du et al. 2004).

In this dissertation, a new iris search method is proposed based on a robust iris segmentation approach for improving iris recognition performance (Verma et al. 2011a), (Verma and Liu 2011a). Major improvements are proposed to the iris segmentation phase. In particular, (i) power-law transformations are implemented for more accurate detection of the pupil region. Additionally, (ii) with the proposed technique the candidate limbic boundary search space can be reduced considerably, leading to a significant increase in the accuracy and speed of the segmentation. The segmentation performance is further enhanced with the application of thresholding. Furthermore, (iii) for higher accuracy and speed, the limbic circle having a center within close range of the pupil center is selectively detected. Additionally, (iv) proposed eyelid detection approach is shown to improve performance. The effectiveness of the proposed method is evaluated on a grand challenge, large scale database: the Iris Challenge Evaluation (ICE) (Phillips 2006) dataset.

Proposed method is able to correctly segment the pupil for 99.8% of the images in the dataset. Iris region detection is 98.5% for the right eye and 98.8% for the left eye. The rank-one recognition rate for proposed method is 3.5% and 2.7% higher than that of the irisBEE method (Phillips et al. 2008) for the right eye and the left eye respectively. Furthermore, proposed method improves upon the ND_IRIS (Liu et al. 2005) by a significant 1.9% on the rank-one recognition rate for the left eye. The verification rate is about 10% higher than the irisBEE method (Phillips et al. 2008) for each eye at a much lower equal error rate; this emphasizes the higher accuracy of proposed method.

The rest of the dissertation is structured as follows: Chapter 2 presents a brief overview of several representative works on color image search, color image representation, learning and classification techniques, and object and scene search and retrieval followed by a review on image acquisition, segmentation, feature encoding and matching for the iris image search. Chapter 3 describes the new color SIFT and LBP descriptors, presents an overview of five color spaces, the details of the EFM feature extraction technique, and classification by the novel EFM-KNN classifier. Followed by a description of datasets used in the experiments along with a detailed evaluation of color descriptors and classification methodology is presented in Chapter 4. Chapter 5 describes the iris dataset used in the experiments along with the implementation details of the proposed improved iris recognition method based on robust iris segmentation and image enhancement followed by a thorough performance evaluation of the proposed method and a detailed analysis of the experimental results. The conclusions and future work are presented in Chapter 6, where the major contributions of this dissertation are summarized and future research directions are discussed.

# CHAPTER 2

# RELATED WORK

## 2.1 Color Image Search

### 2.1.1 Image-level Global and Local Feature Descriptors

In recent years, use of color as a means to biometric image retrieval (Liu and Yang 2009), (Liu 2006), (Shih and Liu 2005) and object and scene search has gained popularity. Color features can capture discriminative information by means of the color invariants, color histogram, color texture, etc. The earliest methods for object and scene classification were mainly based on the global descriptors such as the color and texture histogram (Niblack et al. 1993), (Pontil and Verri 1998), (Schiele and Crowley 2000). One of the earlier works is the color indexing system designed by Swain and Ballard, which uses the color histogram for image inquiry from a large image database (Swain and Ballard 1991). Such methods are sensitive to viewpoint and lighting changes, clutter and occlusions. For this reason, global methods were gradually replaced by the part-based methods, which became one of the popular techniques in the object recognition community. Part-based models combine appearance descriptors from local features along with their spatial relationship. Harris interest point detector was used for local feature extraction; such features are only invariant to translation (Agarwal and Roth 2002), (Weber et al. 2000). Afterwards, local features with greater invariance were developed, which were found to be robust against scale changes (Fergus et al. 2003) and affine deformations (Lazebnik et al. 2004). Learning and inference for spatial relations poses a challenging problem in terms of its complexity and computational cost. Whereas, the orderless bag-of-words methods (Fergus et al. 2003), (Leung and Malik 2001), (Jurie and Triggs 2005) are simpler and computationally efficient, though they are not able to represent the geometric structure of the object or to distinguish between foreground and background features. For these reasons, the bag-of-words methods are not

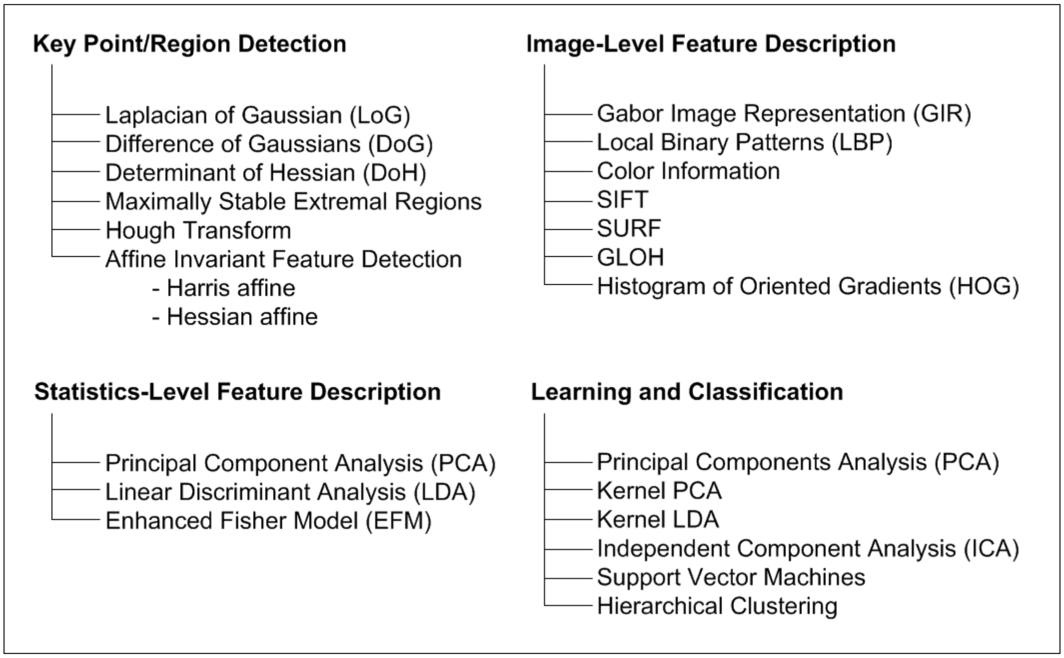


**Figure 2.1** An overview of the various techniques in color image search.

robust to clutter. One way to overcome this drawback is to design kernels that can yield high discriminative power in presence of noise and clutter (Grauman and Darrell 2005).

More recent work on color based image classification appears in (Liu and Yang 2009), (Yang and Liu 2008), (Liu 2008) that propose several new color spaces and methods for face classification and in (Bosch et al. 2008) the HSV color space is used for the scene category recognition. Evaluation of local color invariant descriptors is performed in (Burghouts and Geusebroek 2009). Fusion of color models, color region detection and color edge detection have been investigated for representation of color images (Stokman and Gevers 2007). Key contributions in color, texture, and shape abstraction have been discussed in Datta et al. (Datta et al. 2008).

As discussed before, many recent techniques for the description of images have considered local features. The most successful local image descriptor so far is Lowe's SIFT descriptor (Lowe 2004). The SIFT descriptor encodes the distribution of Gaussian gradients within an image region. It is a 128-bin histogram that summarizes the local oriented gradients over 8 orientations and over 16 locations. This can efficiently represent the spatial intensity pattern, while being robust to small deformations and localization errors. Currently, several modifications to the SIFT features have been proposed; among them are the PCA-SIFT (Ke and Sukthankar 2004), GLOH (Mikolajczyk and Schmid 2005), and SURF (Bay et al. 2008). These region-based descriptors have achieved a high degree of invariance to the overall illumination conditions for planar surfaces. Although, designed to retrieve identical object patches, SIFT-like features turn out to be quite successful in the bag-of-words approaches for general scene and object classification (Bosch et al. 2008).

Lately, several methods based on LBP features have been proposed for image representation and classification (Zhu et al. 2010), (Crosier and Griffin 2008). Extraction of LBP features is computationally efficient and with the use of multi-scale filters; invariance to scaling and rotation can be achieved (Zhu et al. 2010). Fusion of different features has been shown to achieve a good retrieval success rate (Banerji et al. 2011), (Crosier and Griffin 2008), (Zhang et al. 2007). Local image descriptors have also been shown to perform well for texture based image retrieval (Banerji et al. 2011), (Chen et al. 2010), (Zhang et al. 2007). In a $3 \times 3$ neighborhood of an image, the basic LBP operator assigns a binary label 0 or 1 to each surrounding pixel by thresholding at the gray value of the central pixel and replacing its value with a decimal number converted from the 8-bit binary number.

The Pyramid of Histograms of Orientation Gradients (PHOG) descriptor Bosch et al. (2007) is able to represent an image by its local shape and the spatial layout of the shape. The local shape is captured by the distribution over edge orientations within a region, and the spatial layout by tiling the image into regions at multiple resolutions. The distance between two PHOG image descriptors then reflects the extent to which the images contain similar shapes and correspond in their spatial layout.

### 2.1.2 Statistics-level Feature Descriptors

The color images reside usually in a high-dimensional image space. There is a great demand to find meaningful and compact patterns in such a space for developing robust image recognition methods so as to meet two requirements: enhanced discrimination ability and computational efficiency. Therefore, most appearance-based image recognition algorithms usually start with the dimensionality reduction by using some popular linear subspace methods. In the following sections, several major statistical methods are introduced.

*Principal Component Analysis (PCA)*

As an optimal linear transformation in the sense of minimum Mean Square Error (MSE), Principal Component Analysis (PCA) has been a leading technique for dimensionality reduction of input data. Given a set of $d$-dimensional column image vectors $\{X_{ij}\}$, where $X_{ij} \in \mathbb{R}^d$ is the $j$-th image of class $i$. Let the training set consist of $c$ persons and $l_i$ sample images for person $i$. Thus, the number of training samples is $m = \sum_{i=1}^{c} l_i$. For image recognition, each class has a prior probability of $\lambda_i$. The within-class scatter matrix is defined as:

$$S_w = \sum_{i=1}^{c} \frac{\lambda_i}{l_i} \sum_{j=1}^{l_i} (X_{ij} - \overline{X_i})(X_{ij} - \overline{X_i})^T \tag{2.1}$$

where $\overline{X_i} = \frac{1}{l_i} \sum_{j=1}^{l_i} X_{ij}$ is the mean of class $i$. The between-class scatter matrix $S_b$ and the total (mixture) scatter matrix $S_t$ are defined respectively as:

$$S_b = \sum_{i=1}^{c} \lambda_i (\overline{X_i} - \overline{X})(\overline{X_i} - \overline{X})^T \tag{2.2}$$

$$S_t = \sum_{i=1}^{c} \frac{\lambda_i}{l_i} \sum_{j=1}^{l_i} (X_{ij} - \overline{X})(X_{ij} - \overline{X})^T \tag{2.3}$$

where $\overline{X} = \frac{1}{m} \sum_{i=1}^{c} \sum_{j=1}^{l_i} X_{ij}$ is the grand mean.

PCA seeks a principal subspace of lower dimensionality to maximize the data reconstruction capability of the features. As a result, the features in this subspace can represent the original data accurately. The objective function of PCA can be defined as:

$$W^* = \underset{\|W\|=1}{\operatorname{argmax}} |W^T S_t W| \tag{2.4}$$

Maximizing the above equation can be solved via eigenvalue-eigenvector analysis. That is, the matrix $W^*$ can be constructed by obtaining the $k$ principal eigenvectors corresponding to the $k$ largest eigenvalues of $S_t$.

### *Linear Discriminant Analysis (LDA)*

The best representation of data may not perform well from the classification point of view, as the total scatter matrix consists of both the within- and between-class variations. To obtain the discrimination of features for differentiating images of one category from another, one needs to manipulate the within- and between-class variations separately. To that end, face recognition using Linear Discriminant Analysis (LDA) has been an area of increasing interest. LDA is also known as Fisher Linear Discriminant (FLD). The objective function of LDA can be defined as:

$$W^* = \underset{W}{\operatorname{argmax}} \frac{|W^T S_b W|}{|W^T S_w W|} \tag{2.5}$$

Equation (2.5) is called the Fisher criterion. To maximize the ratio value of this criterion, LDA seeks an optimal subspace $W^*$ that separates the different classes as far as possible and compresses the same classes as compactly as possible. To derive $W^*$, LDA solves the generalized eigenvectors of $S_b W = \lambda S_w W$, and chooses the $k$ principal eigenvectors corresponding to the $k$ largest eigenvalues.

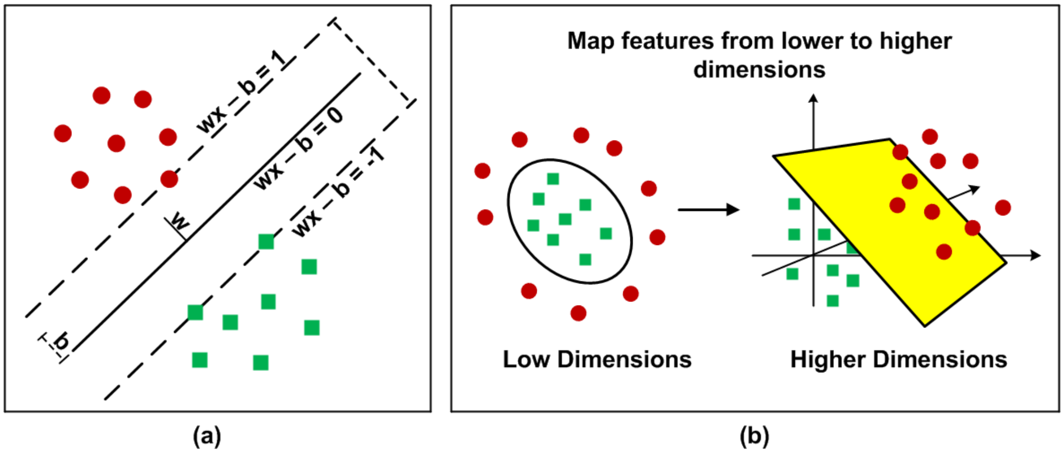


**Figure 2.2** Maximum-margin hyperplane and margins for a SVM in 2D space trained with samples from two classes. Samples on the margin are called the support vectors. (b) Map features from lower dimensions to higher dimensions for easier separability.

### 2.1.3 Learning and Classification

Efficient retrieval requires a robust feature extraction method that has the ability to learn meaningful low-dimensional patterns in spaces of very high dimensionality (Liu 2003), (Liu and Wechsler 2003), (Liu and Wechsler 2000a). Low-dimensional representations are also important when one considers the intrinsic computational aspect. PCA has been widely used to perform dimensionality reduction for image indexing and retrieval (Liu and Wechsler 2000b), (Liu 2004b). Recently, Support Vector Machine (SVM) classifier for multiple category recognition has gained popularity (Zhang et al. 2007), (Bosch et al. 2008) though it suffers from the drawback of being computationally too expensive on large scale image classification tasks. The EFM methodology has achieved good success for the task of image based recognition (Liu and Wechsler 2002), (Liu and Wechsler 2001), (Liu 2004a). See Figure 2.1 for an overview of the various techniques for color image search.

The SVM performs classification by constructing an N-dimensional hyperplane that optimally separates the data into two categories. SVMs are closely related to classical multilayer perceptron neural networks. The goal of SVM modeling is to find the optimal hyperplane that separates clusters of sample vectors in such a way that samples from different target classes lie on different sides of the plane. See Figure 2.2 for an overview of a two class SVM classification. Let the training data be a set of $n$ points of the form:

$$D = \{(x_i, c_i) | x_i \in R^p, c_i \in \{-1, 1\}\}_{i=1}^{n} \tag{2.6}$$

where $c_i$ has a value of -1 or 1, which indicates the class to which $x_i$ belongs. Each $x_i$ is a $p$ dimensional vector. The goal is to find the maximum-margin hyperplane that divides the points having $c_i = 1$ from those having $c_i = -1$. Any hyperplane can be written as a set of points $x$ satisfying:

$$w \bullet x - b = 0 \tag{2.7}$$

where the vector $w$ is a vector normal to the hyperplane and $\bullet$ denotes the dot product. The parameter $b / \parallel w \parallel$ determines the offset of the hyperplane from the origin along the normal vector $w$. The $w$ and $b$ are to be chosen in such a way that it maximizes the margin, or the distance between the parallel hyperplanes that are as far apart as possible while still separating the data. The primal form of the SVM problem is to minimize (in $w$, $b$):
$\frac{1}{2} \parallel w \parallel^2$ subject to (for any $i = 1, 2, \ldots, n$)

$$c_i(w \bullet x_i - b) \geq 1 \tag{2.8}$$

Using non-negative Lagrange multipliers $\alpha_i$ the solution can be expressed as:

$$w = \sum_{i=0}^{n} \alpha_i c_i x_i \tag{2.9}$$

and

$$b = \frac{1}{N_{SV}} \sum_{i=1}^{N_{SV}} (w \bullet x_i - c_i) \tag{2.10}$$

where $N_{SV}$ is the set of all support vectors. Using the fact, that $\parallel w \parallel^2 = w \bullet w$ and substituting, one can show that the dual of the SVM reduces to the following optimization problem:

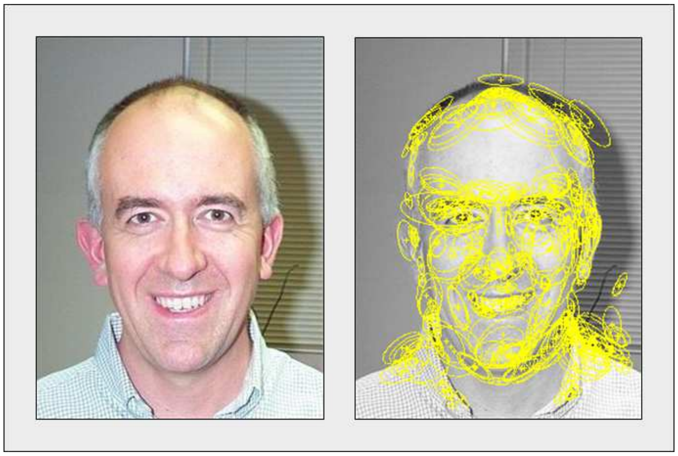


**Figure 2.3** The left image3.5 from the Faces category in the Caltech 256 dataset. Image on the right shows the keypoints and elliptical regions detected by the Harris-affine detector.

Maximize (in $\alpha_i$)

$$\tilde{L}(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i,j} \alpha_i \alpha_j c_i c_j x_i^T x_j = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i,j} \alpha_i \alpha_j c_i c_j k(x_i x_j) \qquad (2.11)$$

subject to (for any $i = 1, 2, \ldots, n$) $\alpha_i \geq 0$ and $\sum_{i=1}^{n} \alpha_i c_i = 0$. Here the kernel is defined by: $k(x_i x_j) = x_i \bullet x_j$

The original optimal hyperplane algorithm proposed by Vladimir Vapnik (1963) was a linear classifier. Boser et al. (1992) suggested a way to create non-linear classifiers by applying the kernel trick (Aizerman et al. 1964) to maximum-margin hyperplanes. Multiclass SVM aims to assign labels to instances by using support vector machines, where the labels are drawn from a finite set of several elements. The approach that is commonly used is to reduce the single multiclass problem into multiple two-class problems. The two methods commonly employed to build such binary classifiers are one-versus-all and one-versus-one. One popular implementation of the SVM is the LibSVM (Chang and Lin 2011).

### 2.1.4 Key Region Detection by Affine Detectors

The affine key point detector detects interest points in scale-space, and then determines an elliptical region for each point. Interest points maybe detected with the Harris detector. The elliptical region is determined with the second moment matrix of the intensity gradient (Lindeberg and Garding 1997). The second moment matrix is often used for feature detection and for describing local image structures. It is used for point detection by Harris detector and for the estimation of surrounding area about the point. The Harris detector (Harris and Stephens 1988) is based on the following principle. First, the local image derivatives are computed with Gaussian kernels of differentiation scale. Second, the derivatives are averaged in the neighborhood of the point by smoothing with a Gaussian window of integration scale. The eigenvalues of this matrix represent two principal signal changes in a neighborhood of the point. Those points for which the signal change is significant in orthogonal directions are extracted. Such points are stable in arbitrary lighting conditions and represent the image fairly well.

After the extraction of a set of initial points the next step is to perform the iterative estimation of elliptical affine region (Lindeberg and Garding 1997). The region is then normalized to the circular one. Harris-affine and the Hessian-affine detectors are robust to light intensity changes, blurring, scale, and viewpoint changes (Mikolajczyk et al. 2005). See Figure 2.3 for the points and regions detected by the Harris-affine detector.

## 2.2 Iris Image Search

A general approach to iris image search consists mainly of four stages: (1) image acquisition, (2) iris segmentation, (3) texture analysis, and (4) matching of texture patterns. Several notable contributions to the aforementioned stages are summarized here.

One of the earlier systems proposed by Flom and Safir (1987) detected the pupil region by finding large connected regions of pixels with intensity values below a given threshold. Iris descriptors were extracted using the difference operator, edge detection algorithms, and the Hough transform. Wildes (1997) system used low light level camera along with diffuse source and polarization for image acquisition. Iris region was segmented by computing the binary edge map followed by the Hough transform. For matching, it applied the Laplacian of Gaussian filter at multiple scales to produce a template and computes the normalized correlation as a similarity measure.

Masek (2003) performed segmentation of iris by canny edge detection and circular Hough transform. Encoding was performed by 1D Log-Gabor wavelets and matching was based on hamming distance. Liu et al. (2005) proposed the ND_IRIS method based on Masek's implementation, hamming distance was used to compare two iris templates. The method proposed by Daugman (2006, 2007) performed segmentation of the iris region with the integro-differential operator followed by its normalization. The normalized iris image was convolved with the 2D Gabor filters to extract the texture information, which was quantized into a compact 256 byte binary iriscode. The iriscodes were compared using the normalized Hamming distance.

The issue of noise detection and handling for non-cooperative iris recognition was explored by Proenca (2006); Proenca and Luis (2007). Bayesian approach to matching of warped iris patterns was discussed by Thornton et al. (2007). More updated methods in image understanding for iris biometrics were reviewed by Bowyer et al. (2008). Vatsa et al. (2008) proposed a curve evolution approach to segment a non-ideal iris image using the modified Mumford-Shah functional. Beacon guided search for fast iris matching was discussed by Hao et al. (2008) and use of short-length iris codes from the most descriptive regions of the iris for fast matching was proposed by Gentile et al. (2009). He et al. (2009) proposed an Adaboost-cascade iris detector for fast iris segmentation. Spline based edge fitting scheme was used for non-circular iris boundary detection. Eyelashes and shadows were detected via a learned prediction model. Baker et al. (2010) explored the issue of degraded iris biometrics performance with non-cosmetic contact lenses. Proenca (2010) proposed a segmentation method to handle the degraded images acquired in less constrained conditions for real-time applications.  Puhan et al. (2011) proposed a fast iris segmentation technique based on the Fourier spectral density for noisy frontal view eye images captured with minimum cooperation from the subjects.

# CHAPTER 3

# COLOR SPACES, NEW COLOR DESCRIPTORS AND THE NOVEL EFM-KNN CLASSIFIER

This chapter first presents a review of five color spaces in which the new color SIFT and new color LBP descriptors are defined followed by a discussion on clustering, visual vocabulary tree, and visual words for SIFT descriptors. Thereafter, five conventional SIFT descriptors are presented: the RGB-SIFT, the rgb-SIFT, the HSV-SIFT, the YCbCr-SIFT, and the grayscale-SIFT descriptors and four new color SIFT descriptors are presented: the oRGB-SIFT, the Color SIFT Fusion (CSF), the Color Grayscale SIFT Fusion (CGSF), and the CGSF+PHOG descriptors for image classification with special applications to biometrics. Second, four novel color Local Binary Pattern (LBP) descriptors are presented for scene image and image texture classification. Specifically, the oRGB-LBP descriptor is derived in the oRGB color space. The other three color LBP descriptors, namely, the Color LBP Fusion (CLF), the Color Grayscale LBP Fusion (CGLF), and the CGLF+PHOG descriptors, are obtained by integrating the oRGB-LBP descriptor with some additional image features. Followed by a detailed discussion on the novel EFM-KNN classification methodology.

## 3.1 Color Spaces

A color image contains three component images, and each pixel of a color image is specified in a color space, which serves as a color coordinate system. The commonly used color space is the RGB color space. Other color spaces are usually calculated from the RGB color space by means of either linear or nonlinear transformations.

To reduce the sensitivity of the RGB images to luminance, surface orientation, and other photographic conditions, the rgb color space is defined by normalizing the $R$, $G$, and

$B$ components:

$$\begin{aligned} r &= R/(R+G+B) \\ g &= G/(R+G+B) \\ b &= B/(R+G+B) \end{aligned} \tag{3.1}$$

Due to the normalization $r$ and $g$ are scale-invariant and thereby invariant to light intensity changes, shadows and shading (Gevers et al. 2006).

The HSV color space is motivated by human vision system because humans describe color by means of hue, saturation, and brightness. Hue and saturation define chrominance, while intensity or value specifies luminance (Gonzalez and Woods 2001). The HSV color space is defined as follows (Smith 1978):

$$\begin{aligned} &Let \begin{cases} MAX = max(R,G,B) \\ MIN = min(R,G,B) \\ \delta = MAX - MIN \end{cases} \\ &V = MAX \\ &S = \begin{cases} \frac{\delta}{MAX} & \text{if } MAX \neq 0 \\ 0 & \text{if } MAX = 0 \end{cases} \\ &H = \begin{cases} 60(\frac{G-B}{\delta}) & \text{if } MAX = R \\ 60(\frac{B-R}{\delta}+2) & \text{if } MAX = G \\ 60(\frac{R-G}{\delta}+4) & \text{if } MAX = B \\ not\ defined & \text{if } MAX = 0 \end{cases} \end{aligned} \tag{3.2}$$

The YCbCr color space is developed for digital video standard and television transmissions. In YCbCr, the RGB components are separated into luminance, chrominance

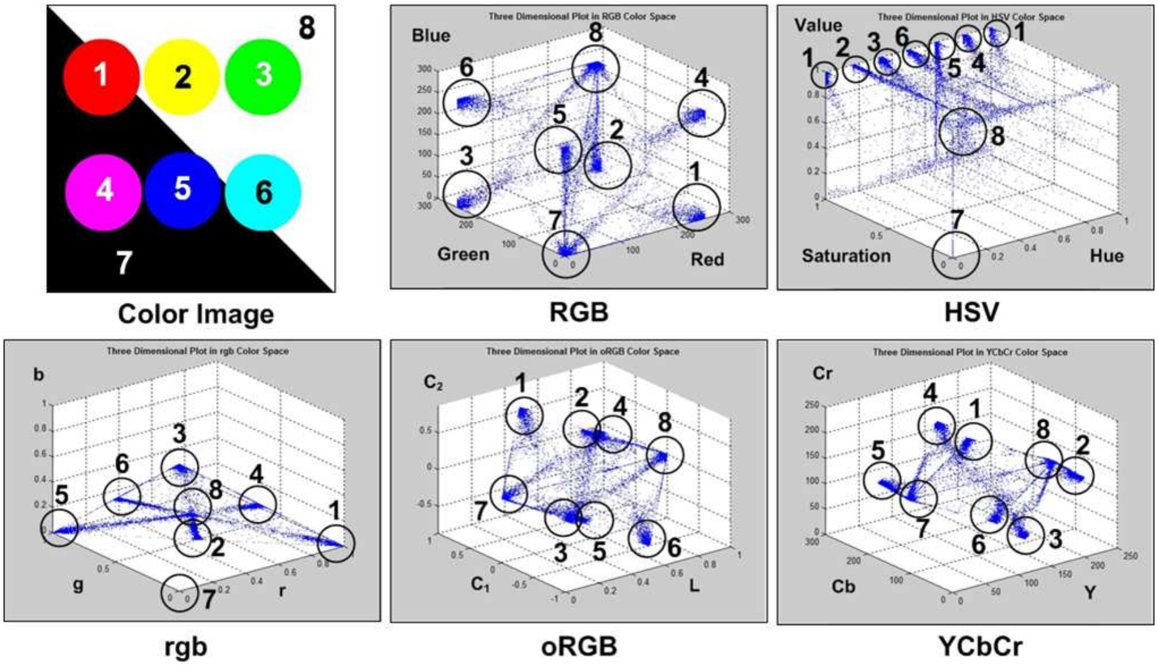


**Figure 3.1** Visualizing eight different colors in various color spaces. Top left is the image with eight colors numbered from 1 to 8. Left to right and top to bottom is the depiction of colors in RGB space, HSV space, rgb space, oRGB space and YCbCr space.

blue, and chrominance red:

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 65.4810 & 128.5530 & 24.9660 \\ -37.7745 & -74.1592 & 111.9337 \\ 111.9581 & -93.7509 & -18.2072 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{3.3}$$

where the $R, G, B$ values are scaled to $[0, 1]$.

The oRGB color space (Bratkova et al. 2009) has three channels $L$, $C1$ and $C2$. The primaries of this model are based on the three fundamental psychological opponent axes: white-black, red-green, and yellow-blue. The color information is contained in $C1$ and $C2$. The value of $C1$ lies within $[-1, 1]$ and the value of $C2$ lies within $[-0.8660, 0.8660]$. The

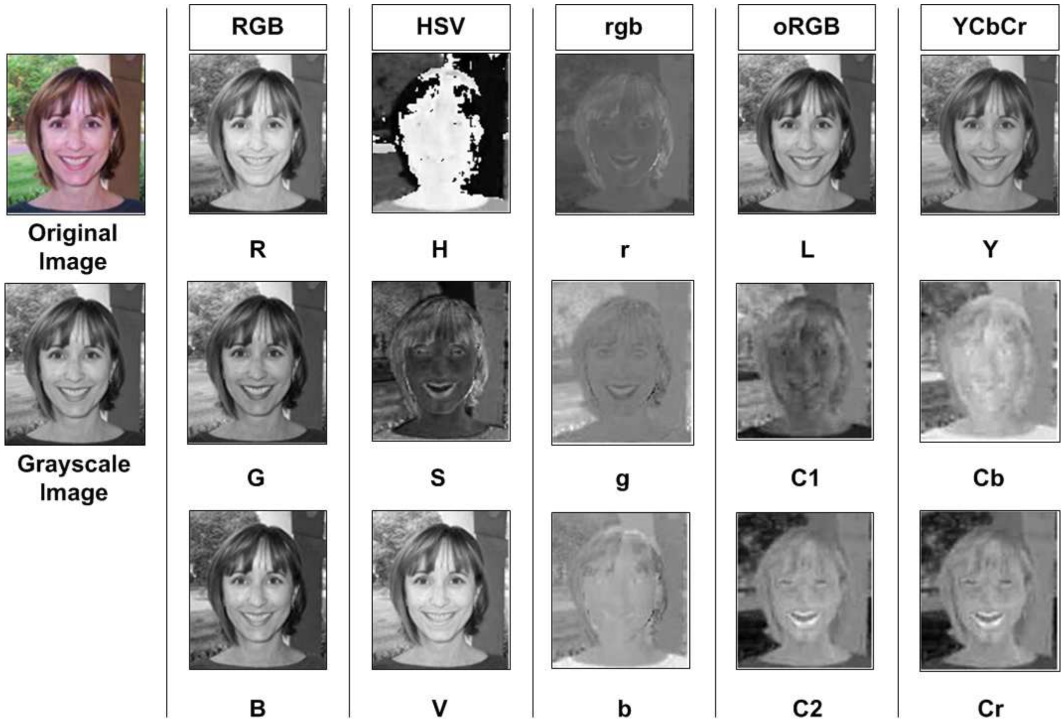


**Figure 3.2** Color component images in the five color spaces: RGB, HSV, rgb, oRGB, and YCbCr. The color image is from the Caltech 256 dataset, whose grayscale image is displayed as well.

$L$ channel contains the luminance information and its values range between $[0,1]$:

$$\begin{bmatrix} L \\ C1 \\ C2 \end{bmatrix} = \begin{bmatrix} 0.2990 & 0.5870 & 0.1140 \\ 0.5000 & 0.5000 & -1.0000 \\ 0.8660 & -0.8660 & 0.0000 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{3.4}$$

Figure 3.1 shows eight different colors in various color spaces. Figure 3.2 shows the color component images in the five color spaces: RGB, HSV, rgb, oRGB, and YCbCr.

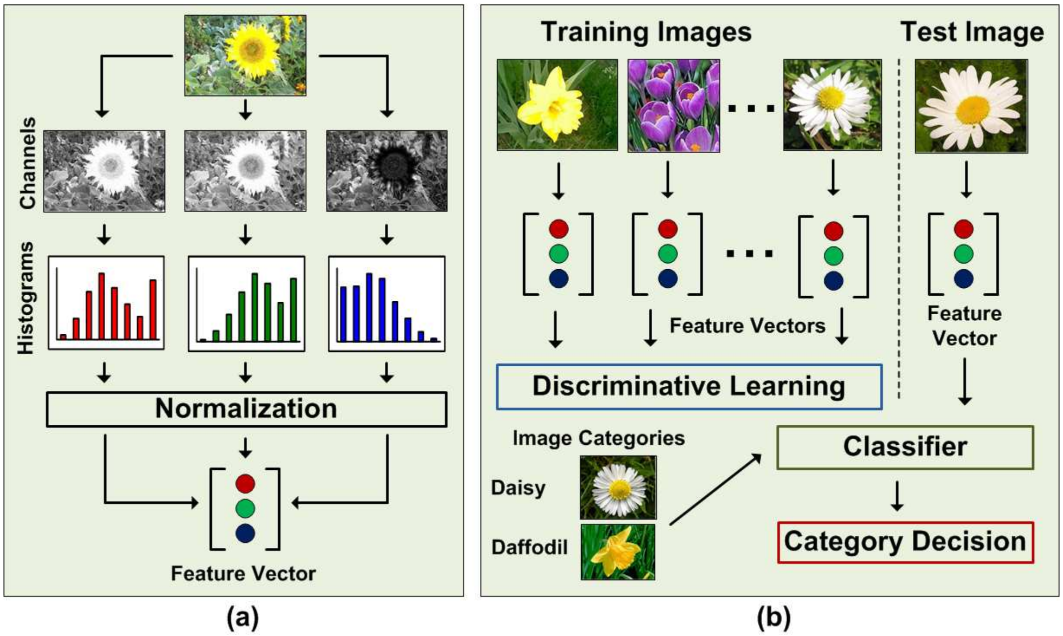


**Figure 3.3** (a) Various stages in representing an image as a feature vector of dense color histogram and (b) learning and classification stages.

## 3.2 Dense Color Histogram Descriptors

Here, an image is defined in terms of a dense histogram of the three color channels. This allows image representation by way of its predominant color features and classification based on its most discriminative color features. Figure 3.3(a) illustrates various stages in extracting a feature vector from a color image. The system starts with a color image as an input and first splits it into three separate color component images. Next step is to form a histogram from each of the color channels. In order to make the three histograms independent of the pixel count in an image, they are normalized to sum to one. The normalized histograms are then concatenated to form a compact fixed length feature vector. The following eight dense histograms are presented: Grayscale-DH, oRGB-DH, RGB-DH, HSV-DH, rgb-DH, YCbCr-DH, Color Histogram Fusion (CHF) that is formed from the combination of five dense color histograms, Color Gray Histogram Fusion (CGHF) is formed by combining Grayscale-DH with CGH. Figure 3.3(b) gives an overview of the learning and classification stages.

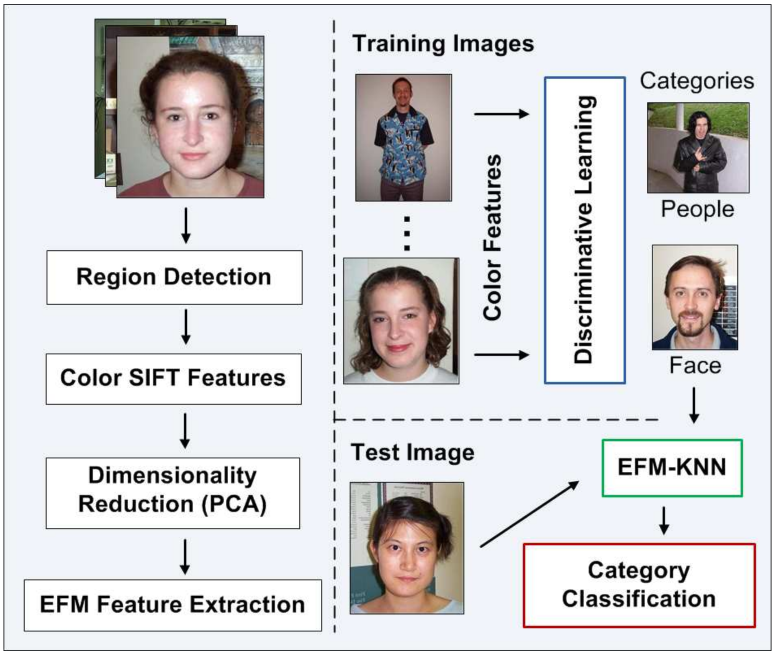


**Figure 3.4** An overview of SIFT feature extraction, visual words, learning and classification stages.

## 3.3 SIFT Feature Extraction, Clustering, Visual Vocabulary Tree, and Visual Words

This section first gives details of the SIFT feature extraction procedure. The next phase deals with the formation of visual vocabulary tree and visual words, here the normalized SIFT features are quantized with the vocabulary tree such that each image is represented as a collection of visual words, provided from a visual vocabulary. The visual vocabulary is obtained by vector quantizing descriptors computed from the training images using $k$-means clustering. See Figure 3.4 for an overview of the processing pipeline.

### 3.3.1 SIFT Feature Extraction

Image similarity may be defined in many ways based on the need of the application. It could be based on shape, texture, resolution, color or some other spatial features. The experiments here compute the SIFT descriptors extracted from the scale invariant points

(Zhang et al. 2007) on aforementioned color spaces. Such descriptors are called sparse descriptors, they have been previously used in (Csurka et al. 2004), (Lazebnik et al. 2003). Scale invariant points are obtained with the Hessian-affine point detector on the intensity channel. For the experiments, the Hessian-affine point detector is used because it has shown good performance in category recognition (Mikolajczyk et al. 2005). The remaining portion of feature extraction is then implemented according to the SIFT feature extraction pipeline of Lowe (Lowe 2004). Canonical directions are found based on an orientation histogram formed on the image gradients. SIFT descriptors are then extracted relative to the canonical directions.

### 3.3.2 Clustering, Visual Vocabulary Tree, and Visual Words

The visual vocabulary tree defines a hierarchical quantization that is constructed with the hierarchical $k$-means clustering. A large set of representative descriptor vectors taken from the training images are used in the unsupervised training of the tree. Instead of $k$ defining the final number of clusters or quantization cells, $k$ defines the branch factor (number of children of each node) of the tree. First, an initial $k$-means process is run on the training data, defining $k$ cluster centers. The training data is then partitioned into $k$ groups, where each group consists of the descriptor vectors closest to a particular cluster center. The same process is then recursively applied to each group of descriptor vectors, recursively defining clusters by splitting each cluster into $k$ new parts. The tree is determined level by level, up to some maximum number of levels say $L$, and each division into $k$ parts is only defined by the distribution of the descriptor vectors that belong to the parent cluster. This process is illustrated in Figure 3.5(a). Once the tree is computed, its leaf nodes are used for quantizing descriptors from the training and test images.

It has been experimentally observed that most important for the retrieval quality is to have a large vocabulary (large number of leaf nodes). While the computational cost of increasing the size of the vocabulary in a non-hierarchical manner would be very high, the

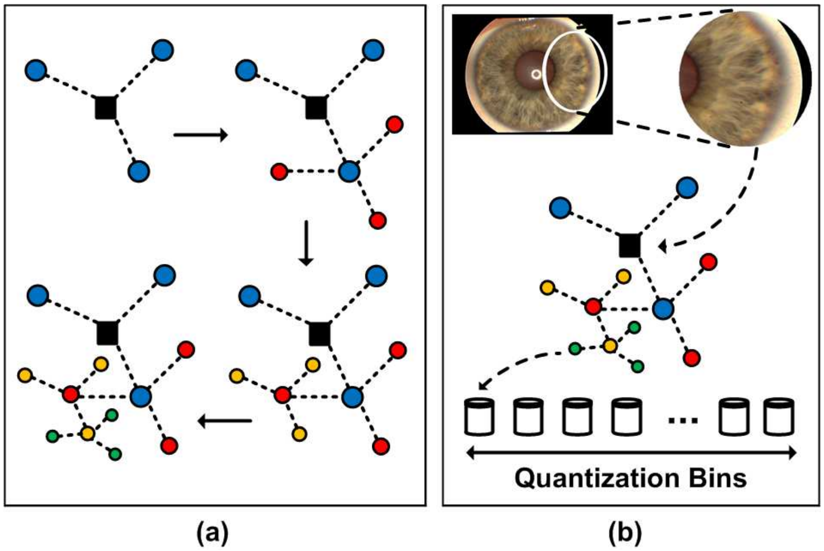


**Figure 3.5** (a) An illustration of the process of constructing a vocabulary tree by hierarchical $k$-means. The hierarchical quantization is defined at each level by $k$ centers (in this case $k = 3$). (b) A large number of elliptical regions are extracted from the image and normalized to circular regions. A SIFT descriptor vector is computed for each region. The descriptor vector is then hierarchically quantized by the vocabulary tree. The number of quantization bins is the number of leaf nodes in the vocabulary tree; this is the length of the final feature vector as well.

computational cost in the hierarchical approach is logarithmic in the number of leaf nodes. The memory usage is linear in the number of leaf nodes $kL$. The current implementation builds a tree of $6,561$ leaf nodes and $k = 9$.

To obtain fixed-length feature vectors per image, the visual words model is used (Bosch et al. 2008), (Csurka et al. 2004). The visual words model performs vector quantization of the color descriptors in an image against a visual vocabulary. In the quantization phase, each descriptor vector is simply propagated down the tree at each level by comparing the descriptor vector to the $k$ candidate cluster centers (represented by $k$ children in the tree) and choosing the closest one till it is assigned to a particular leaf node. This is a simple matter of performing $k$ dot products at each level, resulting in a total of $kL$ dot products, which is very efficient if $k$ is not too large. See Figure 3.5(b) for an overview of the quantization process.

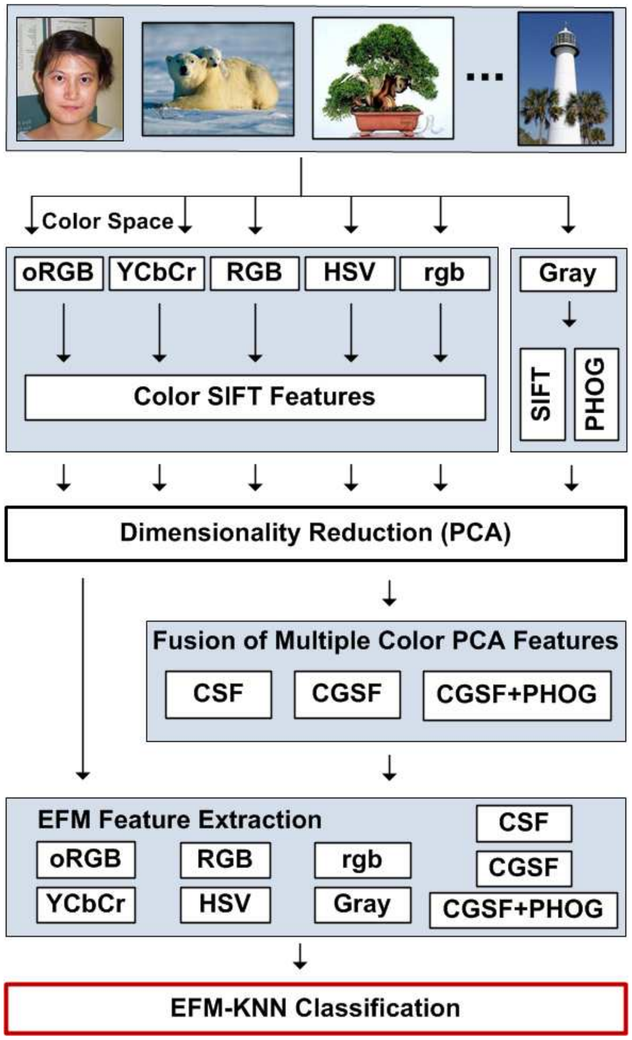


**Figure 3.6** Multiple Color SIFT features fusion methodology using the EFM feature extraction.

After all the SIFT features from an image are quantized, a fixed length feature vector would be obtained. The feature vector is normalized to zero mean and unit standard deviation. The advantage of representing an image as a fixed length feature vector lies in the fact that it allows to effectively compare images that vary in size.

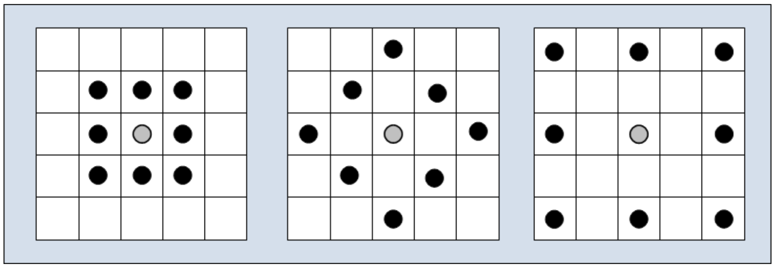


**Figure 3.7** The multi-scale LBP operators.

## 3.4 New Color SIFT Descriptors

The SIFT descriptor proposed by Lowe transforms an image into a large collection of feature vectors, each of which is invariant to image translation, scaling, and rotation, partially invariant to the illumination changes, and robust to local geometric distortion (Lowe 2004). The key locations used to specify the SIFT descriptor are defined as maxima and minima of the result of the difference of Gaussian function applied in the scale-space to a series of smoothed and resampled images. SIFT descriptors robust to local affine distortions are then obtained by considering pixels around a radius of the key location.

The grayscale SIFT descriptor is defined as the SIFT descriptor applied to the grayscale image. A color SIFT descriptor in a given color space is derived by individually computing the SIFT descriptor on each of the three component images in the specific color space. This produces a 384 dimensional descriptor that is formed from concatenating the 128 dimensional vectors from the three channels. As a result, four color SIFT descriptors are defined: the RGB-SIFT, the YCbCr-SIFT, the HSV-SIFT, and the rgb-SIFT descriptors.

The four new color SIFT descriptors are defined in the oRGB color space and the fusion in different color spaces. In particular, the oRGB-SIFT descriptor is constructed by concatenating the SIFT descriptors of the three component images in the oRGB color space. The Color SIFT Fusion (CSF) descriptor is formed by fusing the RGB-SIFT, the YCbCr-SIFT, the HSV-SIFT, the oRGB-SIFT, and the rgb-SIFT descriptors. The Color Grayscale SIFT Fusion (CGSF) descriptor is obtained by fusing further the CSF descriptor

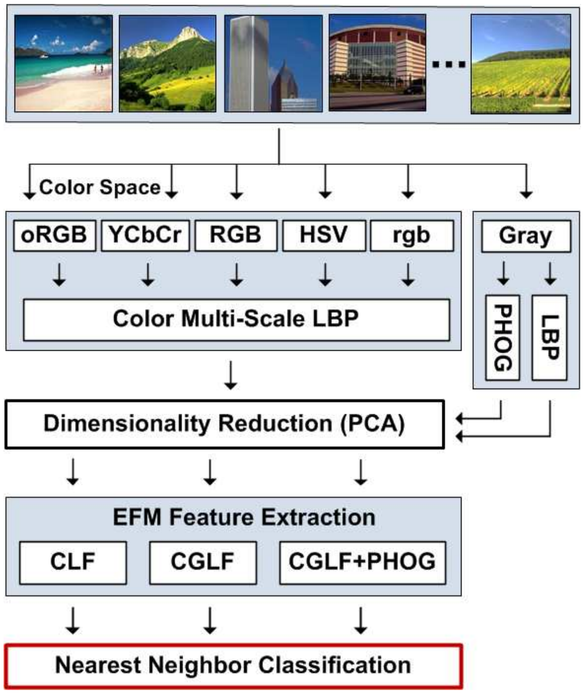


**Figure 3.8** Multiple Color LBP features fusion methodology using the EFM feature extraction.

and the grayscale-SIFT descriptor. The CGSF is combined with the Pyramid of Histograms of Orientation Gradients (PHOG) descriptor to obtain the CGSF+PHOG descriptor. See Figure 3.6 for multiple Color SIFT features fusion methodology.

## 3.5 New Color LBP Descriptors

The LBP descriptor proposed by Ojala et al. (1994, 1996) assigns an intensity value to each pixel of an image based on the intensity values of the eight neighboring pixels. Choosing multiple neighborhoods of different distances from the target pixel and orientations for each pixel has been shown to achieve partial invariance to scaling and rotation (Zhu et al. 2010). Using the multi-scale LBP operator shown in Figure 3.7, three LBP images from the three neighborhoods are generated. The normalized histograms from the LBP images are used as feature vectors and they are independent of the image size. The fused histograms of multi-scale LBP images give a feature vector that is partially invariant to image translation, scaling, and rotation.

The grayscale-LBP descriptor is defined as the LBP descriptor applied to the grayscale image. A color LBP descriptor in a given color space is derived by individually computing the LBP descriptor on each of the three component images in the specific color space. This produces a 2,304 dimensional descriptor that is formed from concatenating the 768 dimensional vectors from the three channels. As a result, the four color LBP descriptors are defined: the RGB-LBP, the YCbCr-LBP, the HSV-LBP, and the rgb-LBP descriptors.

The four new color LBP descriptors are defined in the oRGB color space and the fusion in different color spaces. In particular, the oRGB-LBP descriptor is constructed by concatenating the LBP descriptors of the three component images in the oRGB color space. The Color LBP Fusion (CLF) descriptor is formed by fusing the RGB-LBP, the YCbCr-LBP, the HSV-LBP, the oRGB-LBP, and the rgb-LBP descriptors. The Color Grayscale LBP Fusion (CGLF) descriptor is obtained by fusing further the CLF descriptor and the grayscale-LBP descriptor. The CGLF is combined with the PHOG to obtain the CGLF+PHOG descriptor. See Figure 3.8 for multiple Color LBP features fusion methodology.

## 3.6 The Novel EFM-KNN Classifier

Image classification using the new descriptors introduced in the preceding section is implemented using a novel EFM-KNN classifier (Liu and Wechsler 2002), (Liu and Wechsler 2000b), which combines the Enhanced Fisher Model (EFM) and the K Nearest Neighbor (KNN) decision rule (Fukunaga 1990). Let $\mathscr{X} \in \mathbb{R}^N$ be a random vector whose covariance matrix is $\Sigma_{\mathscr{X}}$:

$$\Sigma_{\mathscr{X}} = \mathscr{E}\{[\mathscr{X} - \mathscr{E}(\mathscr{X})][\mathscr{X} - \mathscr{E}(\mathscr{X})]^t\} \tag{3.5}$$

where $\mathscr{E}(\cdot)$ is the expectation operator and $t$ denotes the transpose operation. The eigenvectors of the covariance matrix $\Sigma_{\mathscr{X}}$ can be derived by PCA:

$$\Sigma_{\mathscr{X}} = \Phi\Lambda\Phi^t \tag{3.6}$$

where $\Phi = [\phi_1\phi_2\dots\phi_N]$ is an orthogonal eigenvector matrix and $\Lambda = diag\{\lambda_1, \lambda_2, \dots, \lambda_N\}$ a diagonal eigenvalue matrix with diagonal elements in decreasing order. An important application of PCA is dimensionality reduction:

$$\mathscr{Y} = P^t \mathscr{X} \tag{3.7}$$

where $P = [\phi_1\phi_2\dots\phi_K]$, and $K < N$. $\mathscr{Y} \in \mathbb{R}^K$ thus is composed of the most significant principal components. PCA, which is derived based on an optimal representation criterion, usually does not lead to good image classification performance. To improve upon PCA, the Fisher Linear Discriminant (FLD) analysis (Fukunaga 1990) is introduced to extract the most discriminating features.

The FLD method optimizes a criterion defined on the within-class and between-class scatter matrices, $S_w$ and $S_b$ (Fukunaga 1990):

$$S_w = \sum_{i=1}^{L} P(\omega_i)\mathscr{E}\{(\mathscr{Y} - M_i)(\mathscr{Y} - M_i)^t|\omega_i\} \tag{3.8}$$

$$S_b = \sum_{i=1}^{L} P(\omega_i)(M_i - M)(M_i - M)^t \tag{3.9}$$

where $P(\omega_i)$ is *a priori* probability, $\omega_i$ represent the classes, and $M_i$ and $M$ are the means of the classes and the grand mean, respectively. The criterion the FLD method optimizes is $J_1 = tr(S_w^{-1}S_b)$, which is maximized when $\Psi$ contains the eigenvectors of the matrix $S_w^{-1}S_b$ (Fukunaga 1990):

$$S_w^{-1}S_b\Psi = \Psi\Delta \tag{3.10}$$

where $\Psi, \Delta$ are the eigenvector and eigenvalue matrices of $S_w^{-1}S_b$, respectively. The FLD discriminating features are defined by projecting the pattern vector $\mathscr{Y}$ onto the eigenvectors of $\Psi$:

$$\mathscr{Z} = \Psi^t \mathscr{Y} \tag{3.11}$$

$\mathscr{Z}$ thus is more effective than the feature vector $\mathscr{Y}$ derived by PCA for image classification.

The FLD method, however, often leads to overfitting when implemented in an inappropriate PCA space. To improve the generalization performance of the FLD method, a proper balance between two criteria should be maintained: the energy criterion for adequate image representation and the magnitude criterion for eliminating the small-valued trailing eigenvalues of the within-class scatter matrix (Liu and Wechsler 2000b). A new method, the Enhanced Fisher Model (EFM), is capable of improving the generalization performance of the FLD method (Liu and Wechsler 2000b). Specifically, the EFM method improves the generalization capability of the FLD method by decomposing the FLD procedure into a simultaneous diagonalization of the within-class and between-class scatter matrices (Liu and Wechsler 2000b). The simultaneous diagonalization is stepwise equivalent to two operations as pointed out by Fukunaga (1990): whitening the within-class scatter matrix and applying PCA to the between-class scatter matrix using the transformed data. The stepwise operation shows that during whitening the eigenvalues of the within-class scatter matrix appear in the denominator. Since the small (trailing) eigenvalues tend to capture noise (Liu and Wechsler 2000b), they cause the whitening step to fit for misleading variations, which leads to poor generalization performance. To achieve enhanced performance, the EFM method preserves a proper balance between the need that the selected eigenvalues account for most of the spectral energy of the raw data (for representational adequacy), and the requirement that the eigenvalues of the within-class scatter matrix (in the reduced PCA space) are not too small (for better generalization performance) (Liu and Wechsler 2000b).

Image classification is implemented using the EFM-KNN and the EFM-NN (nearest neighbor) classifiers, Figure 3.6 and Figure 3.8 show the fusion methodology of multiple descriptors using the EFM feature extraction methodology.

# CHAPTER 4

# EXPERIMENTAL EVALUATION OF NEW COLOR DESCRIPTORS AND THE NOVEL EFM-KNN CLASSIFIER

This chapter first describes the five datasets used in the experiments. Next, an evaluation of dense color histogram is performed, followed by an evaluation of new color SIFT descriptors and new color LBP descriptors upon the various datasets.

## 4.1 Datasets

### 4.1.1 Caltech 256 Object Categories Dataset

The Caltech 256 dataset (Griffin et al. 2007) holds 30,607 images divided into 256 categories and a clutter class. The images have high intra-class variability and high object location variability. Each category contains at least 80 images, a maximum of 827 images and the mean number of images per category is 119. The images have been collected from Google and PicSearch, they represent a diverse set of lighting conditions, poses, backgrounds, image sizes, and camera systematics. The various categories represent a wide

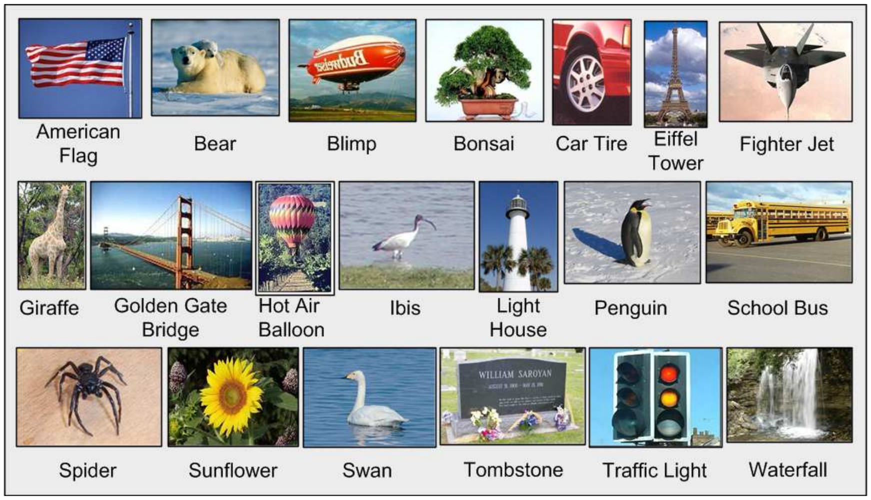


**Figure 4.1** Example images from the Caltech 256 object categories dataset.

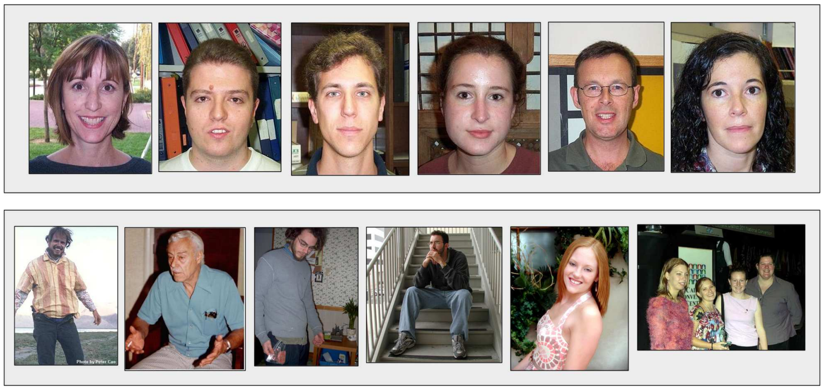


**Figure 4.2** Example images from the Faces and People classes of the Caltech 256 object categories dataset.

variety of natural and artificial objects in various settings. The images are in color, in JPEG format with only a small number of grayscale images. The average size of each image is 351x351 pixels. See Figure 4.1 for some images from the object categories and Figure 4.2 for some sample images from the Faces and People categories.

### 4.1.2 UPOL Iris Dataset

The UPOL iris dataset (Dobes et al. 2006) contains 128 unique eyes (or classes) belonging to 64 subjects with each class containing three sample images. The images of the left and right eyes of a person belong to different classes. The irises were scanned by a TOPCON TRC50IA optical device connected with a SONY DXC-950P 3CCD camera. The iris images are in 24-bit PNG format (color) and the size of each image is 576x768 pixels. See Figure 4.3 for some sample images from this dataset.

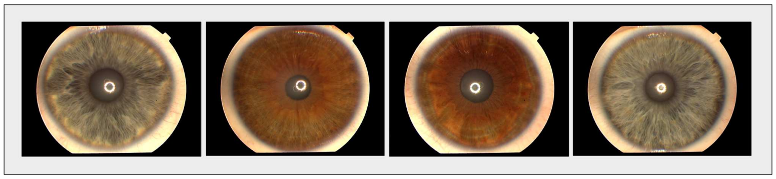


**Figure 4.3** Example images from the UPOL Iris dataset.

**Figure 4.4** Example images from the MIT Scene dataset.

### 4.1.3 MIT Scene Dataset

The MIT scene dataset (Oliva and Torralba 2001) has 2,688 images classified as eight categories: 360 coast, 328 forest, 374 mountain, 410 open country, 260 highway, 308 inside of cities, 356 tall buildings, and 292 streets. All of the images are in color, in JPEG format, and the average size of each image is 256x256 pixels. There is a large variation in light, pose and angles, along with a high intra-class variation. The sources of the images vary (from commercial databases, websites, and digital cameras) (Oliva and Torralba 2001). See Figure 4.4 for some sample images from this dataset.

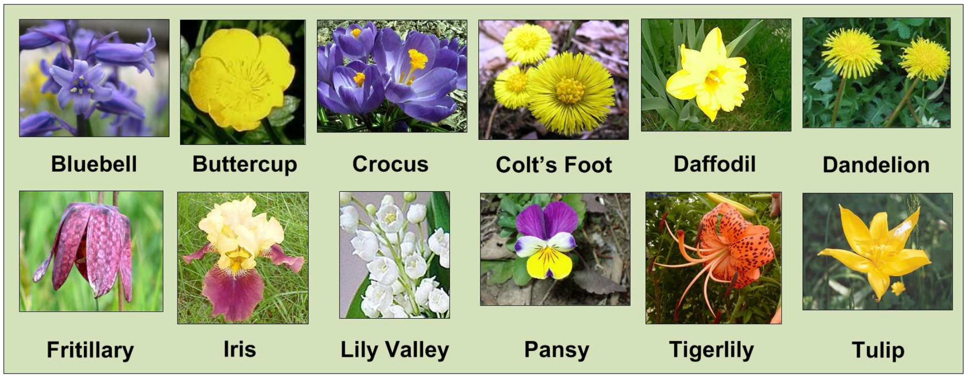


**Figure 4.5** Example images from the Oxford Flower dataset.

**Figure 4.6** Example images from the KTH-TIPS Texture dataset.

### 4.1.4 Oxford Flower Dataset

This data set consists of 17 species of flowers with 80 images in each category and a total of 1,360 images. All the images are in color in JPEG format and the average image size of each image is 560x560 pixels. There are species that have a very unique visual appearance, for example Fritillaries and Tigerlilies, as well as species with very similar appearance, for example Dandelions and Coltsfoot. There are large viewpoint, scale, and illumination variations. The large intra-class variability and the small inter-class variability make this data set very challenging. The flower categories are deliberately chosen to have some ambiguity on each aspect. For example, some classes cannot be distinguished on color alone (e.g., Dandelion and Buttercup), others cannot be distinguished on shape alone (e.g., Bluebell and Crocus). The flower images were retrieved from various websites and personal photographs (Nilsback and Zisserman 2006). Figure 4.5 shows some of the categories.

### 4.1.5 KTH-TIPS Texture Dataset

The KTH-TIPS dataset (Hayman et al. 2004) consists of ten classes of textures with 81 images per class. All the images are in color, PNG format and the maximum image size is 200x200 pixels. All ten textures have been photographed at nine scales and nine illumination conditions for each scale. Some of the classes have a very similar visual appearance,

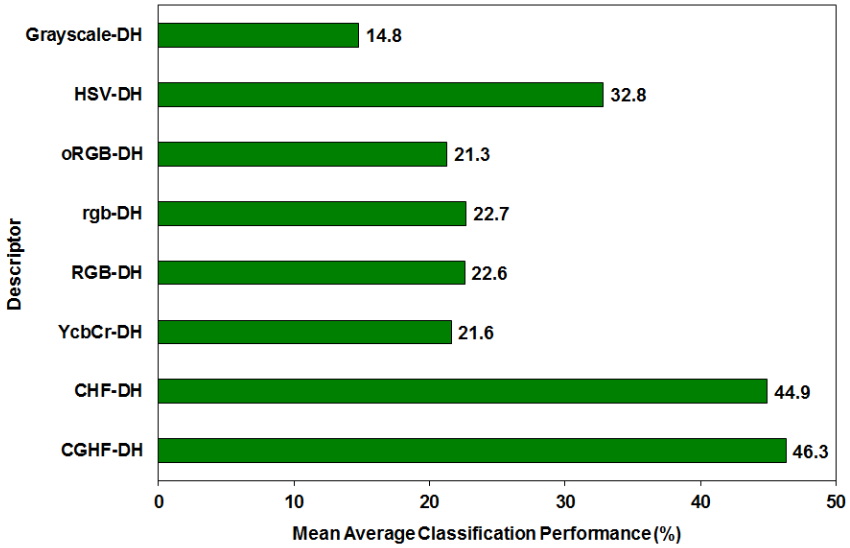


**Figure 4.7** The mean average classification performance of the eight descriptors: the oRGB-DH, the YCbCr-DH, the RGB-DH, the HSV-DH, the rgb-DH, the grayscale-DH, the CHF, and the CGHF descriptors on the Oxford Flower Dataset.

like cotton and linen, and brown bread and sponge which makes this dataset moderately challenging. The KTH-TIPS2-b (Caputo et al. 2005) dataset is a more challenging extension of the KTH-TIPS dataset with 11 classes of materials and four samples for each material. Each of these samples has 108 images with 432 images per class and a total of

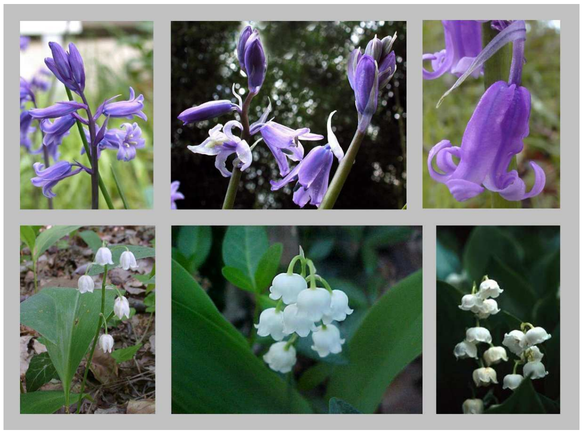


**Figure 4.8** Image recognition using the EFM-NN classifier: Examples of correctly classified images of the Bluebell (top) and Lily Valley (bottom) categories from the Oxford Flower dataset.

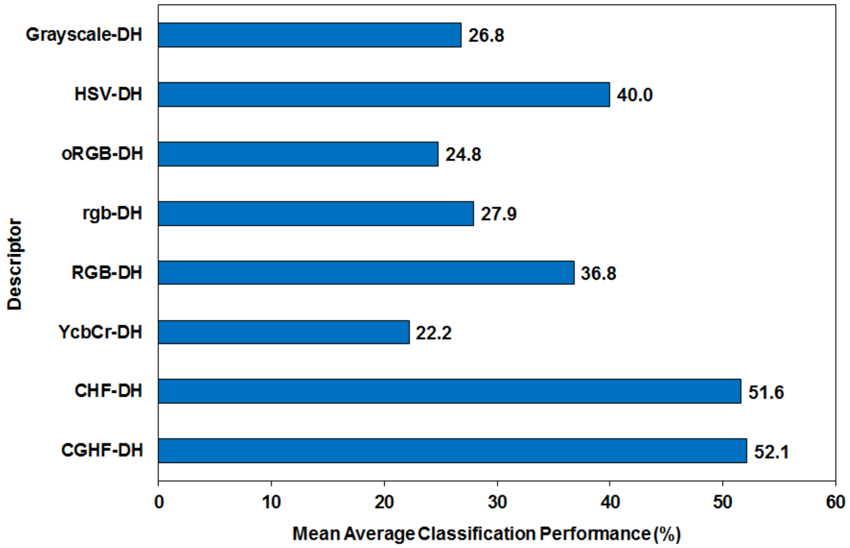


**Figure 4.9** The mean average classification performance of the eight descriptors: the oRGB-DH, the YCbCr-DH, the RGB-DH, the HSV-DH, the rgb-DH, the grayscale-DH, the CHF, and the CGHF descriptors on the MIT Scene Dataset.

4,752 images. Some of the images in the classes like wool and cotton are from differently colored samples leading to very high intra-class variation between samples, while some samples from different classes like cork and cracker have the same color and general appearance lowering the inter-class variation. See Figure 4.6 for some sample images from this dataset.

## 4.2 Evaluation of Dense Color Histogram on: Oxford Flower, MIT Scene, and Caltech 256 Datasets

The method starts with a color image as an input and first splits it into three separate color component images. Next step is to compute histograms from each of the color channels. After normalization the individual histograms are concatenated to form a compact fixed length feature vector. The following eight dense histogram (DH) descriptors are evaluated: the oRGB-DH, the YCbCr-DH, the RGB-DH, the HSV-DH, the rgb-DH, the grayscale-DH, the Color Histogram Fusion (CHF), and the Color Grayscale Histogram Fusion (CGHF)

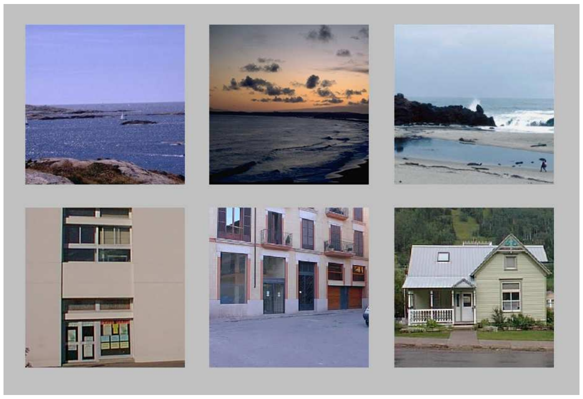


**Figure 4.10** Image recognition using the EFM-NN classifier: Examples of correctly classified images of the Coast (top) and Inside City (bottom) categories from the MIT Scene dataset.

descriptors. Learning and classification is performed with the EFM-NN (nearest neighbor) classifier.

### 4.2.1 Experimental Results on the Oxford Flower Dataset

Experimental setup consists of three sets of 40 training images and 20 test images per class. See Figure 4.7 for the classification performance across eight descriptors. On dense histogram, HSV features give a success rate of 32.8% followed by rgb-DH at 22.7% and RGB-DH at 22.6%. The YCbCr-DH and oRGB-DH come next at 21.6% and 21.3%, respectively. The grayscale-DH achieves 14.8% success rate. Combined color histograms reach a performance of 44.9% and fusing color and gray histogram reaches 46.3%.

Figure 4.8 shows some example images that were classified correctly using the EFM-NN classifier from the Bluebell and Lily Valley categories of the Oxford Flower dataset.

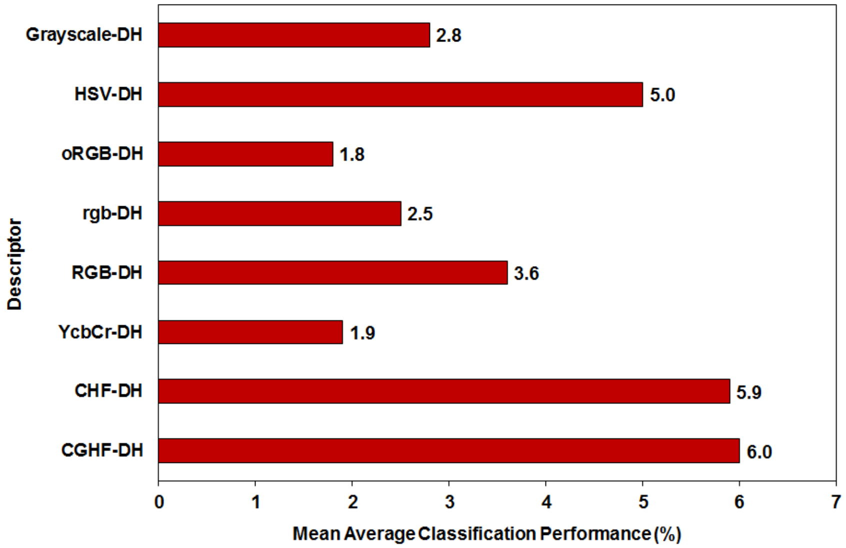


**Figure 4.11** The mean average classification performance of the eight descriptors: the oRGB-DH, the YCbCr-DH, the RGB-DH, the HSV-DH, the rgb-DH, the grayscale-DH, the CHF, and the CGHF descriptors on the Caltech 256 Dataset.

### 4.2.2 Experimental Results on the MIT Scene Dataset

From each class, 100 images for training, 50 images for testing are selected. Five-fold cross validation is performed on the following dense histogram descriptors: the oRGB-DH, the YCbCr-DH, the RGB-DH, the HSV-DH, the rgb-DH, the grayscale-DH, the Color Histogram Fusion (CHF), and the Color Grayscale Histogram Fusion (CGHF) descriptors. See Figure 4.9 for the classification performance across eight descriptors. HSV features give a success rate of 40% followed by RGB-DH at 36.8% and rgb-DH at 27.9%. The grayscale-DH achieves 26.8% success rate. The oRGB-DH and YCbCr-DH come next at 24.8% and 22.2%, respectively. Combined color histograms reach a performance of 51.6% and fusing color and gray histogram reaches 52.1%.

Figure 4.10 shows some example images that were classified correctly using the EFM-NN classifier from the Coast and Inside City categories of the MIT Scene dataset.

### 4.2.3 Experimental Results on the Caltech 256 Dataset

On the Caltech 256 dataset (Griffin et al. 2007) experiments are performed to evaluate the performance of the following dense histogram descriptors: the oRGB-DH, the YCbCr-DH, the RGB-DH, the HSV-DH, the rgb-DH, the grayscale-DH, the Color Histogram Fusion (CHF), and the Color Grayscale Histogram Fusion (CGHF) descriptors. From each class, 50 images are used for training and 25 images for testing. Figure 4.11 shows the detailed performance of the EFM-NN classification technique on this dataset. HSV features give a success rate of 5% followed by RGB-DH at 3.6% and grayscale-DH at 2.8%. The rgb-DH, YCbCr-DH and oRGB-DH come next at 2.5%, 1.9% and 1.8%, respectively. Combined color histograms reach a performance of 5.9% and fusing color and gray histogram reaches 6%. This dataset has very high intra-class variability, low inter-class variability, and in several cases the object occupies a small portion of the full image. This makes it challenging for the dense histogram to attain good success. Such issues are later addressed in this dissertation and new robust methods with improved classification performance are proposed.

Figure 4.12 shows some example images that were classified correctly using the EFM-NN classifier from the Bat and Swiss Army Knife categories of the Caltech 256 dataset.

## 4.3 Evaluation of SIFT Descriptors on the Caltech 256 and the UPOL Iris Datasets

### 4.3.1 Experimental Methodology

The following two publicly accessible datasets are used to evaluate the proposed descriptors and classification method: the Caltech 256 object categories (Griffin et al. 2007) and the UPOL iris dataset (Dobes et al. 2006). In order to make a thorough comparative assessment of the proposed descriptors and methods; from the above two databases, three datasets are generated: the Biometric 20, the Biometric 100, and the Biometric 257. The Biometric

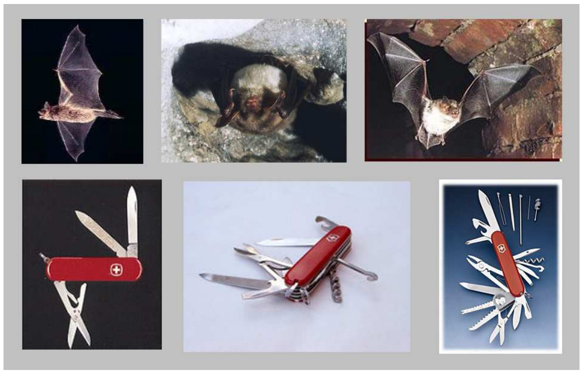


**Figure 4.12** Image recognition using the EFM-NN classifier: Examples of correctly classified images of the Bat (top) and Swiss Army Knife (bottom) categories from the Caltech 256 dataset.

20 dataset with 20 categories includes the Iris category from the UPOL dataset, Faces and People categories and 17 randomly chosen categories from the Caltech 256 dataset. The Biometric 100 dataset with 100 categories includes the Iris category from the UPOL dataset, Faces and People categories and 97 randomly chosen categories from the Caltech 256 dataset. The Biometric 257 dataset with 257 categories includes all categories from the UPOL dataset and the Caltech 256 dataset. The latter two datasets are of high difficulty due to the large number of classes with high intra-class and low inter-class variations.

The classification task is to assign each test image to one of a number of categories. The performance is measured using a confusion matrix, and the overall performance rates are measured by the average value of the diagonal entries of the confusion matrix. Each dataset is split randomly into two separate sets of images for training and testing. From each class 60 images for training and 20 images for testing are randomly selected. There is no overlap in the images selected for training and testing. The classification scheme on the dataset compares the overall and category wise performance of ten different descriptors: the oRGB-SIFT, the YCbCr-SIFT, the RGB-SIFT, the HSV-SIFT, the rgb-SIFT, the PHOG, the grayscale-SIFT, the CSF, the CGSF, and the CGSF+PHOG descriptors. Classification is implemented using a novel EFM-KNN classifier, which combines the Enhanced Fisher Model (EFM) and the K Nearest Neighbor (KNN) decision rule.

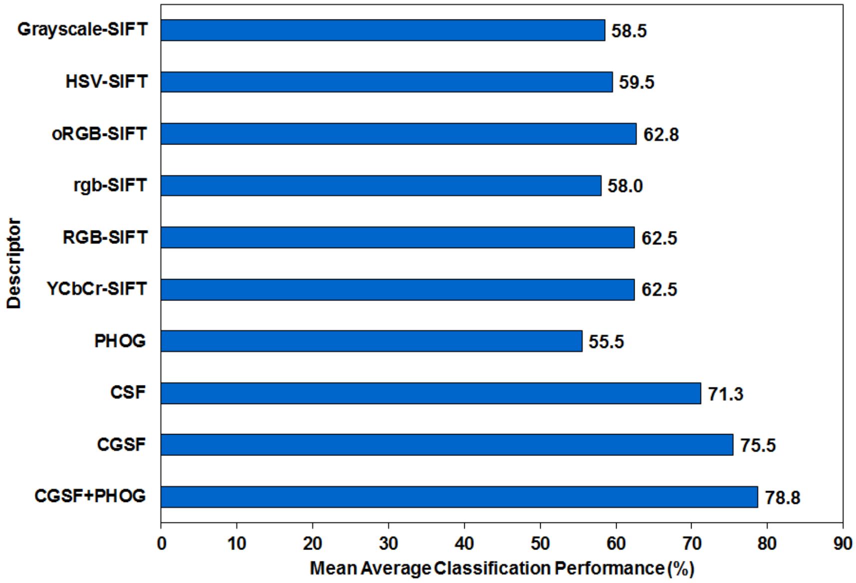


**Figure 4.13** The mean average classification performance of the ten descriptors: the oRGB-SIFT, the YCbCr-SIFT, the RGB-SIFT, the HSV-SIFT, the rgb-SIFT, the grayscale-SIFT, the CSF, the CGSF, and the CGSF+PHOG descriptors on the Biometric 20 dataset.

### 4.3.2 Experimental Results on the Biometric 20 Categories Dataset

*Evaluation of Overall Classification Performance of Descriptors with the EFM-KNN Classifier*

The first set of experiments assesses the overall classification performance of the ten descriptors on the Biometric 20 Dataset with 20 categories. Note that for each category a five-fold cross validation is implemented for each descriptor using the EFM-KNN classification technique to derive the average classification performance. As a result, each descriptor yields 20 average classification rates corresponding to the 20 image categories. The mean value of these 20 average classification rates is defined as the mean average classification performance for the descriptor. Figure 4.13 shows the mean average classification performance of the eight descriptors: the oRGB-SIFT, the YCbCr-SIFT, the RGB-SIFT, the HSV-SIFT, the rgb-SIFT, the grayscale-SIFT, the PHOG, the CSF, the CGSF, and the CGSF+PHOG descriptors.

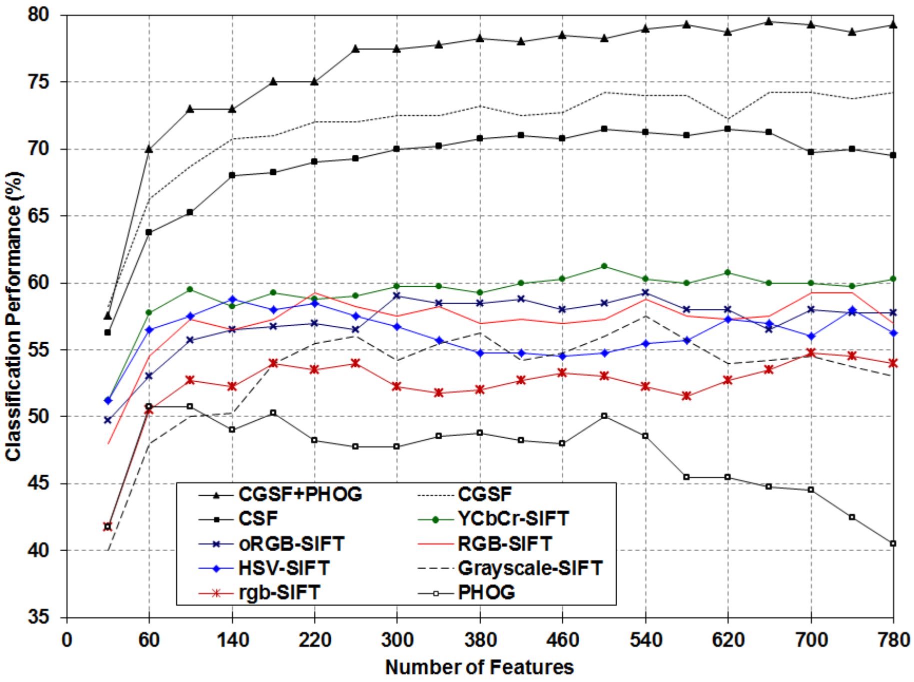


**Figure 4.14** Classification results using the PCA method across the ten descriptors with varying number of features on the Biometric 20 dataset.

The best recognition rate that is obtained is 78.8% from the CGSF+PHOG, which is a very respectable value for a dataset of this size and complexity. The oRGB-SIFT achieves the classification rate of 62.8%. It outperforms other two color descriptors (HSV-SIFT and rgb-SIFT) while showing roughly the same success rate as the YCbCr-SIFT and RGB-SIFT, both are in second place with 62.5%. It is noted that fusion of the color SIFT descriptors (CSF) improves upon the grayscale-SIFT by a huge 12.8% margin. The grayscale-SIFT descriptor improves the fusion (CGSF) result by a good 4.2% margin upon the CSF descriptor.

### *Evaluation of PCA and EFM-KNN Results upon Varying Number of Features*

The second set of experiments evaluates the classification performance using the PCA and the EFM-KNN methods respectively by varying the number of features over the following ten descriptors: CGSF+PHOG, CGSF, CSF, YCbCr-SIFT, oRGB-SIFT, RGB-

**Table 4.1** Comparison of Classifiers across Ten Descriptors (%) on the Biometric 20 Dataset

| Descriptor | PCA | EFM-NN | EFM-KNN |
|---|---|---|---|
| RGB-SIFT | 58.0 | 60.5 | **62.5** |
| HSV-SIFT | 55.0 | 57.8 | **59.5** |
| rgb-SIFT | 52.3 | 56.3 | **58.0** |
| oRGB-SIFT | 59.3 | 61.5 | **62.8** |
| YCbCr-SIFT | 62.0 | 60.8 | **62.5** |
| Grayscale-SIFT | 56.0 | 57.5 | **58.5** |
| PHOG | 51.0 | 54.8 | **55.5** |
| CSF | **71.5** | 71.3 | 71.3 |
| CGSF | 74.5 | **75.5** | **75.5** |
| CGSF+PHOG | **79.5** | 78.3 | 78.8 |

SIFT, HSV-SIFT, Grayscale-SIFT, PHOG, and rgb-SIFT. Also, a comparison between the overall success rates of the three classification techniques: PCA, EFM-NN (nearest neighbor), and EFM-KNN can be seen from Table 4.1.

Classification performance is computed for up to 780 features with the PCA method. From Figure 4.14 it can be seen that the success rate for the CGSF+PHOG stays consistently above that of the CGSF and CSF over varying number of features. These three descriptors show an increasing trend till 660 features and start to dip slightly thereafter. The YCbCr-SIFT and oRGB-SIFT show a similar increasing trend and decline only towards the latter half. The HSV-SIFT and RGB-SIFT dip in the middle and gain steadily thereafter. Performance of the grayscale-SIFT varies more sharply over the increasing number of features peaking at 540 features.

Using the EFM-KNN method, success rates for up to 19 features are computed. From Figure 4.15 it can be seen that the success rate for the CGSF+PHOG stays consistently above that of the CGSF and the CSF over varying number of features and peaks between 18 and 19 features. These two descriptors by and large show an increasing trend throughout. The oRGB-SIFT, YCbCr-SIFT, and RGB-SIFT show an increasing trend and outperform the rest of the descriptors. The grayscale-SIFT maintains its higher perfor-

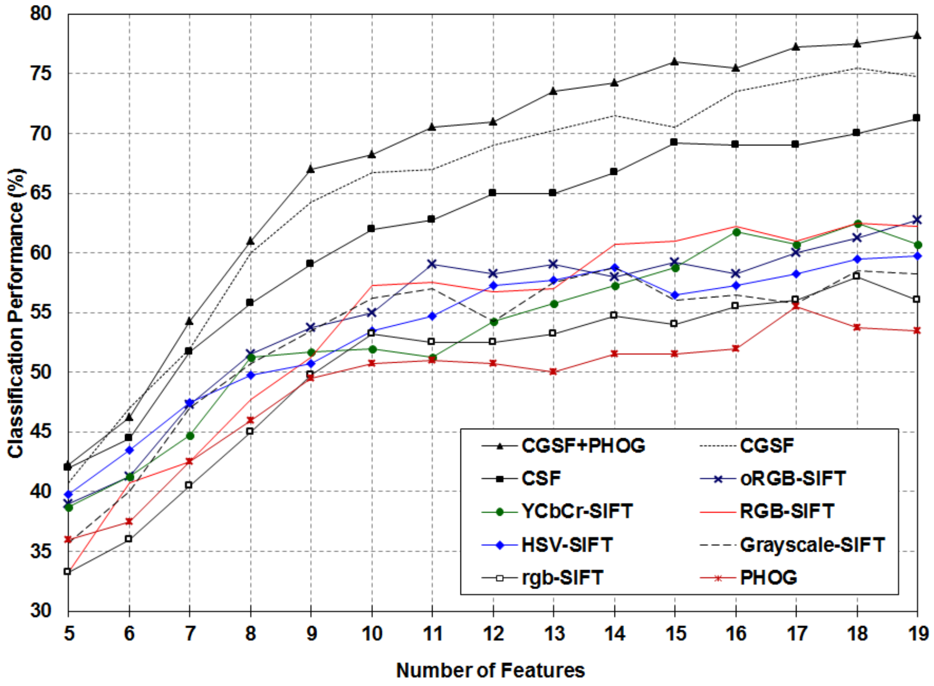


**Figure 4.15** Classification results using the EFM-KNN method across the ten descriptors with varying number of features on the Biometric 20 dataset.

mance over the rgb-SIFT for the varying number of features.

*Evaluation of Descriptors and Classifier on Individual Image Categories*

The third set of experiments assesses the ten descriptors using the EFM-KNN classifier on individual image categories. Here a detailed analysis is performed for the performance of the descriptors with the EFM-KNN classifier over all the twenty image categories. First the classification results on the three biometric categories are presented. Table 4.2 shows that the Iris category has a 100% recognition rate across all the descriptors. For the Faces category, the color SIFT descriptors outperform the grayscale-SIFT descriptor by 15% to 20% and the fusion of all color descriptors (CSF) reaches a 100% success rate. The People category achieves a high success rate of 80% with the CGSF+PHOG, which is a respectable recognition rate when considering very high intra-class variability due to the challenging background, variable postures, variable appearance, occlusion, multiple

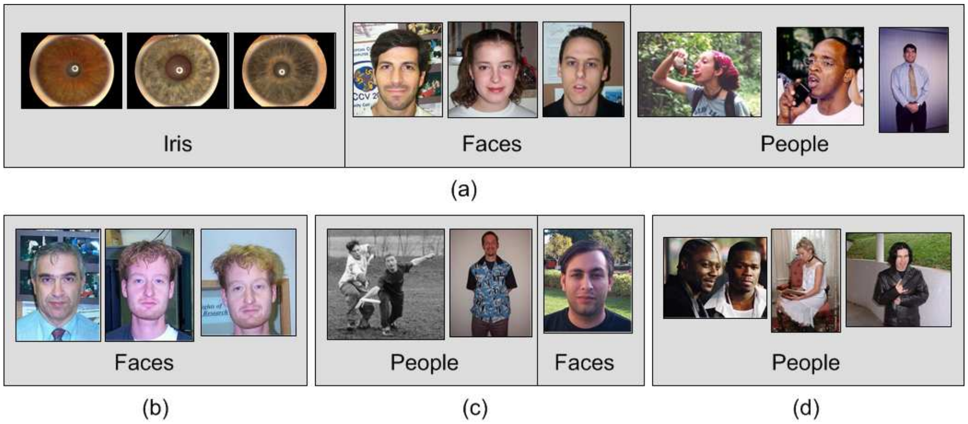


**Figure 4.16** Image recognition using the EFM-KNN classifier on the Biometric 20 dataset: (a) examples of the correctly classified images from the three biometric image categories; (b) images unrecognized using the grayscale-SIFT descriptor but recognized using the oRGB-SIFT descriptor; (c) images unrecognized using the oRGB-SIFT descriptor but recognized using the CSF descriptor; (d) images unrecognized using the CSF but recognized using the CGSF+PHOG descriptor.

humans in the same image, and different illumination conditions. Fusion of the individual color SIFT descriptors (CSF) improves the classification performance, which indicates that various color descriptors are not redundant for recognition of the People category.

The average success rate for the CGSF+PHOG descriptor over the top 15 categories is 89.7% with only five categories below the 80% mark. Individual color SIFT features improve upon the grayscale-SIFT features for most of the categories, in particular for the Grapes, the Roulette wheel, the Waterfall, and the Rotary phone categories. The CSF descriptor almost always improves upon the grayscale-SIFT descriptor, with the exception of only a few categories where it performs at par or slightly below. The CGSF descriptor either is at par or improves upon the CSF descriptor for all categories with the exception of the Waterfall and snake categories.

*Evaluation of Descriptors and Classifier Based on Correctly Recognized Images*

The final set of experiments further assesses the performance of the descriptors based on the correctly recognized images. Figure 4.16(a) for some examples of the cor-

**Table 4.2** Category Wise Descriptor Performance (%) Split-out with the EFM-KNN Classifier on the Biometric 20 Dataset (Note That the Categories are Sorted on the CGSF+PHOG Results)

| Category | CGSF+ PHOG | CGSF | CSF | oRGB SIFT | YCbCr SIFT | RGB SIFT | Gray SIFT | PHOG |
|---|---|---|---|---|---|---|---|---|
| iris | **100** | **100** | **100** | **100** | **100** | **100** | **100** | **100** |
| faces | **100** | **100** | **100** | 95 | 90 | 95 | 75 | 95 |
| people | **80** | 70 | 60 | 40 | 40 | 35 | 45 | 20 |
| cartman | **100** | **100** | 95 | 90 | **100** | 95 | 90 | 60 |
| grand piano | 95 | **100** | 95 | 85 | 85 | 70 | 90 | 85 |
| grapes | **95** | 90 | 90 | 70 | **95** | 80 | 60 | 70 |
| roulette wheel | **95** | **95** | **95** | 90 | 75 | 85 | 75 | 40 |
| waterfall | **95** | 90 | **95** | 80 | 75 | 85 | 75 | 65 |
| rainbow | 90 | 85 | 80 | 55 | 35 | 60 | 75 | **95** |
| cockroach | **85** | 75 | 70 | 50 | 50 | 60 | 55 | 20 |
| human skeleton | 85 | **90** | 80 | 70 | 60 | 75 | 60 | 40 |
| laptop | 85 | 85 | 80 | 75 | **90** | 70 | 65 | 65 |
| centipede | **80** | 75 | 65 | 55 | 60 | 55 | 45 | 20 |
| mountain bike | 80 | 80 | 80 | 75 | 70 | 80 | **85** | 75 |
| rotary phone | **80** | **80** | **80** | 60 | 75 | 45 | 45 | 65 |
| buddha | **70** | 50 | 40 | 40 | 65 | 45 | 45 | **70** |
| owl | **60** | **60** | 45 | 40 | 45 | 30 | 25 | 35 |
| jesus christ | **50** | 40 | 30 | 35 | 10 | 30 | 20 | 45 |
| snake | 25 | 20 | 25 | 25 | 20 | **30** | 15 | 25 |
| wheelbarrow | **25** | **25** | 20 | **25** | 10 | **25** | **25** | 20 |
| **Mean** | **78.8** | **75.5** | **71.3** | **62.8** | **62.5** | **62.5** | **58.5** | **55.5** |

rectly classified images from the Iris, Faces, and People categories. Notice the high intra-class variability for the Faces and People classes. Figure 4.16(b) shows some example images from the Faces class that are not recognized by the EFM-KNN classifier using the grayscale-SIFT descriptor but are correctly recognized using the oRGB-SIFT descriptor. This reaffirms the importance of color and the distinctiveness of the oRGB-SIFT descriptor for image category recognition. Figure 4.16(c) shows some images that are not recognized by the EFM-KNN classifier using the oRGB-SIFT descriptor but are correctly recognized by using the CSF descriptor. Figure 4.16 (d) shows images unrecognized using the CSF but recognized using the CGSF+PHOG descriptor.

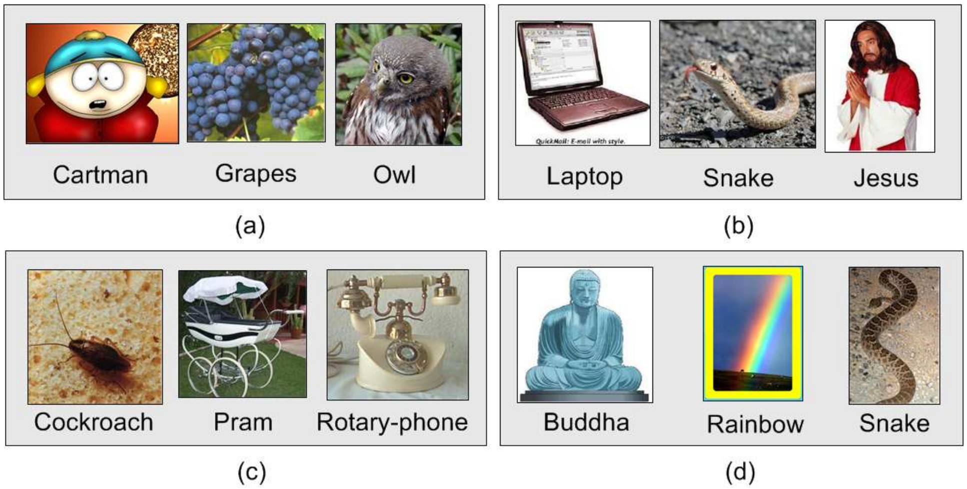


**Figure 4.17** Image recognition using the EFM-KNN classifier on the Biometric 20 dataset: (a) example images unrecognized using the grayscale-SIFT descriptor but recognized using the oRGB-SIFT descriptor; (b) example images unrecognized using the oRGB-SIFT descriptor but recognized using the CSF descriptor; (c) example images unrecognized using the CSF but recognized using the CGSF+PHOG descriptor. (d) Example images unrecognized using the EFM-KNN but recognized using the PCA with the CGSF+PHOG descriptor.

Figure 4.17(a) shows some example images that are not recognized by the EFM-KNN classifier using the grayscale-SIFT descriptor but are correctly recognized using the oRGB-SIFT descriptor. Figure 4.17(b) displays some images that are not recognized by the EFM-KNN classifier using the oRGB-SIFT descriptor but are correctly recognized using the CSF descriptor. Figure 4.17(c) shows example images unrecognized using the CSF but recognized using the CGSF+PHOG descriptor. Figure 4.17(d) shows example images unrecognized using the EFM-KNN but recognized using the PCA with the CGSF+PHOG descriptor.

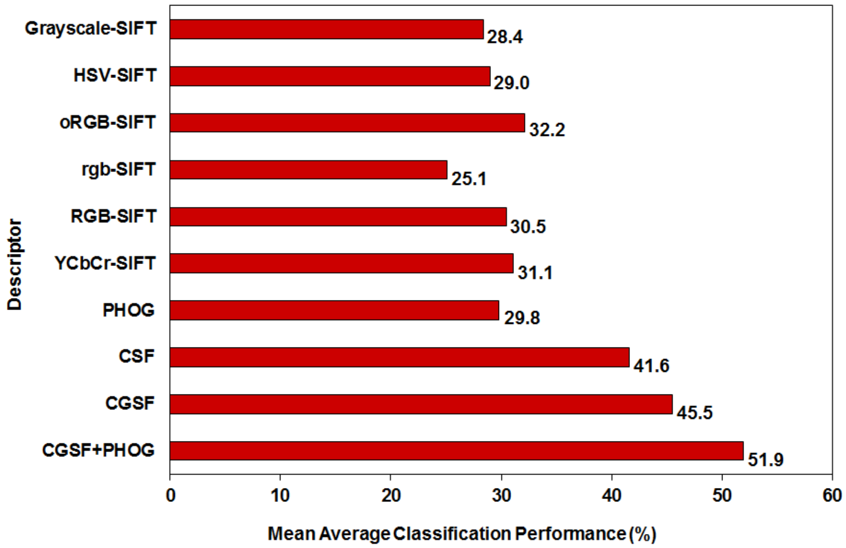


**Figure 4.18** The mean average classification performance of the ten descriptors: the oRGB-SIFT, the YCbCr-SIFT, the RGB-SIFT, the HSV-SIFT, the rgb-SIFT, the grayscale-SIFT, the PHOG, the CSF, the CGSF, and the CGSF+PHOG descriptors on the Biometric 100 dataset.

### 4.3.3 Experimental Results on the Biometric 100 Categories Dataset

*Evaluation of Overall Classification Performance of Descriptors with the EFM-NN Classifier*

The first set of experiments assesses the overall classification performance of the ten descriptors on the Biometric 100 Dataset with 100 categories. Note that for each category a five-fold cross validation is implemented for each descriptor using the EFM-NN classification technique to derive the average classification performance. As a result, each descriptor yields 100 average classification rates corresponding to the 100 image categories. The mean value of these 100 average classification rates is defined as the mean average classification performance for the descriptor. Figure 4.23 shows the mean average classification performance of various descriptors.

The best recognition rate that is obtained is 51.9% from the CGSF+PHOG, which is a very respectable value for a dataset of this size and complexity. The oRGB-SIFT achieves the classification rate of 32.2% and hence once again outperforms other color descriptors. The success rate for YCbCr-SIFT comes in second place with 31.1% followed by the RGB-SIFT at 30.5%. Fusion of color SIFT descriptors (CSF) improves over the grayscale-SIFT by a huge 13.2%. Again, the grayscale-SIFT shows more distinctiveness than the rgb-SIFT, and improves the fusion (CGSF) result by a good 3.9% over the CSF. Fusing the CGSF and PHOG further improves the recognition rate over the CGSF by 6.4%.

### *Comparison of PCA and EFM-NN Results*

The second set experiments compares the classification performance of the PCA and the EFM-NN (nearest neighbor) classifiers. Table 4.3 shows the results of the two classifiers across various descriptors. It can be seen that the EFM-NN technique improves over the PCA technique by 2% to 3% on the color SIFT descriptors, by 2.1% on the grayscale-SIFT, and by 1.9% on the PHOG. The improvement on fused descriptors is in the range of 1%-2.6%. These results reaffirm the superiority of the EFM-NN classifier over the PCA technique.

**Table 4.3** Comparison of Classifiers across Ten Descriptors (%) on the Biometric 100 Dataset

| Descriptor | PCA | EFM-NN |
|---|---|---|
| RGB-SIFT | 27.9 | **30.5** |
| HSV-SIFT | 26.1 | **29.0** |
| rgb-SIFT | 23.1 | **25.1** |
| oRGB-SIFT | 29.4 | **32.2** |
| YCbCr-SIFT | 28.2 | **31.1** |
| SIFT | 26.3 | **28.4** |
| PHOG | 28.0 | **29.8** |
| **CSF** | 40.2 | **41.6** |
| **CGSF** | 44.6 | **45.5** |
| **CGSF+PHOG** | 49.4 | **51.9** |

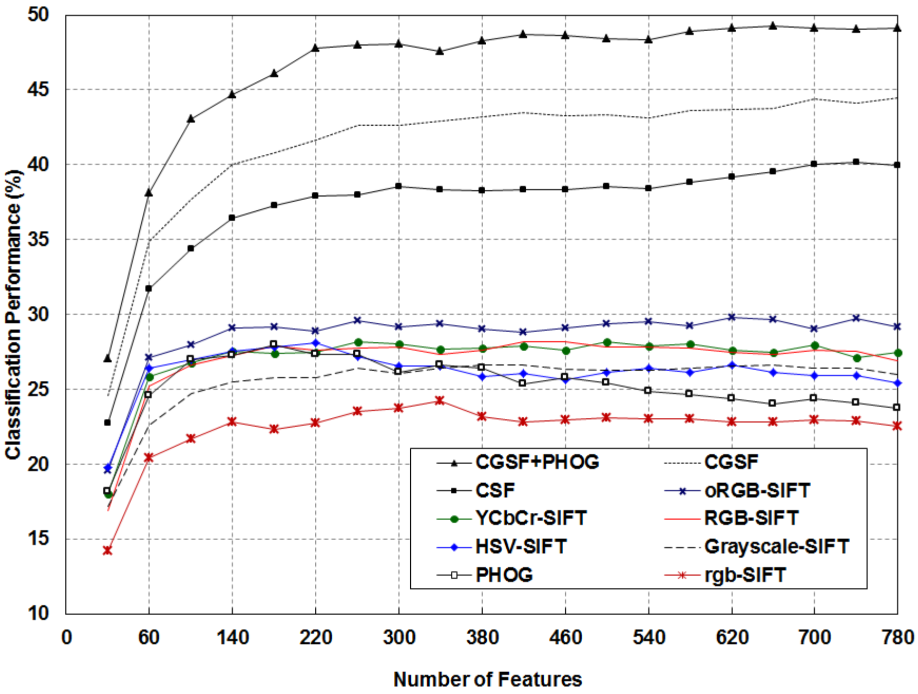


**Figure 4.19** Classification results using the PCA method across the ten descriptors with varying number of features on the Biometric 100 dataset.

### *Evaluation of PCA and EFM-NN Results upon Varying Number of Features*

The third set of experiments evaluates the classification performance using the PCA and the EFM-NN methods respectively by varying the number of features over the following ten descriptors: CGSF+PHOG, CGSF, CSF, YCbCr-SIFT, oRGB-SIFT, RGB-SIFT, HSV-SIFT, Grayscale-SIFT, rgb-SIFT, and PHOG.

Classification performance is computed for up to 780 features with the PCA classifier. From Figure 4.19 it can be seen that the success rate for the CGSF+PHOG stays consistently above that of the CGSF and CSF over varying number of features and peaks at around 660 features. These three descriptors show an increasing trend overall and flatten out toward the end. The oRGB-SIFT, YCbCr-SIFT, RGB-SIFT, and grayscale-SIFT show a similar increasing trend and flatten toward the end. The oRGB-SIFT descriptor consistently stays above other color SIFT descriptors. The HSV-SIFT and PHOG peak in the first half of the graph and show a declining trend thereafter. The grayscale-SIFT maintains its

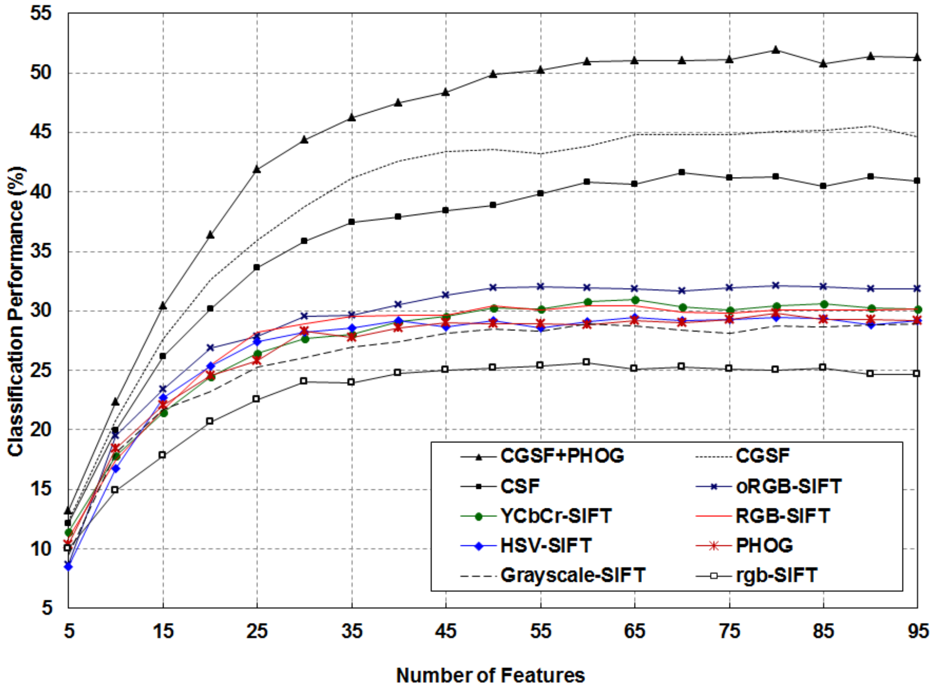


**Figure 4.20** Classification results using the EFM-NN method across the ten descriptors with varying number of features on the Biometric 100 dataset.

superior performance upon the rgb-SIFT on the varying number of features.

With the EFM-NN classifier, the success rates are computed for up to 95 features. From Figure 4.20 it can be seen that the success rate for the CGSF+PHOG stays consistently above that of the CGSF and CSF over varying number of features and peaks at about 80 features. These three descriptors show an increasing trend throughout and tend to flatten above 65 features. The oRGB-SIFT consistently stays above the rest of the descriptors. The grayscale-SIFT improves over the rgb-SIFT but falls below the PHOG.

*Evaluation of Descriptors and Classifier on Individual Image Categories*

The fourth set of experiments assesses the eight descriptors using the EFM-NN classifier on individual image categories. Here a detailed analysis of the performance of the descriptors is performed with the EFM-NN classifier over all the 100 image categories. First the classification results on the three biometric categories are presented. From Ta-

**Table 4.4** Category Wise Descriptor Performance (%) Split-out with the EFM-NN Classifier on the Biometric 100 Dataset (Note That the Categories are Sorted on the CGSF+PHOG Results)

| Category | CGSF+ PHOG | CGSF | CSF | oRGB SIFT | YCbCr SIFT | RGB SIFT | Gray SIFT | PHOG |
|---|---|---|---|---|---|---|---|---|
| iris | **100** | **100** | **100** | **100** | **100** | **100** | **100** | **100** |
| faces | **95** | 90 | 90 | 90 | **95** | 90 | 85 | **95** |
| people | **40** | **40** | 25 | 20 | 20 | 15 | 30 | 10 |
| hibiscus | **100** | **100** | 95 | 70 | 80 | 85 | 75 | 55 |
| french horn | **95** | 85 | 85 | 85 | 65 | 80 | 90 | 20 |
| leopards | 95 | 90 | **100** | 90 | 95 | 95 | **100** | 90 |
| saturn | **95** | **95** | **95** | **95** | 85 | 90 | **95** | 55 |
| school bus | **95** | **95** | **95** | 75 | 85 | **95** | 80 | 60 |
| swiss army knife | **95** | 90 | 80 | 65 | 75 | 65 | 65 | 25 |
| watch | **95** | 60 | 55 | 45 | 40 | 45 | 30 | 85 |
| zebra | **95** | 80 | 60 | 60 | 35 | 40 | 45 | 60 |
| galaxy | **90** | 85 | 85 | 85 | 70 | 65 | 80 | 15 |
| american flag | **85** | **85** | 80 | 55 | 75 | 65 | 40 | 5 |
| cartman | **85** | 75 | 75 | 40 | 55 | 65 | 55 | 30 |
| desk-globe | **85** | 75 | 75 | 60 | 65 | 65 | 45 | 80 |
| harpsichord | **85** | 80 | **85** | 50 | 80 | 70 | 60 | 55 |
| ketch | **85** | **85** | **85** | 45 | 50 | 45 | 50 | 70 |
| roulette wheel | **85** | 80 | 75 | 70 | 65 | 75 | 55 | 35 |
| hawksbill | **80** | **80** | 75 | 55 | 60 | 70 | 55 | 40 |
| iris flower | **80** | 75 | 75 | 35 | 65 | **80** | 65 | 30 |
| mountain bike | 80 | 85 | **90** | 70 | 65 | 85 | 75 | 70 |

ble 4.4 it can be seen that the Iris has a 100% recognition rate across all the descriptors. For the Faces category the color SIFT descriptors outperform the grayscale-SIFT by 5% to 10% and the fusion of all descriptors (CGSF+PHOG) reaches a 95% success rate. The People category achieves a high success rate of 40% with the CGSF+PHOG, surprisingly grayscale-SIFT outperforms the color descriptors by 10% to 20%. The fusion of individual SIFT descriptors (CGSF) improves the classification performance for the People category.

The average success rate for the CGSF+PHOG over the top 20 categories is 90% with ten categories above the 90% mark. Individual color SIFT features improve upon the grayscale-SIFT on most of the categories, in particular for the Swiss army knife, Watch,

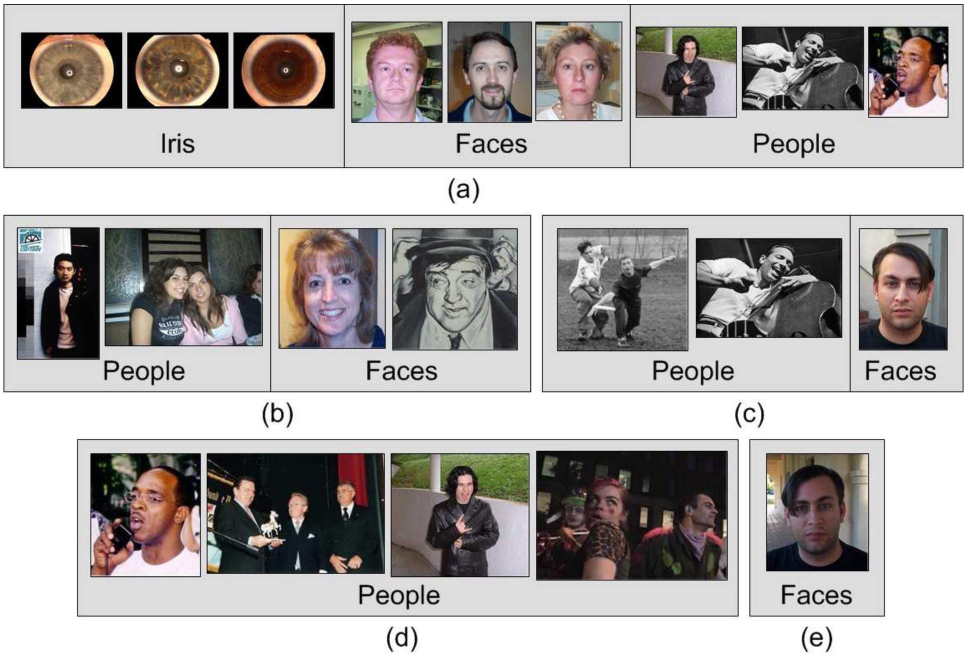


**Figure 4.21** Image recognition using the EFM-NN classifier on the Biometric 100 dataset: (a) examples of the correctly classified images from the three biometric image categories; (b) images unrecognized using the grayscale-SIFT descriptor but recognized using the oRGB-SIFT descriptor; (c) images unrecognized using the oRGB-SIFT descriptor but recognized using the CSF descriptor; (d) images unrecognized using the CSF but recognized using the CGSF+PHOG; (e) images unrecognized by PCA but recognized by EFM-NN on the CGSF+PHOG descriptor.

American flag, and Roulette wheel categories. The CSF almost always improves over the grayscale-SIFT, with the exception of People and French horn categories. The CGSF either is at par or improves over the CSF for all categories with the exception of two of the categories. Most categories perform at their best when the PHOG is combined with the CGSF.

*Evaluation of Descriptors and Classifier Based on Correctly Recognized Images*

The final set of experiments further assesses the performance of the descriptors based on the correctly recognized images. See Figure 4.21(a) for some examples of the correctly classified images from the Iris, Faces, and People categories. Once again no-

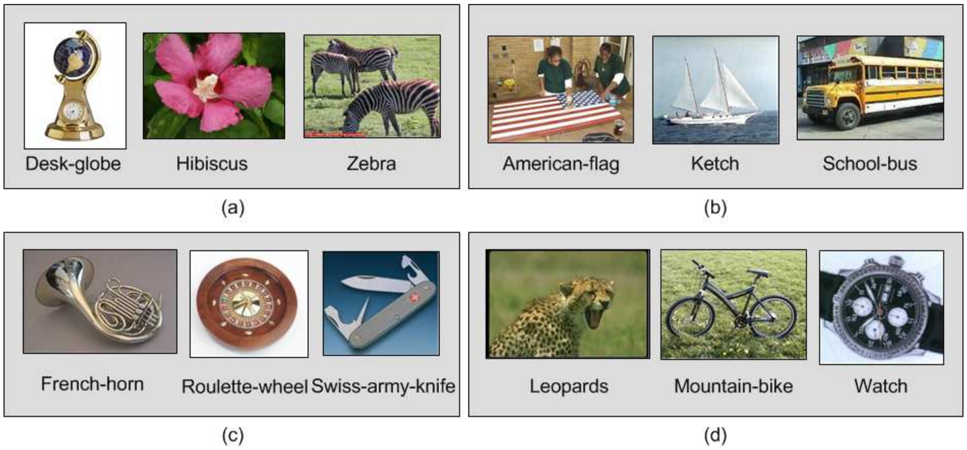


**Figure 4.22** Image recognition using the EFM-NN classifier on the Biometric 100 dataset: (a) example images unrecognized using the grayscale-SIFT descriptor but recognized using the oRGB-SIFT descriptor; (b) example images unrecognized using the oRGB-SIFT descriptor but recognized using the CSF descriptor; (c) images unrecognized using the CSF but recognized using the CGSF+PHOG. (d) Images unrecognized using the PCA but recognized using the EFM-NN on the CGSF+PHOG descriptor.

tice the high intra-class variability in the recognized images for the Faces and People class. Figure 4.21(b) shows some images from the Faces and People categories that are not recognized by the grayscale-SIFT but are correctly recognized by the oRGB-SIFT. Figure 4.21(c) shows some images that are not recognized by the oRGB-SIFT but are correctly recognized by the CSF. Figure 4.21(d) shows some images from the People class, which are not recognized by the CSF but are correctly recognized by the CGSF+PHOG descriptor. Thus, combining grayscale-SIFT, PHOG, and CSF lends more discriminative power. Lastly in Figure 4.21(e) a face image unrecognized by the PCA but recognized by the EFM-NN classifier on the CGSF+PHOG descriptor.

See Figure 4.22(a) for some examples of the images unrecognized by the grayscale-SIFT but are correctly recognized by the oRGB-SIFT. Figure 4.22(b) shows some images that are not recognized by the oRGB-SIFT but are correctly recognized by the CSF. Figure 4.22(c) shows some images unrecognized by the CSF but are correctly recognized by the CGSF+PHOG descriptor. Lastly in Figure 4.22(d) images unrecognized by the PCA but recognized by the EFM-NN classifier on the CGSF+PHOG descriptor.

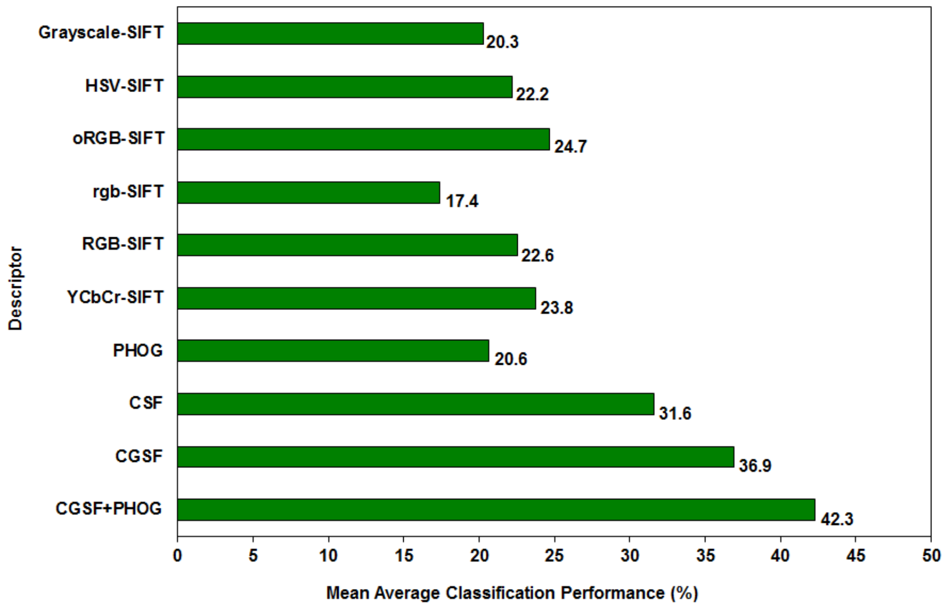


**Figure 4.23** The mean average classification performance of the ten descriptors: the oRGB-SIFT, the YCbCr-SIFT, the RGB-SIFT, the HSV-SIFT, the rgb-SIFT, the grayscale-SIFT, the PHOG, the CSF, the CGSF, and the CGSF+PHOG descriptors on the Biometric 257 dataset.

### 4.3.4 Experimental Results on the Biometric 257 Categories and the Caltech 256 Datasets

*Evaluation of Overall Classification Performance of Descriptors with the EFM-NN Classifier*

The first set of experiments assesses the overall classification performance of the ten descriptors on the Biometric 257 Dataset with 257 categories. Note that for each category a five-fold cross validation is implemented for each descriptor using the EFM-NN classification technique to derive the average classification performance. As a result, each descriptor yields 257 average classification rates corresponding to the 257 image categories. The mean value of these 257 average classification rates is defined as the mean average classification performance for the descriptor. Figure 4.23 shows the mean average classification performance of various descriptors.

The best recognition rate that is obtained is 42.3% from the CGSF+PHOG, which is a very respectable value for a dataset of this size and complexity. The oRGB-SIFT achieves the classification rate of 24.7%. It outperforms the other color SIFT descriptors. It is noted that fusion of the color SIFT descriptors (CSF) improves upon the grayscale-SIFT by a huge 11.3% margin. The grayscale-SIFT descriptor improves the fusion (CGSF) result by a good 5.3% margin upon the CSF descriptor.

### *Comparison of PCA and EFM-NN Results*

The second set of experiments compares the classification performance of the PCA and the EFM-NN (nearest neighbor) classifiers. Table 4.5 shows the results of the two classifiers across various descriptors. It can be seen that the EFM-NN technique improves over the PCA technique by 1.5% to 2.6% upon the color SIFT descriptors, by 2.1% upon the grayscale-SIFT descriptor, and by 1.1% upon the PHOG descriptor. The improvement on fused descriptors is in the range of 1.6% to 2.2%.

**Table 4.5** Comparison of Classifiers across Ten Descriptors (%) on the Biometric 257 Dataset

| Descriptor | PCA | EFM-NN |
|---|---|---|
| RGB-SIFT | 20.5 | **22.6** |
| HSV-SIFT | 20.2 | **22.2** |
| rgb-SIFT | 16.0 | **17.4** |
| oRGB-SIFT | 22.4 | **24.7** |
| YCbCr-SIFT | 21.2 | **23.8** |
| SIFT | 18.2 | **20.3** |
| PHOG | 19.5 | **20.6** |
| **CSF** | 30.1 | **31.6** |
| **CGSF** | 35.3 | **36.9** |
| **CGSF+PHOG** | 40.1 | **42.3** |

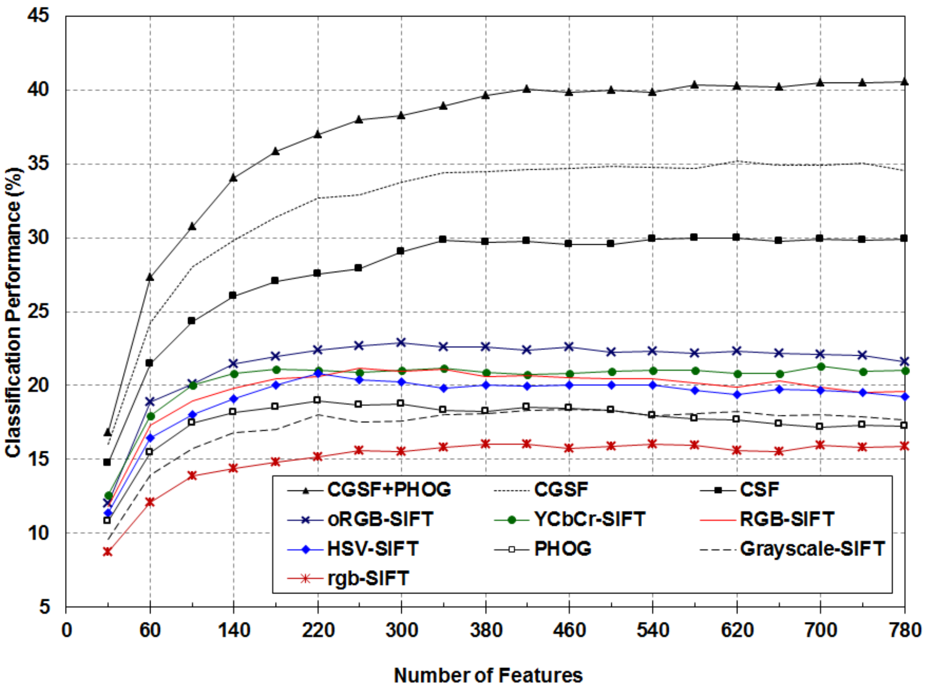


**Figure 4.24** Classification results using the PCA method across the ten descriptors with varying number of features on the Biometric 257 dataset.

*Evaluation of PCA and EFM-NN Results upon Varying Number of Features*

The third set of experiments evaluates the classification performance using the PCA and the EFM-NN methods respectively by varying the number of features over the following ten descriptors: CGSF+PHOG, CGSF, CSF, YCbCr-SIFT, oRGB-SIFT, RGB-SIFT, HSV-SIFT, Grayscale-SIFT, rgb-SIFT, and PHOG.

Classification performance is computed for up to 780 features with the PCA classifier. From Figure 4.24 it can be seen that the success rate for the CGSF+PHOG descriptor stays consistently above that of the CGSF and CSF descriptors on varying number of features and peaks at around 700 features. These three descriptors show an increasing trend during the first half and flatten out toward the second half. The oRGB-SIFT descriptor consistently stays above other color SIFT descriptors. The RGB-SIFT and HSV-SIFT descriptors have a similar success rate throughout. The grayscale-SIFT descriptor maintains its superior performance upon the rgb-SIFT and PHOG descriptors on the varying number

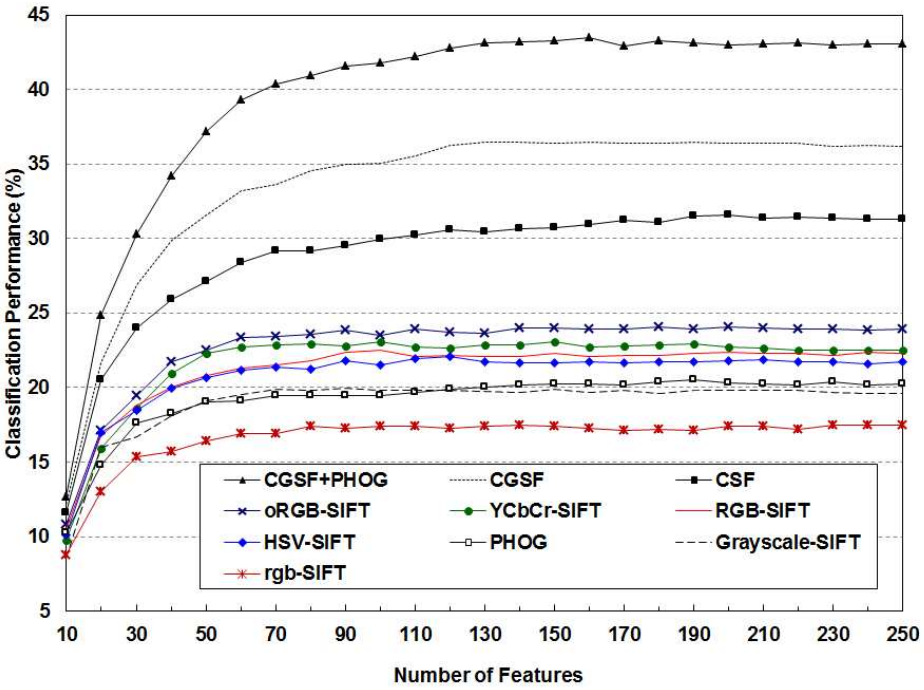


**Figure 4.25** Classification results using the EFM-NN method across the ten descriptors with varying number of features on the Biometric 257 dataset.

of features.

With the EFM-NN classifier, the success rates are computed for up to 250 features. From Figure 4.25 it can be seen that the success rate for the CGSF+PHOG stays consistently above that of the CGSF and CSF descriptors on varying number of features. These two descriptors show an increasing trend till 200 features and start to dip slightly thereafter. The YCbCr-SIFT and oRGB-SIFT descriptors show a similar increasing trend and decline only toward the latter half and continue to perform better than the rest of the descriptors. The grayscale-SIFT descriptor maintains its higher performance over the rgb-SIFT descriptor on the varying number of features.

*Evaluation of Descriptors and Classifier on Individual Image Categories*

The fourth set of experiments assesses the eight descriptors using the EFM-NN classifier on individual image categories. Here a detailed analysis of the performance of

**Table 4.6** Category Wise Descriptor Performance (%) Split-out with the EFM-NN Classifier on the Biometric 257 Dataset (Note That the Categories are Sorted on the CGSF+PHOG Results)

| Category | CGSF+ PHOG | CGSF | CSF | oRGB SIFT | YCbCr SIFT | RGB SIFT | Gray SIFT | PHOG |
|---|---|---|---|---|---|---|---|---|
| iris | **100** | **100** | **100** | **100** | **100** | **100** | 97 | 95 |
| faces | **97** | **97** | **97** | 92 | 60 | 92 | 73 | 95 |
| people | **17** | 14 | 13 | 10 | 8 | 10 | 7 | 11 |
| car side | **100** | **100** | **100** | 93 | 67 | **100** | **100** | 95 |
| leopards | **100** | 98 | **100** | 70 | 70 | 93 | 98 | 97 |
| motorbikes | **98** | 92 | 90 | 82 | 77 | 70 | 73 | 97 |
| sunflower | **97** | **97** | 93 | 88 | 68 | 95 | 90 | 53 |
| trilobite | **95** | 80 | 67 | 62 | 62 | 60 | 50 | 83 |
| lawn mower | **93** | 80 | 77 | 78 | 77 | 70 | 78 | 77 |
| american flag | **90** | 88 | 85 | 60 | 60 | 70 | 40 | 5 |
| zebra | **90** | 87 | 67 | 75 | 85 | 27 | 30 | 38 |
| chess board | 88 | **92** | 88 | 90 | 80 | 78 | 83 | 13 |
| tower pisa | 88 | **92** | 85 | 82 | 87 | 77 | 77 | 77 |
| swiss army knife | **87** | 82 | 77 | 60 | 60 | 75 | 68 | 8 |
| airplanes | **85** | 70 | 63 | 60 | 62 | 38 | 48 | 82 |
| saturn | **85** | 83 | 72 | 83 | 80 | 67 | 52 | 53 |
| cereal box | 83 | **88** | 72 | 73 | 52 | 52 | 47 | 17 |
| french horn | 83 | **85** | 78 | 72 | 67 | 68 | 70 | 32 |
| ketch | **83** | 62 | 57 | 28 | 30 | 32 | 37 | 65 |
| pci card | **83** | 80 | 78 | 68 | 50 | 58 | 58 | 10 |
| hibiscus | 82 | **85** | 80 | 65 | 72 | 72 | 62 | 48 |

the descriptors is performed with the EFM-NN classifier over all the 257 image categories. First the classification results on the three biometric categories are presented. Table 4.6 shows that the Iris category has a 100% recognition rate across all the descriptors with the exception of PHOG. For the Faces category the three fused descriptors reach a 97% success rate. The People category achieves a success rate of 17% with the CGSF+PHOG indicating the effect of very high intra-class variability due to the challenging background, variable postures, variable appearance, occlusion, multiple humans in the same image, and different illumination conditions. Fusion of the individual color SIFT descriptors (CSF) improves the classification performance, which indicates that various color descriptors are

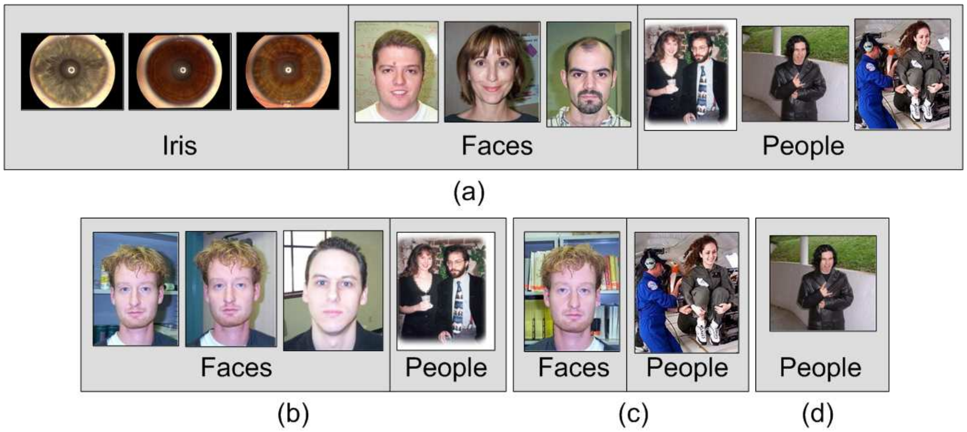


**Figure 4.26** Image recognition using the EFM-NN classifier on the Biometric 257 dataset: (a) examples of the correctly classified images from the three biometric image categories; (b) images unrecognized using the grayscale-SIFT descriptor but recognized using the oRGB-SIFT descriptor; (c) images unrecognized using the oRGB-SIFT descriptor but recognized using the CSF descriptor; (d) images unrecognized using the CSF but recognized using the CGSF+PHOG descriptor.

not redundant for recognition of the People category.

The average success rate for the CGSF+PHOG over the top 20 categories is 90.35% with ten categories at or above the 90% mark. Three categories have a 100% recognition rate. Individual color SIFT features improve over the grayscale-SIFT for most of the categories, in particular for the Trilobite, American flag, Tower Pisa, Saturn, and Hibiscus categories. The CSF almost always improves over the grayscale-SIFT. The CGSF either is at par or improves over the CSF on all categories with the exception of Leopards category. Most categories perform at their best when the PHOG is combined with the CGSF.

*Evaluation of Descriptors and Classifier Based on Correctly Recognized Images*

The final set of experiments further assesses the performance of the descriptors based on the correctly recognized images. See Figure 4.26(a) for some examples of the correctly classified images from the Iris, Faces, and People categories. Notice the high intra-class variability for the Faces and People classes. Figure 4.26(b) shows some example images from the Faces class that are not recognized by the EFM-NN classifier using the

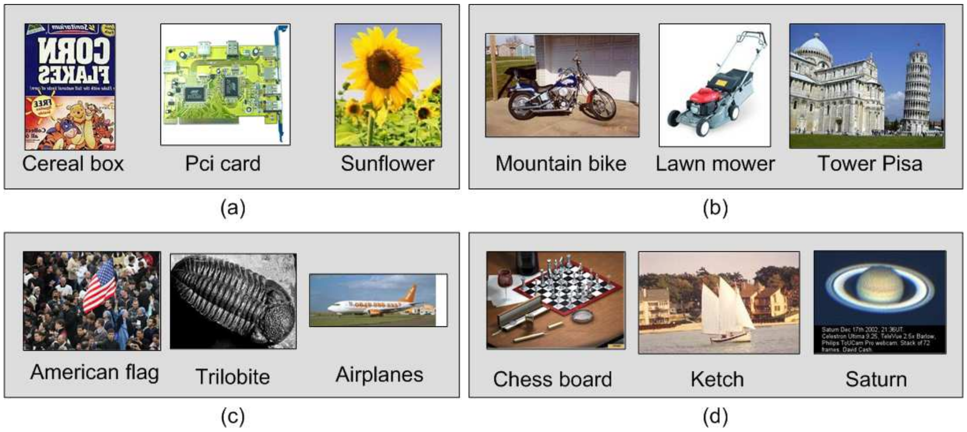


**Figure 4.27** Image recognition using the EFM-NN classifier on the Biometric 257 dataset: (a) example images unrecognized using the grayscale-SIFT descriptor but recognized using the oRGB-SIFT descriptor; (b) example images unrecognized using the oRGB-SIFT descriptor but recognized using the CSF descriptor; (c) images unrecognized using the CSF but recognized using the CGSF+PHOG. (d) Images unrecognized using the PCA but recognized using the EFM-NN on the CGSF+PHOG descriptor.

grayscale-SIFT descriptor but are correctly recognized using the oRGB-SIFT descriptor. This reaffirms the importance of color and the distinctiveness of the oRGB-SIFT descriptor for image category recognition. Figure 4.26(c) shows images unrecognized using the oRGB-SIFT descriptor but recognized using the CSF descriptor and Figure 4.26(d) shows images unrecognized using the CSF but recognized using the CGSF+PHOG.

See Figure 4.27(a) for some examples of the images unrecognized by the EFM using the grayscale-SIFT but are correctly recognized by the oRGB-SIFT. Figure 4.27(b) shows some images that were previously not recognized by the oRGB-SIFT but are correctly recognized by the CSF. In Figure 4.27(c) see some images unrecognized by the CSF but are correctly recognized by the CGSF+PHOG descriptor. Lastly in Figure 4.27(d) images unrecognized by the PCA but recognized by the EFM-NN classifier on the CGSF+PHOG descriptor.

**Table 4.7** Classification Performance (%) Comparison on the Caltech 256 and the Biometric 257 Categories Datasets on Groups of 15, 30, 45, and 60 Training Images From Each Class

| Dataset | Method | 15 | 30 | 45 | 60 |
|---|---|---|---|---|---|
| Caltech 256 | SPM (Griffin et al. 2007) | - | 34.1 | - | - |
| Caltech 256 | KC (Gemert et al. 2008) | - | 27.2 | - | - |
| Caltech 256 | KSPM (Yang et al. 2009) | 23.3 | 29.5 | - | - |
| Caltech 256 | LSPM (Yang et al. 2009) | 13.2 | 15.5 | 16.4 | 16.6 |
| Caltech 256 | ScSPM (Yang et al. 2009) | 27.7 | 34.0 | 37.5 | 40.1 |
|  | Proposed Methods: |  |  |  |  |
| Caltech 256 | CGSF+PHOG (PCA) | 27.1 | 33.8 | 37.3 | 39.9 |
| Caltech 256 | CGSF+PHOG (EFM-NN) | **30.0** | **36.5** | **39.7** | **42.1** |
| Biometric 257 | CGSF+PHOG (PCA) | 27.5 | **34.1** | **37.6** | 40.1 |
| Biometric 257 | CGSF+PHOG (EFM-NN) | **30.4** | **36.9** | **40.0** | **42.3** |

*Comparison with other Methods on the Caltech 256 and the Biometric 257 Datasets*

Finally the results of proposed methods are presented on varying number of training images per class over the Caltech 256 dataset and the Biometric 257 dataset. The number of test images per class is fixed at 20 for each of the experiments. From Table 4.7 on the Biometric 257 dataset it can be seen that on the 15 training images a success rate of 30.39% is attained. This improves over the previous best result on the Caltech 256 by 2.66%. For the 30 and 45 training images, there is an improvement on the previous best on the Caltech 256 by 2.85% and 2.58%, respectively. On the 60 training images, a success rate of 42.28% is achieved and this improves over the previous best by 2.14%.

From Table 4.7 on the Caltech 256 dataset it can be seen that on the 15 training images per class a success rate of 29.97% is achieved. This improves over the previous best result by 2.24%. For the set of 30 and 45 training images, an improvement of 2.39% and 2.28%, respectively is achieved on the previous best. On the 60 training images, a success rate of 42.06% is achieved, which improves over the previous best by 1.92%.

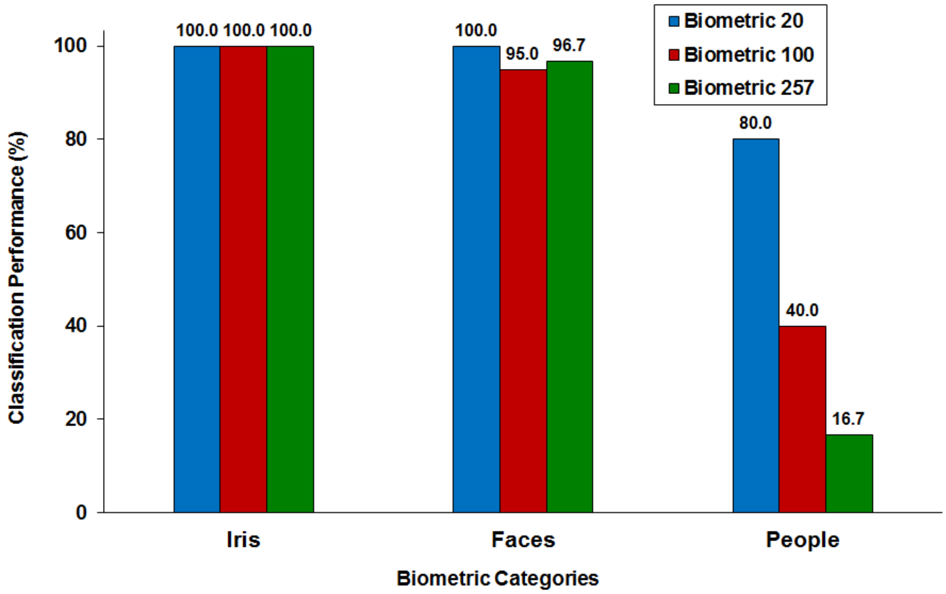


**Figure 4.28** Classification results on the three biometric categories. For each biometric category, from left to right bars in the triad denote the results from the Biometric 20 dataset, the Biometric 100 dataset, and the Biometric 257 dataset.

### 4.3.5 Panoramic Assessment of the Performance of Biometric Categories across Three Datasets and Five Descriptors

Finally, an overall view of the performance of the three biometric categories across three datasets is provided. From Figure 4.28 it can be seen that the Iris category has a success rate of 100% across all three datasets. The Faces category performs above the 95% mark on all three datasets and reaches 100% on the Biometric 20 dataset. The success rate for the People category varies widely from 16.7% on the Biometric 257 dataset to 80% on the Biometric 20 dataset.

Figure 4.29 shows the classification results on the CGSF+PHOG, CGSF, CSF, oRGB-SIFT, and grayscale-SIFT descriptors across all three datasets. It can be seen that the performance of each descriptor across the three datasets changes in a consistent manner.

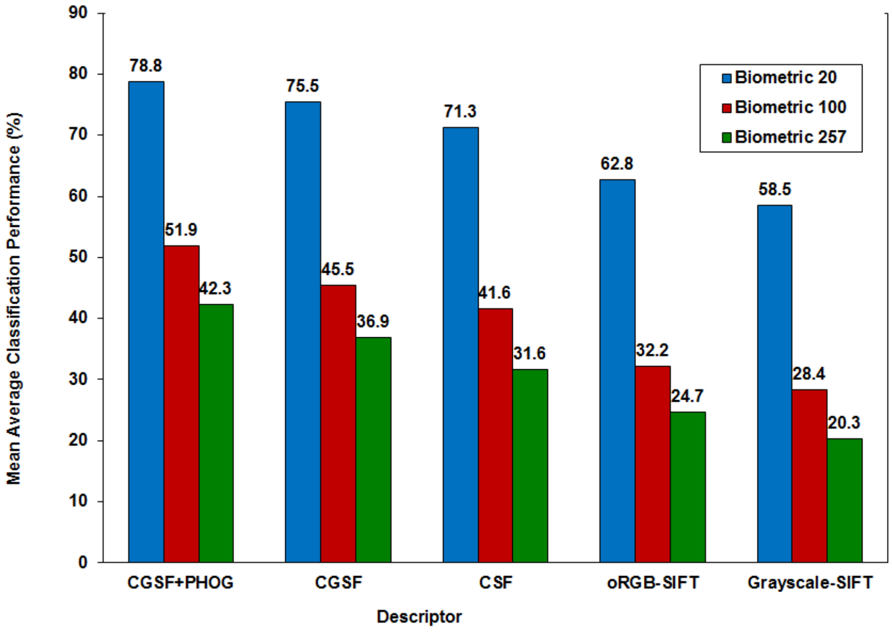


**Figure 4.29** Classification results on the five descriptors. For each descriptor, from left to right bars in the triad denote the results from the Biometric 20 dataset, the Biometric 100 dataset, and the Biometric 257 dataset.

## 4.4 Evaluation of SIFT Descriptors on the MIT Scene Dataset

### 4.4.1 Experimental Methodology

The classification task is to assign each test image to one of a number of categories. The performance is measured using a confusion matrix, and the overall performance rates are measured by the average value of the diagonal entries of the confusion matrix. For the MIT scene dataset five sets are selected randomly and each set consists of 2,000 images for training (250 images per class) and the rest 688 images for testing. Within each set there is no overlap in the images selected for training and testing. The classification scheme on the datasets compares the overall and category wise performance of ten different descriptors: the oRGB-SIFT, the YCbCr-SIFT, the RGB-SIFT, the HSV-SIFT, the rgb-SIFT, the grayscale-SIFT, the PHOG, the CSF, the CGSF, and the CGSF+PHOG descriptors. Classification is implemented using the EFM-KNN classifier.

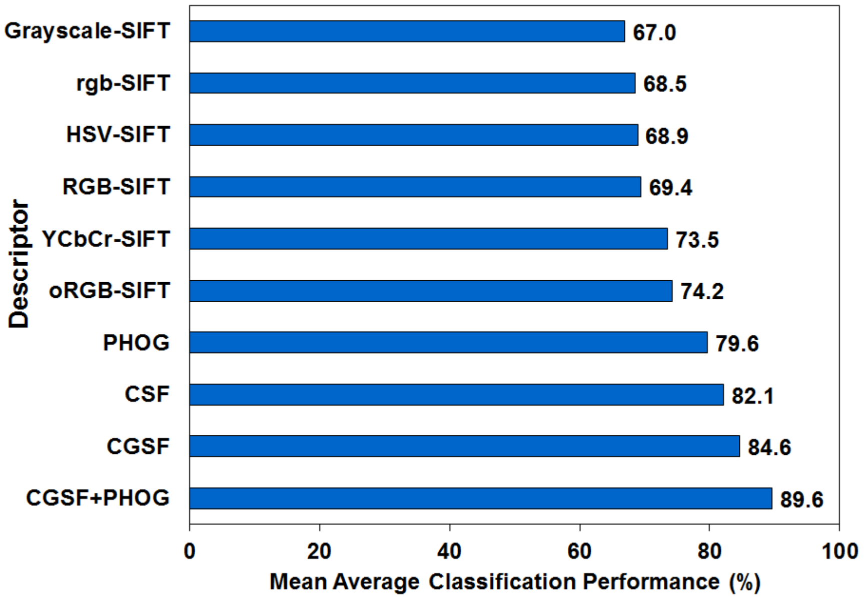


**Figure 4.30** The mean average classification performance of the ten descriptors using the EFM-KNN classifier on the MIT scene dataset: the oRGB-SIFT, the YCbCr-SIFT, the RGB-SIFT, the HSV-SIFT, the rgb-SIFT, the grayscale-SIFT, the PHOG, the CSF, the CGSF, and the CGSF+PHOG descriptors.

### 4.4.2 Experimental Results

The first set of experiments on this dataset assesses the overall classification performance of the ten descriptors. Note that for each category a five-fold cross validation is implemented for each descriptor using the EFM-KNN classifier to derive the average classification performance. Figure 4.30 shows the mean average classification performance of various descriptors. The best recognition rate that is obtained is 89.6% from the CGSF+PHOG, which is a very respectable value for a dataset of this size and complexity. The oRGB-SIFT achieves the classification rate of 74.2%. It outperforms the other color SIFT descriptors. It is noted that fusion of the color SIFT descriptors (CSF) improves upon the grayscale-SIFT by a huge 15.1% margin. The grayscale-SIFT descriptor improves the fusion (CGSF) result by a good 2.5% margin upon the CSF descriptor.

The second set of experiments assesses the five best descriptors and the grayscale-SIFT using the EFM-KNN classifier on individual image categories. From Table 4.8 it

**Table 4.8** Category Wise Descriptor Performance (%) Split-out with the EFM-KNN Classifier on the MIT Scene Dataset (Note That the Categories are Sorted on the CGSF+PHOG Results)

| Category | CGSF+ PHOG | CGSF | CSF | oRGB SIFT | Gray SIFT | PHOG |
|---|---|---|---|---|---|---|
| Highway | **100** | **100** | **100** | **100** | 82 | 91 |
| Forest | **99** | 97 | 97 | 88 | 90 | 95 |
| Inside City | **97** | 90 | 88 | 75 | 86 | 80 |
| Coast | **91** | 77 | 66 | 63 | 65 | 84 |
| Street | 90 | **93** | 89 | 90 | 60 | 86 |
| Mountain | **88** | 82 | 79 | 69 | 60 | 75 |
| Tall Building | **85** | 81 | 80 | 65 | 62 | 71 |
| Open Country | **67** | 58 | 57 | 43 | 32 | 56 |
| **Mean** | **89.6** | **84.6** | **82.1** | **74.2** | **67** | **79.6** |

can be seen that the top five categories achieve a success rate of over 90%. The Highway category achieves a success rate of 100% across the best four descriptors. Individual color SIFT features improve upon the grayscale-SIFT on most of the categories. The CSF results on each of the eight categories show improvement upon the grayscale-SIFT and the CGSF improves upon the CSF. Integration of PHOG with the CGSF to obtain the CGSF+PHOG highly benefits most categories and in particular there is a significant increase in the classification performance upon the CGSF results for the Coast, Inside City and Open Country categories where the increment is in the range of 7% to 14%.

The final set of experiments further assesses the performance of the descriptors

**Table 4.9** Comparison of the Classification Performance (%) with Other Method on the MIT Scene Dataset

| # Train | # Test | Proposed Method | | (Oliva and Torralba 2001) |
|---|---|---|---|---|
| 2000 | 688 | CSF | **82.1** | |
| | | CGSF | **84.6** | |
| | | CGSF+PHOG | **89.6** | - |
| 800 | 1888 | CSF | 79.4 | |
| | | CGSF | 81.9 | |
| | | CGSF+PHOG | **86.7** | 83.7 |

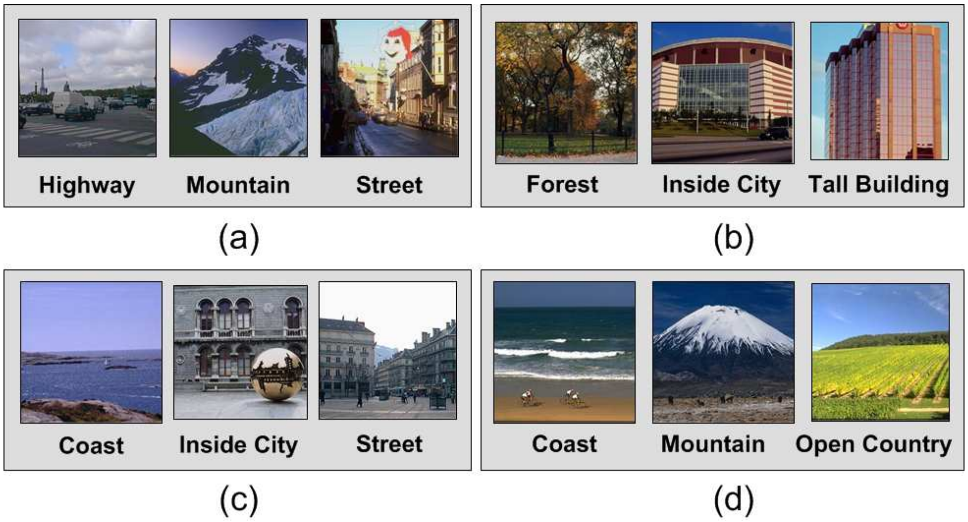


**Figure 4.31** Image recognition using the EFM-KNN classifier on the MIT scene dataset: (a) example images unrecognized using the grayscale-SIFT descriptor but recognized using the oRGB-SIFT descriptor; (b) example images unrecognized using the oRGB-SIFT descriptor but recognized using the CSF descriptor; (c) images unrecognized using the CSF but recognized using the CGSF descriptor; (d) images unrecognized using the CGSF but recognized using the CGSF+PHOG descriptor.

based on the correctly recognized images. See Figure 4.31(a) for some example images that are not recognized by the EFM-KNN classifier using the grayscale-SIFT descriptor but are correctly recognized using the oRGB-SIFT descriptor. Figure 4.31(b) shows images unrecognized using the oRGB-SIFT descriptor but recognized using the CSF descriptor, Figure 4.31(c) shows images unrecognized using the CSF but recognized using the CGSF descriptor and Figure 4.31(d) shows images unrecognized using the CGSF but recognized using the CGSF+PHOG descriptor.

From Table 4.9 it can be seen that on the 800 training images (100 images per class) and 1,688 testing images a success rate of 86.7% is achieved with the CGSF+PHOG descriptor. This improves over the result of authors in (Oliva and Torralba 2001) by a good 3%.

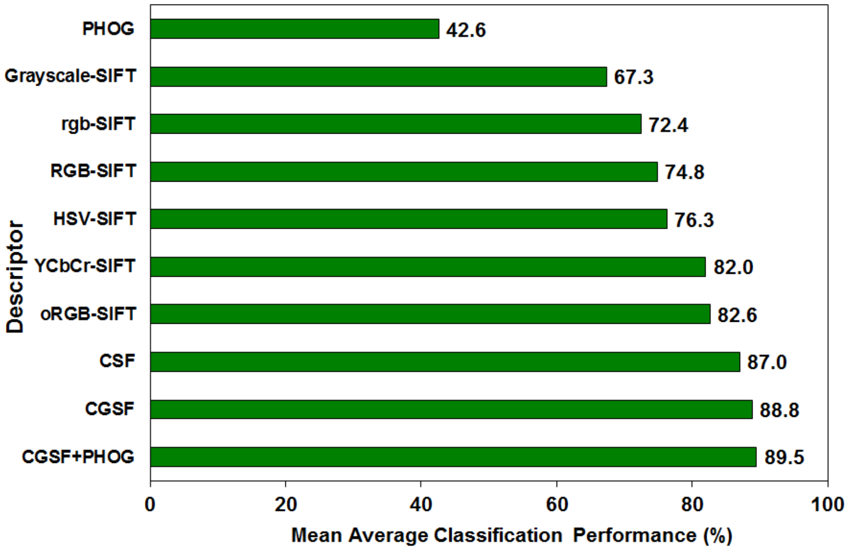


**Figure 4.32** The mean average classification performance of the ten descriptors fusing the EFM-KNN classifier on the Oxford flower dataset: the oRGB-SIFT, the YCbCr-SIFT, the RGB-SIFT, the HSV-SIFT, the rgb-SIFT, the grayscale-SIFT, the PHOG, the CSF, the CGSF, and the CGSF+PHOG descriptors.

## 4.5 Evaluation of SIFT Descriptors on the Oxford Flower Dataset

### 4.5.1 Experimental Methodology

The classification task is to assign each test image to one of a number of categories. The performance is measured using a confusion matrix, and the overall performance rates are measured by the average value of the diagonal entries of the confusion matrix. Three sets of 40 training images per class and 20 test images per class are selected (same data splits as used in (Nilsback and Zisserman 2006)). Within each set there is no overlap in the images selected for training and testing. The classification scheme on the datasets compares the overall and category wise performance of ten different descriptors: the oRGB-SIFT, the YCbCr-SIFT, the RGB-SIFT, the HSV-SIFT, the rgb-SIFT, the grayscale-SIFT, the PHOG, the CSF, the CGSF, and the CGSF+PHOG descriptors. Classification is implemented using the EFM-KNN classifier.

### 4.5.2 Experimental Results

The first set of experiments assesses the overall classification performance of the ten descriptors on the Oxford flower dataset. Note that for each category three-fold cross validation is implemented for each descriptor using the EFM-KNN classifier to derive the average classification performance. As a result, each descriptor yields 17 average classification rates corresponding to the 17 image categories. The mean value of these 17 average classification rates is defined as the mean average classification performance for the descriptor. Figure 4.32 shows the mean average classification performance of various descriptors.

The best recognition rate that is obtained is 89.5% from the CGSF+PHOG, which is a very respectable value for a dataset of this size and complexity. The oRGB-SIFT achieves the classification rate of 82.6%. It outperforms the other color SIFT descriptors. It is noted that fusion of the color SIFT descriptors (CSF) improves upon the grayscale-SIFT by a huge 19.7% margin. The grayscale-SIFT descriptor improves the fusion (CGSF) result by a good 1.8% margin upon the CSF descriptor.

The second set of experiments assesses the five best descriptors and the grayscale-SIFT using the EFM-KNN classifier on individual image categories. From Table 4.10 it can be seen that three categories achieve 100% success rate and over 50% of the categories achieve a success rate of more than 90% with the CGSF+PHOG descriptor. Sunflower achieves 100% success rate across the best five descriptors. Crocus and Tulip have a success rate of 75% indicating high intra-class variability and low inter-class variability. The average success rate for the top 10 categories with the CGSF+PHOG descriptor is a respectable 95%. Individual color SIFT features improve upon the grayscale-SIFT on most of the categories. The CSF almost always improves upon the grayscale-SIFT, this indicates that various color descriptors are not redundant. The CGSF improves upon the CSF; furthermore most categories perform at their best when the PHOG is combined with the CGSF.

The final set of experiments further assesses the performance of the descriptors

**Table 4.10** Category Wise Descriptor Performance (%) Split-out with the EFM-KNN Classifier on the Oxford Flower Dataset (Note That the Categories are Sorted on the CGSF+PHOG Results)

| Category | CGSF+ PHOG | CGSF | CSF | oRGB SIFT | YCbCr SIFT | Gray SIFT |
|---|---|---|---|---|---|---|
| Daisy | **100** | 98 | 98 | **100** | 98 | 93 |
| Sunflower | **100** | **100** | **100** | **100** | **100** | 95 |
| Windflower | **100** | 98 | 92 | 92 | 92 | 90 |
| Tigerlily | **98** | **98** | 97 | **98** | 95 | 78 |
| Dandelion | **95** | 93 | 92 | 92 | 92 | 82 |
| Bluebell | **93** | **93** | 90 | 85 | 79 | 49 |
| Colt's Foot | 93 | 93 | **95** | 90 | 93 | 83 |
| Lily Valley | **93** | 91 | 90 | 82 | 80 | 78 |
| Pansy | 91 | **93** | 89 | 76 | 78 | 75 |
| Cowslip | 87 | **90** | 88 | 84 | 88 | 46 |
| Iris | **87** | 85 | 80 | 75 | 72 | 78 |
| Buttercup | 84 | **85** | 84 | 83 | 82 | 49 |
| Fritillary | 84 | **85** | 83 | 80 | 83 | 75 |
| Snowdrop | **83** | 81 | 78 | 62 | 60 | 63 |
| Daffodil | 82 | 80 | **83** | 78 | 73 | 45 |
| Crocus | **75** | 69 | 68 | 63 | 59 | 25 |
| Tulip | **75** | 74 | 73 | 64 | 70 | 37 |
| **Mean** | **89.5** | **88.8** | **87** | **82.6** | **82** | **67.3** |

based on the correctly recognized images. See Figure 4.33(a) for some example images that are not recognized by the EFM-KNN classifier using the grayscale-SIFT descriptor but are correctly recognized using the oRGB-SIFT descriptor. This reaffirms the importance of color and the distinctiveness of the oRGB-SIFT descriptor for image category recognition. Figure 4.33(b) shows images unrecognized using the oRGB-SIFT descriptor but recognized using the CSF descriptor, Figure 4.33(c) shows images unrecognized using the CSF but recognized using the CGSF descriptor and Figure 4.33(d) shows images unrecognized using the CGSF but recognized using the CGSF+PHOG descriptor.

Table 4.11 shows a comparison of the results with those obtained by Nilsback and Zisserman (2006) and Varma and Ray (2007). Proposed technique outperforms the state of the art on this dataset even without combining color descriptors or considering texture

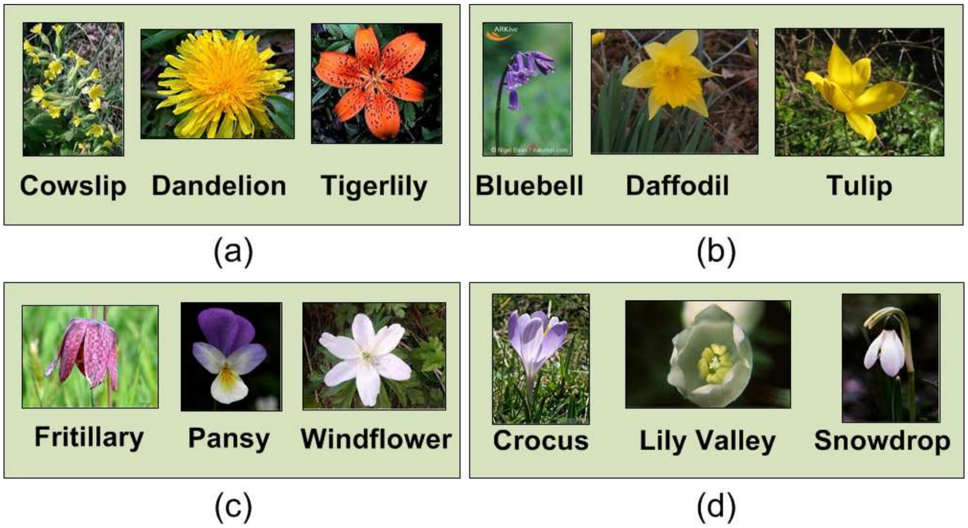


**Figure 4.33** Image recognition using the EFM-KNN classifier on the Oxford flower dataset: (a) example images unrecognized using the grayscale-SIFT descriptor but recognized using the oRGB-SIFT descriptor; (b) example images unrecognized using the oRGB-SIFT descriptor but recognized using the CSF descriptor; (c) images unrecognized using the CSF but recognized using the CGSF descriptor; (d) images unrecognized using the CGSF but recognized using the CGSF+PHOG descriptor.

and shape features independently. Each of the four color SIFT descriptors outperform descriptors in (Nilsback and Zisserman 2006), (Varma and Ray 2007). Combined SIFT descriptors (CSF, CGSF and CGSF+PHOG) improve over the fusion result in (Nilsback and Zisserman 2006) and SVM 1-vs-All fusion result in (Varma and Ray 2007), previously the best result on this dataset.

**Table 4.11** Comparison of the Classification Performance (%) with Other Methods on the Oxford Flower Dataset

| Proposed Method | | Nilsback 2006 | | Varma 2007 | |
|---|---|---|---|---|---|
| RGB-SIFT | 74.8 | Color | 73.7 | Shape | 68.88 |
| HSV-SIFT | 76.3 | Shape | 71.8 | Color | 59.71 |
| YCbCr-SIFT | 82.0 | Texture | 56.0* | Texture | 59.00 |
| oRGB-SIFT | **82.6** | | | | |
| CSF | **87.0** | | | | |
| CGSF | **88.8** | | | | |
| CGSF+PHOG | **89.5** | Fusion | 81.3 | Fusion | 82.55 |

*Approximate value inferred from Figure 12 in Nilsback 2006.

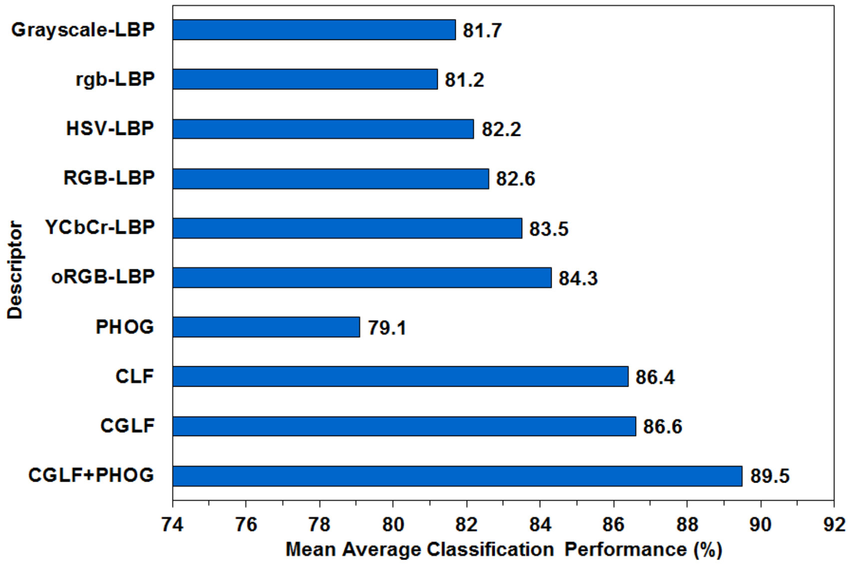


**Figure 4.34** The mean average classification performance of the ten descriptors using the EFM-NN classifier on the MIT scene dataset: the oRGB-LBP, the YCbCr-LBP, the RGB-LBP, the HSV-LBP, the rgb-LBP, the grayscale-LBP, the PHOG, the CLF, the CGLF, and the CGLF+PHOG descriptors.

## 4.6 Evaluation of LBP Descriptors on the MIT Scene Dataset

### 4.6.1 Experimental Methodology

The classification task is to assign each test image to one of a number of categories. The performance is measured using a confusion matrix, and the overall performance rates are measured by the average value of the diagonal entries of the confusion matrix. Five sets are randomly selected and each set consists of 2,000 images for training (250 images per class) and the rest 688 images for testing. Within each set there is no overlap in the images selected for training and testing. The classification scheme on this dataset compares the overall and category wise performance of ten different descriptors: the oRGB-LBP, the YCbCr-LBP, the RGB-LBP, the HSV-LBP, the rgb-LBP, the grayscale-LBP, the CLF, the CGLF, the PHOG and the CGLF+PHOG descriptors. Classification is implemented using the EFM-nearest neighbor (EFM-NN) classifier.

**Table 4.12** Category Wise Descriptor Performance (%) Split-out with the EFM-NN Classifier on the MIT Scene Dataset (Note That the Categories are Sorted on the CGLF+PHOG Results)

| Category | CGLF+ PHOG | CGLF | CLF | oRGB LBP | YCbCr LBP | RGB LBP | Gray LBP | PHOG |
|---|---|---|---|---|---|---|---|---|
| Highway | **97** | 90 | 93 | 90 | 87 | 90 | 93 | 90 |
| Forest | 96 | **97** | **97** | **97** | **97** | 95 | 94 | 94 |
| Coast | **91** | 88 | 87 | 85 | 88 | 83 | 86 | 84 |
| Street | **90** | **90** | 86 | 83 | 83 | 82 | 81 | 86 |
| Mountain | **90** | 85 | 84 | 80 | 81 | 80 | 77 | 75 |
| Tall Building | **90** | 86 | 86 | 86 | 83 | 84 | 79 | 70 |
| Inside City | 86 | **87** | **87** | 86 | 83 | 81 | 83 | 79 |
| Open Country | **76** | 71 | 71 | 68 | 66 | 65 | 61 | 56 |
| **Mean** | **89.5** | **86.6** | **86.4** | **84.2** | **83.5** | **82.6** | **81.7** | **79.1** |

### 4.6.2 Experimental Results

The first set of experiments on this dataset assesses the overall classification performance of the ten descriptors. Note that for each category a five-fold cross validation is implemented for each descriptor using the EFM-nearest neighbor classifier to derive the average classification performance. As a result, each descriptor yields eight average classification rates corresponding to the eight image categories. The mean value of these 8 average classification rates is defined as the mean average classification performance for the descriptor. Figure 4.34 shows the mean average classification performance of various descriptors. The best recognition rate that is obtained is 89.5% from the CGLF+PHOG, which is a very respectable value for a dataset of this size and complexity. The oRGB-LBP achieves the classification rate of 84.3%. It outperforms the other color LBP descriptors. It is noted that fusion of the color LBP descriptors (CLF) improves upon the grayscale-LBP by a significant 4.0% margin. The grayscale-LBP descriptor improves the fusion (CGLF) result slightly upon the CLF descriptor.

The second set of experiments assesses the ten descriptors using the EFM-nearest neighbor classifier on individual image categories. From Table 4.12 it can be seen that

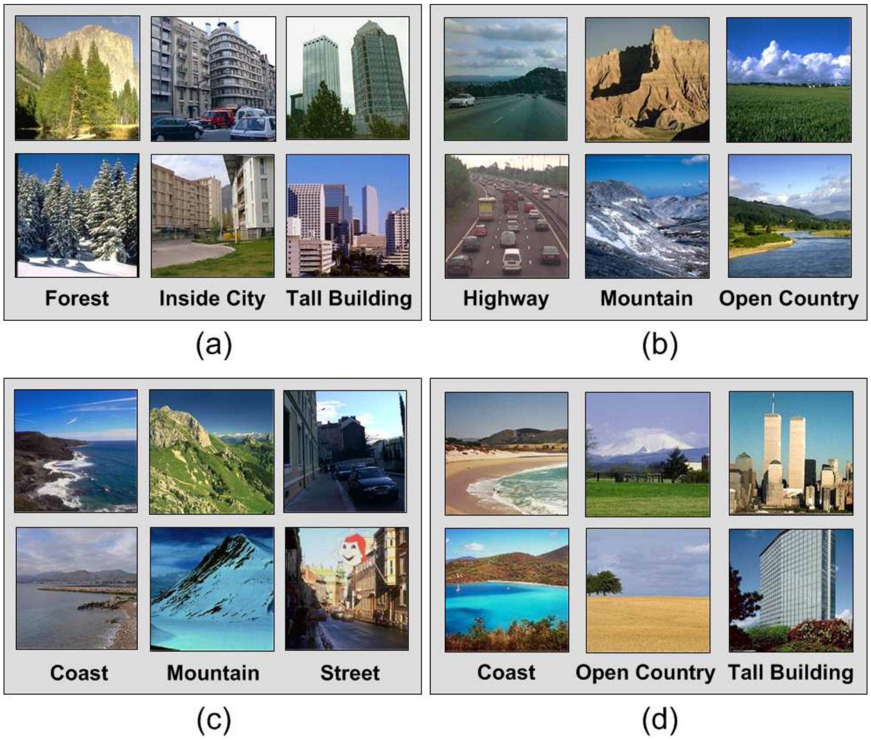


**Figure 4.35** Image recognition using the EFM-NN classifier on the MIT scene dataset: (a) example images unrecognized using the grayscale-LBP descriptor but recognized using the oRGB-LBP descriptor; (b) example images unrecognized using the oRGB-LBP descriptor but recognized using the CLF descriptor; (c) images unrecognized using the CLF but recognized using the CGLF descriptor; (d) images unrecognized using the CGLF but recognized using the CGLF+PHOG descriptor.

the top six categories achieve a success rate of over 90%. The Forest category achieves a success rate of over 90% across all ten descriptors. Individual color LBP features improve upon the grayscale-LBP on most of the categories. The CLF results on each of the eight categories show significant improvement upon the grayscale-LBP and the CGLF slightly improves upon the CLF. Integration of PHOG with the CGLF to obtain the CGLF+PHOG highly benefits most categories and in particular there is a significant increase in the classification performance upon the CGLF results for the Highway, Mountain and Open Country categories where the increment is in the range of 5% to 7%.

The final set of experiments further assesses the performance of the descriptors based on the correctly recognized images. See Figure 4.35(a) for some example images

**Table 4.13** Comparison of the Classification Performance (%) with Other Method on the MIT Scene Dataset

| # Train | # Test | Proposed Method | | (Oliva and Torralba 2001) |
|---|---|---|---|---|
| 2000 | 688 | CLF | **86.4** | |
| | | CGLF | **86.6** | |
| | | CGLF+PHOG | **89.5** | - |
| 800 | 1888 | CLF | 79.3 | |
| | | CGLF | 80.0 | |
| | | CGLF+PHOG | **84.3** | 83.7 |

that are not recognized by the EFM-nearest neighbor classifier using the grayscale-LBP descriptor but are correctly recognized using the oRGB-LBP descriptor. Figure 4.35(b) shows images unrecognized using the oRGB-LBP descriptor but recognized using the CLF descriptor, Figure 4.35(c) shows images unrecognized using the CLF but recognized using the CGLF descriptor and Figure 4.35(d) shows images unrecognized using the CGLF but recognized using the CGLF+PHOG descriptor.

From Table 4.13 it can be seen that on the 800 training images (100 images per class) and 1,688 testing images a success rate of 84.3% is achieved with the CGLF+PHOG descriptor. This improves over the result of authors in (Oliva and Torralba 2001) by 0.6%.

## 4.7 Evaluation of LBP Descriptors on the KTH-TIPS and the KTH-TIPS-2b Datasets

### 4.7.1 Experimental Methodology

The classification task is to assign each test image to one of a number of categories. The performance is measured using a confusion matrix, and the overall performance rates are measured by the average value of the diagonal entries of the confusion matrix. For KTH-TIPS2-b dataset, five random sets of 200 training images per class and 100 testing images per class are used. For the KTH-TIPS dataset, five random sets of 40 training images per class and 41 test images per class are selected (same numbers as used in (Crosier and

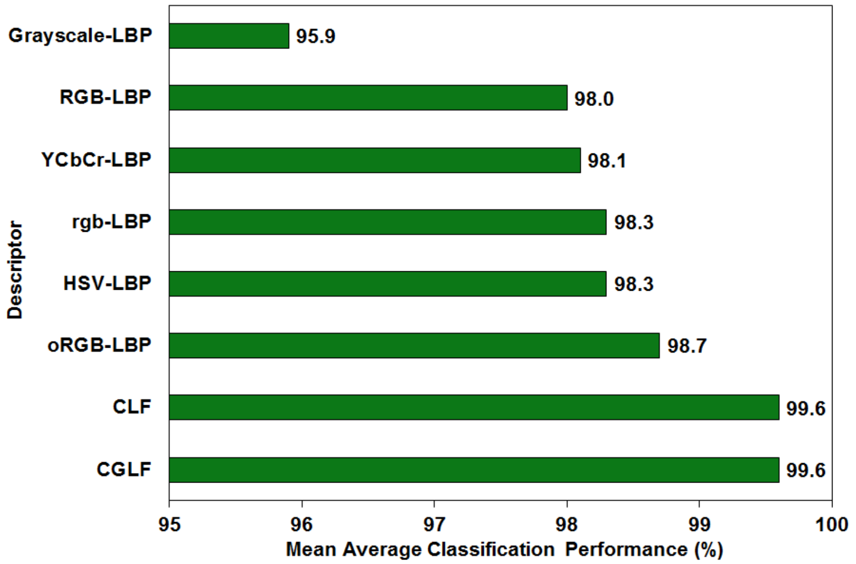


**Figure 4.36** The mean average classification performance of the eight descriptors using the EFM-NN classifier on the KTH-TIPS2-b dataset: the oRGB-LBP, the YCbCr-LBP, the RGB-LBP, the HSV-LBP, the rgb-LBP, the grayscale-LBP, the CLF, and the CGLF descriptors.

Griffin 2008), (Zhang et al. 2007), (Kondra and Torre 2008)). Within each set there is no overlap in the images selected for training and testing. The classification scheme on the datasets compares the overall and category wise performance of ten different descriptors: the oRGB-LBP, the YCbCr-LBP, the RGB-LBP, the HSV-LBP, the rgb-LBP, the grayscale-LBP, the CLF, the CGLF, the PHOG and the CGLF+PHOG descriptors (the final two evaluated on the scene dataset). Classification is implemented using the EFM-nearest neighbor (EFM-NN) classifier.

### 4.7.2 Experimental Results

Here, a detailed experimental evaluation is presented on the KTH-TIPS2-b dataset followed by a comparison of success rate with other research groups on the KTH-TIPS dataset. The first set of experiments assesses the overall classification performance of the eight descriptors on the KTH-TIPS2-b dataset. Note that for each category a five-fold cross validation

**Table 4.14** Category Wise Descriptor Performance (%) Split-out with the EFM-NN Classifier on the KTH-TIPS2-b Dataset (Note That the Categories are Sorted on the CGLF Results)

| Category | CGLF | CLF | oRGB LBP | HSV LBP | rgb LBP | Gray LBP |
|---|---|---|---|---|---|---|
| Aluminium Foil | **100** | **100** | **100** | **100** | **100** | **100** |
| Brown Bread | **100** | **100** | **100** | 99 | 99 | 94 |
| Corduroy | **100** | **100** | **100** | **100** | **100** | 93 |
| Cork | **100** | **100** | **100** | 98 | 98 | 98 |
| Cracker | **100** | **100** | 96 | 93 | 93 | 90 |
| Lettuce Leaf | **100** | **100** | **100** | **100** | **100** | 97 |
| Linen | **100** | **100** | **100** | 99 | 99 | 99 |
| Wood | **100** | **100** | **100** | **100** | **100** | **100** |
| Wool | **100** | **100** | 99 | **100** | **100** | 96 |
| White Bread | **99** | **99** | **99** | **99** | **99** | 97 |
| Cotton | **98** | 97 | 97 | 96 | 96 | 91 |
| **Mean** | **99.6** | **99.6** | **98.7** | **98.3** | **98.3** | **95.9** |

is implemented for each descriptor using the EFM-nearest neighbor classifier to derive the average classification performance. Figure 4.36 shows the mean average classification performance of various descriptors. The best recognition rate that is obtained is 99.6% from the CLF and CGLF descriptors. The oRGB-LBP achieves the classification rate of 98.7%. It outperforms the other color LBP descriptors. It is noted that fusion of the color LBP descriptors (CLF) improves upon the grayscale-LBP by a significant 3.7% margin. The grayscale-LBP descriptor does not have any effect on the fusion (CGLF) result in case of this dataset.

The second set of experiments assesses the five best descriptors and the grayscale-LBP using the EFM-nearest neighbor classifier on individual image categories. From Table 4.14 it can be seen that nine out of eleven categories achieve 100% success rate and all of the categories achieve a success rate of 98% or more with the CGLF descriptor. Aluminium Foil, Corduroy, Lettuce Leaf and Wood achieve 100% success rate across the best five descriptors. Individual color LBP features improve upon the grayscale-LBP on most of the categories. The CLF almost always improves upon the grayscale-LBP, this indicates

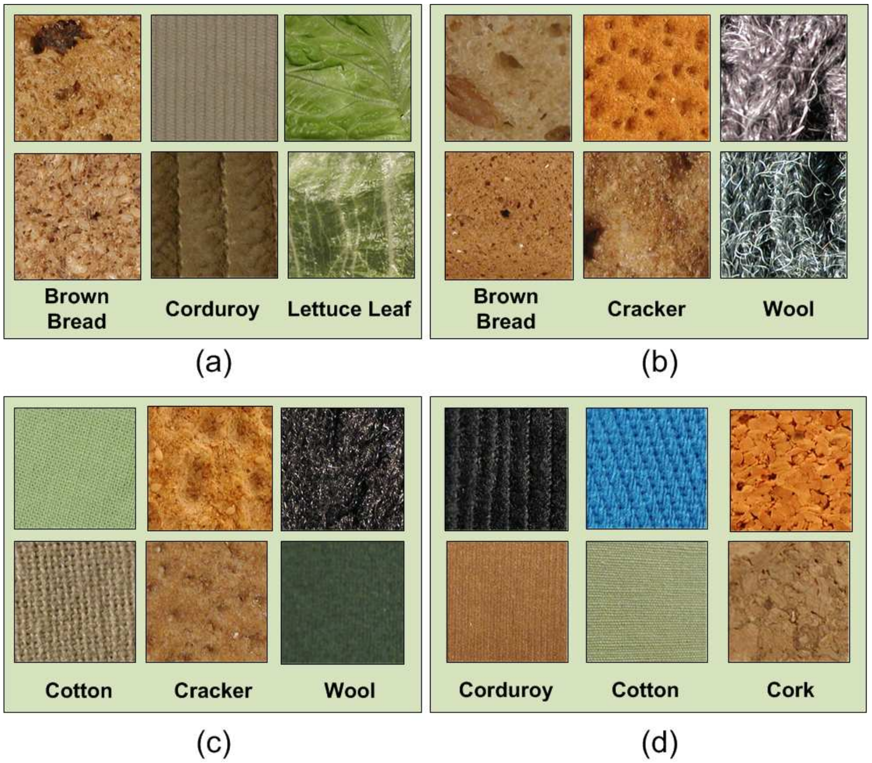


**Figure 4.37** Image recognition using the EFM-NN classifier on the KTH-TIPS2-b dataset: (a) example images unrecognized using the grayscale-LBP descriptor but recognized using the oRGB-LBP descriptor; (b) example images unrecognized using the RGB-LBP descriptor but recognized using the oRGB-LBP descriptor; (c) images unrecognized using the oRGB-LBP but recognized using the CLF descriptor; (d) images unrecognized using the grayscale-LBP but recognized using the CGLF descriptor.

that various color descriptors are not redundant. The CGLF very slightly improves upon the CLF. This, however, does not necessarily indicate that the grayscale information is redundant as almost all the categories show a success rate of 100% with these two descriptors. It only indicates that CLF alone contains enough information to correctly classify the texture images in the case of KTH-TIPS2-b dataset.

The final set of experiments further assesses the performance of the descriptors based on the correctly recognized images. See Figure 4.37(a) for some example images that are not recognized by the EFM-nearest neighbor classifier using the grayscale-LBP descriptor but are correctly recognized using the oRGB-LBP descriptor. This reaffirms the importance of color and the distinctiveness of the oRGB-LBP descriptor for image category

**Table 4.15** Comparison of the Classification Performance (%) with Other Methods on the KTH-TIPS Dataset

| Method | Performance |
|---|---|
| **Proposed Methods:** | |
| **CGLF** | **99.6** |
| **CLF** | **99.6** |
| **oRGB-LBP** | **99.1** |
| Crosier 2008 | 98.5 |
| Kondra 2008 | 97.7 |
| Zhang 2007 | 95.5 |

recognition. Figure 4.37(b) shows images unrecognized using the RGB-LBP descriptor but recognized using the oRGB-LBP descriptor, Figure 4.37(c) shows images unrecognized using the oRGB-LBP but recognized using the CLF descriptor, and Figure 4.37(d) shows images unrecognized using the grayscale-LBP but recognized when combined with the CLF, i.e., the CGLF descriptor.

The same set of experiments is repeated on the KTH-TIPS dataset. The best result on this dataset while using a single color space was once again from the oRGB-LBP descriptor, which achieves a 99.1% classification rate with an improvement of 3% over the grayscale-LBP. The CLF and the CGLF descriptors are tied at 99.6%. Table 4.15 shows a comparison of the results with those obtained from other methods in (Crosier and Griffin 2008), (Zhang et al. 2007), (Kondra and Torre 2008). In the oRGB color space, proposed technique outperforms the state of the art on this dataset even without combining color descriptors. Combined LBP descriptors (CLF and CGLF) improve upon the result in (Crosier and Griffin 2008), previously the best result on this dataset.

# CHAPTER 5

# IRIS IMAGE SEARCH BASED ON ROBUST SEGMENTATION AND IMAGE ENHANCEMENT

## 5.1 Robust Iris Search Method and Its Major Components

A robust iris search method is proposed and implemented here and improvement in iris recognition performance is shown using the Iris Challenge Evaluation (ICE) (Phillips 2006) dataset. First details of the ICE dataset are given in Section 5.1.1. Next the major components of the iris recognition method are presented. These include iris segmentation, iris encoding, and iris matching. Figure 5.1 shows the iris region between the sclera and the pupil. See Figure 5.2 for an overview of the iris recognition method.

Research efforts here mainly focus on improving the iris segmentation stage of the system. This allows to compare the performance of the segmentation stage with that implemented by the irisBEE method (Phillips et al. 2008). The segmentation step performs the localization of the iris region by detecting the pupil and the limbic boundary along with the eyelid detection. The iris encoding and iris matching stage are similar to that implemented by the irisBEE method (Phillips et al. 2008) in the Biometric Experimentation Environment (BEE) system. In comparison to the irisBEE method (Phillips et al. 2008) the proposed method leads to a significant increase in the accuracy of the iris region segmentation along with a much higher overall recognition performance at a lower error rate. Furthermore, proposed method outperforms the rank-one recognition performance achieved by the ND_IRIS (Liu et al. 2005) method.

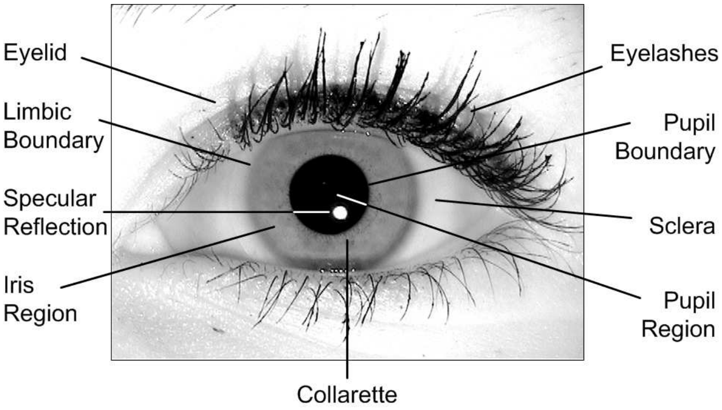


**Figure 5.1** Front view of the human eye. The various parts labeled are important to iris segmentation and recognition.

### 5.1.1 The Iris Challenge Evaluation (ICE) Dataset

The ICE dataset (Phillips 2006) consists of 1,425 right eye images of 124 different subjects and 1,528 left eye images of 120 different subjects. Eye images belong to 132 total subjects with 112 overlapping subjects between the left eye and the right eye images. The iris images are intensity images with a resolution of 640x480 in the TIFF format. The average diameter of an iris is 228 pixels. The images vary in quality due to the percentage of the iris region occlusion, the degree of blur in the image, off angle image, and images with subject wearing the contact lens. Figure 5.3(a) shows some example images of the right eye and Figure 5.3(b) shows some images from the left eye from the ICE dataset. Notice the varying degree of illumination levels, pupil dilation, angle and occlusion.

### 5.1.2 Iris Segmentation

Here the details of the iris segmentation method are presented. In particular, the effect of the power-law transformations on an eye image along with its advantages is discussed. Next details of efficient determination of the pupil region are presented, followed by a discussion on the effective method to determine the limbic boundary and the iris region segmentation. Finally, details of the improved eyelid detection phase are given. See Figure 5.4 for an

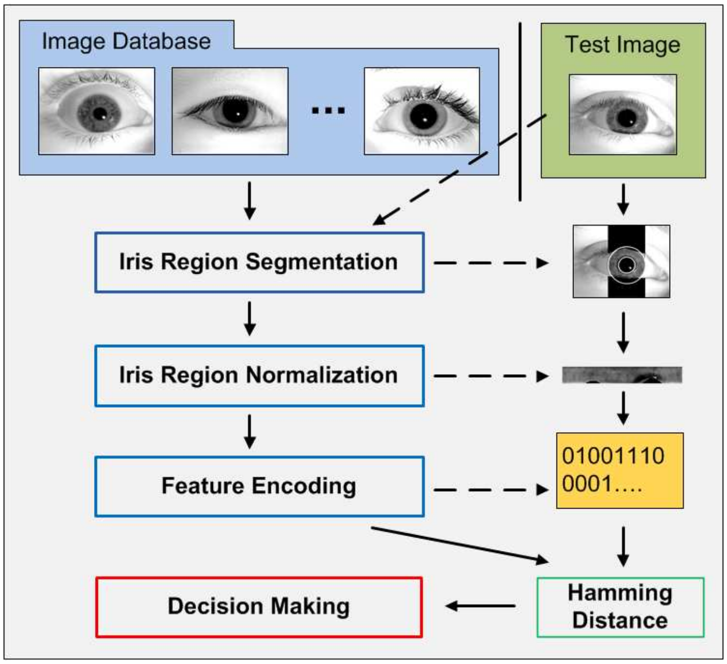


**Figure 5.2** An overview of the iris recognition method.

overview of the three main stages in iris segmentation: the pupil detection, the limbic boundary detection, and the eyelid detection.

*Performing the Power-Law Transformations on an Eye Image*

The power-law transformation when applied to a grayscale image changes its dynamic range. The pixel intensity values in the input image act as the basis, which is raised to a (fixed) power. The operator is defined by the following formula (Gonzalez and Woods 2001):

$$R(\rho) = c{*}I^{\frac{1}{\rho}} \tag{5.1}$$

where $I$ is the intensity value of a pixel in the input image, $c$ is the scaling factor, and $1/\rho$ is the power.

For $\rho < 1$, this operation increases the bandwidth of the high intensity values at the cost of the low pixel values. For $\rho > 1$, this process enhances the low intensity values while

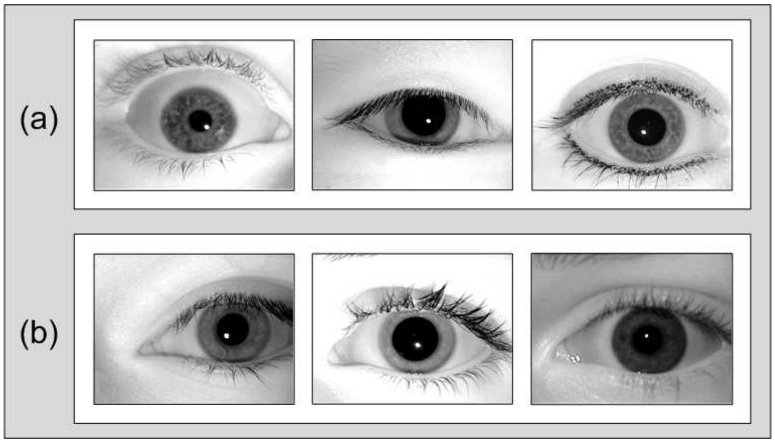


**Figure 5.3** Example images of the (a) right eye and (b) left eye from the ICE dataset, under varying illumination levels, pupil dilation, angle and occlusion.

decreasing the bandwidth of the high intensity values, i.e., boosts the contrast in the dark regions. For $\rho = 1$, the above transformation linearly scales the intensity values.

In Figure 5.5 the plot shows the result of the power-law transformations on the image intensity values at various values of $\rho$. The output pixel intensity value is scaled back to $[0, 255]$. This operation when applied on the input pixel intensity with $\rho = 1$ and $c = 1$ does not have any effect on the output intensity. This can be seen in the plot for $R(\rho)$ at $\rho = 1$. At $\rho = 1.9$ and 2.5 the lower intensity values gain more than the higher intensity values. At $\rho = 0.5$ the intensity values get pulled down and the lower values tend to get mapped into a narrower range.

Assessment of the impact of the power-law transformations on an eye image in terms of the pixel intensity frequency can be seen from Figure 5.6. The original eye image is shown in Figure 5.6(a), transformed images with $\rho$ values as 0.5, 1.9 and 2.5 can be seen in Figure 5.6(b), (c) and (d), respectively. The corresponding pixel intensity frequency plot for the four images is presented in Figure 5.6(e). For $\rho > 1$ many more pixels get mapped into a narrower brighter intensity range as seen in Figure 5.6(e). Also, this effect can be observed from the eye images in Figure 5.6(c) and (d) where the contrast between the pupil and the iris becomes more significant.

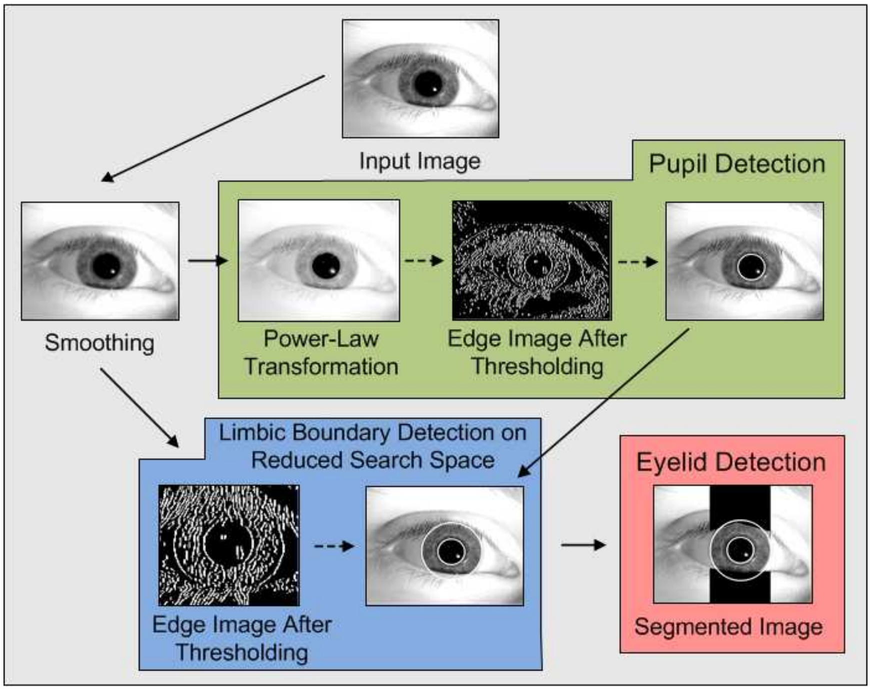


**Figure 5.4** An overview of the three main stages in iris segmentation: the pupil detection, the limbic boundary detection, and the eyelid detection.

*Efficient Determination of the Pupil Region*

Proposed new iris segmentation method first applies the power-law transformation on an eye image and then detects the pupil boundary. Firstly, detection of the pupil boundary is performed followed by the detection of outer iris boundary. The reason for this approach lies in the observation that the contrast between the iris and the pupil is usually larger than that between the iris and the sclera. The contrast is further enhanced from the application of the power-law transformation; this makes it easier to detect the pupil region and thereby increases the accuracy of the pupil segmentation. Proposed method results in the accurate detection of the pupil boundary for 99.8% of the images in the dataset; this includes the entire right eye and the left eye images. The appropriate $\rho$ value for the power-law transformation is selected after analyzing the contrast between the iris and the pupil on a validation set chosen from the ICE dataset.

In Figure 5.4 and Figure 5.7(a) details of the pupil detection stage are presented. As

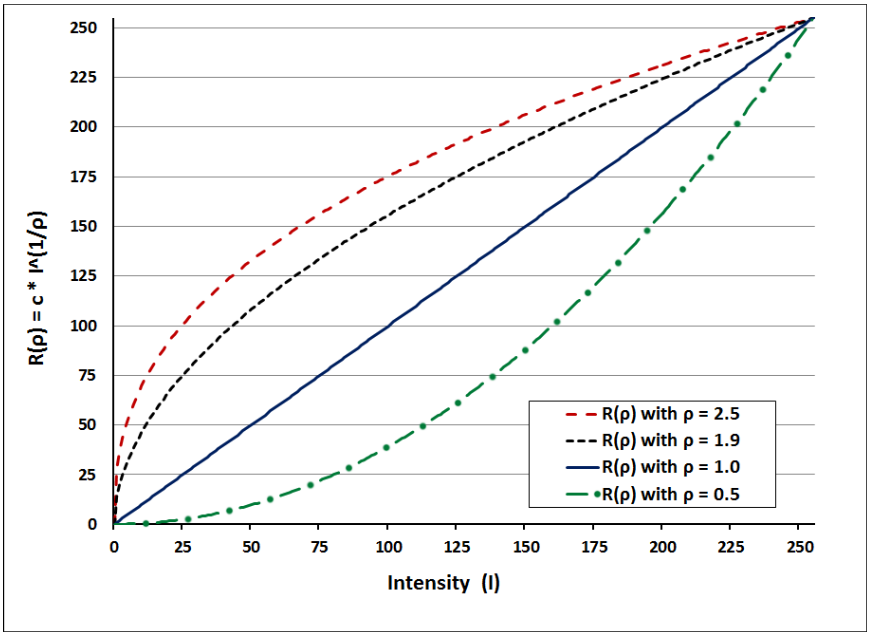


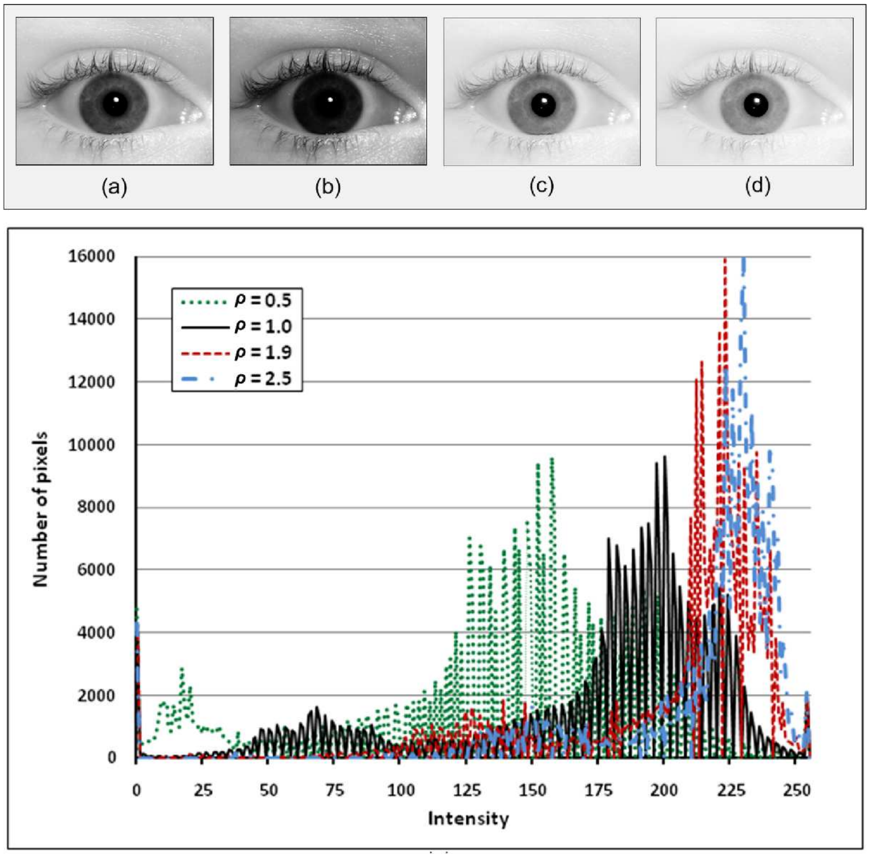
**Figure 5.5** Plot showing the result of the power-law transformations on the image intensity values at various values of $\rho$.

a first step Gaussian filter is applied to the input image for smoothing, in order to get rid of the noise. The Gaussian smoothing filter is defined by the following formula (Forsyth and Ponce 2003):

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{5.2}$$

where $x$ is the distance from the origin in the horizontal axis, $y$ is the distance from the origin in the vertical axis, and $\sigma$ is the standard deviation of the Gaussian distribution. In the next stage, power-law transformation is applied followed by the canny edge detector to detect edges in the image. Thresholding is performed to get rid of the weak edges.

Finally, circular Hough transform is applied on the edge image to detect the pupil boundary. In order to make the pupil search more accurate and fast, a candidate pupil having radius within a narrow range is searched. This range is computed from a validation set chosen from the ICE dataset. See in Figure 5.7(a) the image on the left for the range of the radius and on the right the edge image space to be searched for candidate pupil circles. The circular Hough transform can be described as a transformation of a point in the



(e)

**Figure 5.6** Results of the power-law transformations on (a) input eye image, at $\rho = 0.5$, 1.9 and 2.5 shown in (b), (c) and (d), respectively. (e) Plot of the frequency of intensity of the input image at various $\rho$ values. Plot at $\rho = 1.0$ corresponds to the input image in (a).

$x, y$-plane to the circle parameter space. The parametric representation of the circle is given as:

$$\begin{aligned} x &= a + r\cos(\theta) \\ y &= b + r\sin(\theta) \end{aligned} \tag{5.3}$$

where $a$ and $b$ are the center of the circle in the $x$ and $y$ direction respectively and where $r$ is the radius and $\theta$ is the angle between 0 and $2\pi$.

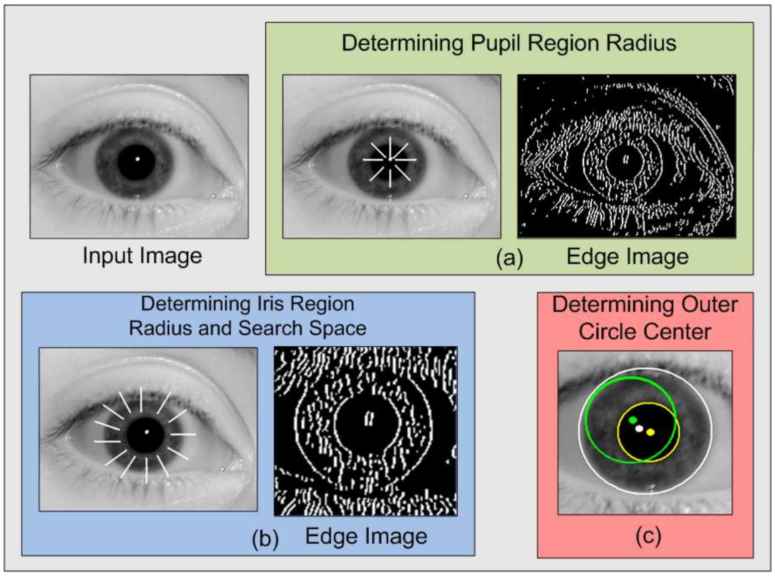


**Figure 5.7** Efficient determination of: (a) the pupil region radius, (b) the iris region radius and search space, and (c) the limbic boundary center.

*Efficient Determination of the Limbic Boundary and the Iris Region*

It is observed that when detecting the limbic boundary the Hough transform often makes incorrect detections. Research reveals that such incorrect detections are due to the presence of a large number of weak edges. Therefore, a thresholding technique is applied to the edge image produced by the Canny edge detector to get rid of the insignificant edge points. This has shown to improve the percentage of the correctly segmented iris region by close to 3% for both the right eye and the left eye images. See Figure 5.4 and Figure 5.7(b) for details.

In order to further improve the accuracy of the Hough transform for detecting the limbic boundary, the circle within a specific region around the detected pupil boundary is searched. Furthermore, a candidate limbic boundary having radius within a narrow range is searched. The range for the radius is estimated on the validation set chosen from the ICE dataset. The reduced search space and the narrow radius range thus considerably increase the speed of the circle detection. See in Figure 5.7(b) the image on the left for the range of the radius and on the right the reduced edge image space that is searched for candidate

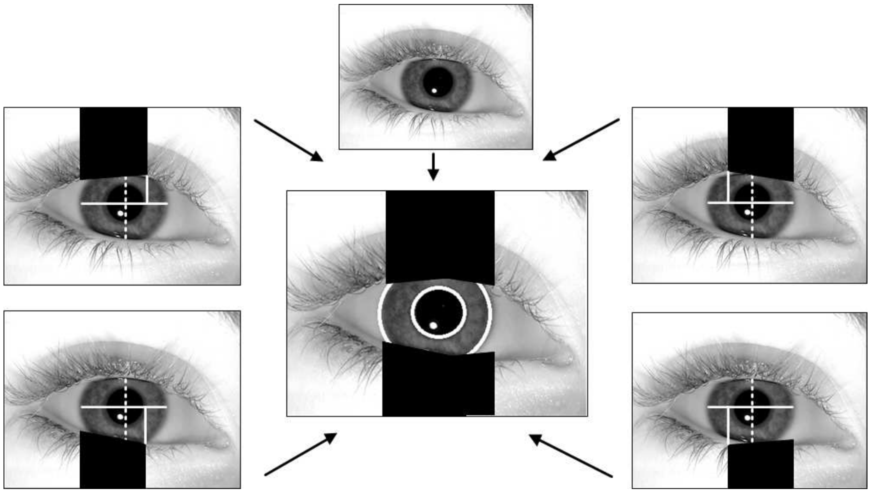


Figure 5.8 Detection of iris region occlusion from the upper and lower eyelid.

limbic circles.

Additionally, another efficient technique is applied to detect the limbic boundary. The Hough transform implemented by the irisBEE method (Phillips et al. 2008) searches the maximum in the parameter space to detect the circle. A technique based on the Hough transform is implemented in order to increase the accuracy of the correct limbic boundary detection by 1.3% for the right eye and by 1.4% for the left eye images. Specifically, when the distance between the center of the detected circle and the center of the detected pupil is more than a predefined threshold value, then the detected circle is rejected. Out of all the non-rejected circles, the one that corresponds to the maximum in the parameter space of the Hough transform and has center coordinates within a predefined threshold value from the pupil center is selected. As a result, proposed heuristic method considerably increases the accuracy of the Hough transform. In Figure 5.7(c) the center of the pupil is pointed in yellow, the incorrect limbic boundary circle with center in green is rejected as it is farther away from the pupil center when compared to the acceptable limbic detection with center displayed in white.

Here the results of time complexity analysis for the segmentation stage are presented. Mean average implementation time of the segmentation stage for an image is com-

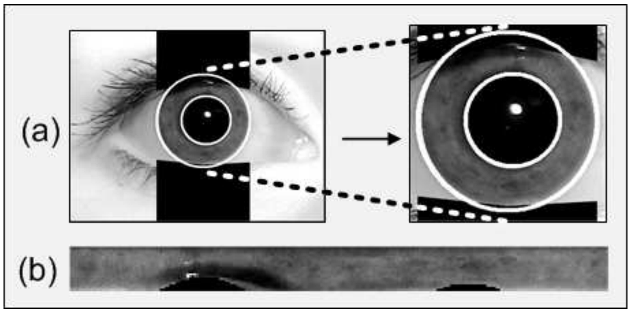


**Figure 5.9** (a) Segmented iris region and (b) its normalized iris region.

puted. First the average processing time for an image from all 2,953 images is computed and then ten runs are made to obtain the mean average implementation time for an image. The experiment is performed on a 3.00 GHz Pentium 4 CPU with 3.2 GB RAM running on Linux. By reducing the search space for Hough transform proposed algorithm takes 748 milliseconds to process an image against 853 milliseconds without this optimization. Hence, proposed efficient limbic boundary detection technique decreases the processing time by approximately 12%.

### *Improved Eyelid Detection*

One of the earlier system proposed by Wildes (1997) detects the eyelids by applying edge detection filter and Hough transform to obtain edge points and then fits parameterized parabolic arcs to localize the upper and lower eyelids. Eyelid detection is implemented by modeling the eyelid as two straight lines using linear Hough transform. Additionaly, power-law transformation is performed on an eye image in order to enhance the contrast between the eyelid and the iris/pupil region. Hence, occlusion from eyelids is more accurately detected.

Furthermore, iris region is split horizontally and vertically resulting in four windows. See Figure 5.8 for the result of eyelid detection. In order to detect the upper eyelid search is performed within the top half of the iris region. Furthermore, to detect the top

**Table 5.1** Correctness of Segmentation (%) for the Pupil and the Iris Region at Various Values of $\rho$

| $\rho$ | Right Eye | | Left Eye | |
|---|---|---|---|---|
| | Pupil Region | Iris Region | Pupil Region | Iris Region |
| 0.7 | 96.3 | 95.5 | 96.8 | 96.0 |
| 1.0 | 98.3 | 97.4 | 98.6 | 97.7 |
| 1.3 | 98.9 | 98.0 | 99.2 | 98.1 |
| 1.6 | 99.2 | 98.2 | 99.5 | 98.4 |
| **1.9** | **99.7** | **98.5** | **99.9** | **98.8** |
| 2.2 | 99.6 | 98.4 | 99.9 | 98.8 |
| 2.5 | 99.6 | 98.4 | 99.8 | 98.7 |

left of the eyelid region, only the left three quarters of the top half of the iris region is considered. The top right of the eyelid region is detected within the right three quarters of the top half of the iris region. Thus, there is an overlap of a portion of the iris region between two splits. Similarly, the bottom eyelid is detected in the lower two windows. After detecting the eyelid in each of these four windows the results are connected together. Proposed approach has shown to improve performance when compared to the method in (Liu et al. 2005) where the splits do not overlap.

### 5.1.3 Feature Encoding and Matching

The feature encoding stage encodes the iris image texture patterns into iriscodes using filters. Iris region is normalized to a constant dimension before encoding. Denoising of the noise regions in the normalized pattern is implemented by means of averaging. This results in a bitwise template, which contains iris information and a noise mask for corrupt areas within the iris pattern. Figure 5.9 shows the result of the normalization of the iris region.

Encoding is implemented by convolving the normalized iris pattern with the 1D Log-Gabor wavelets (Masek 2003). The frequency response of a Log-Gabor filter is given

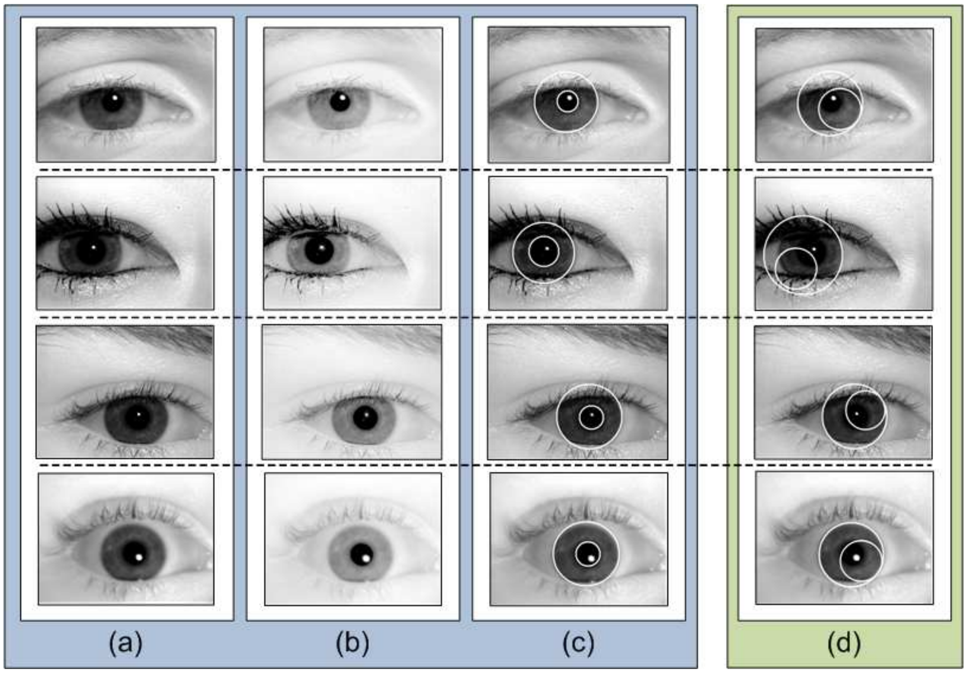


**Figure 5.10** Comparison of the pupil segmentation performance of the proposed improved method with the irisBEE method (Phillips et al. 2008). (a) Input eye images, (b) images after the power-law transformation, (c) examples of correct segmentation of the pupil and iris region by proposed method, and (d) examples of incorrect segmentation by the irisBEE method (Phillips et al. 2008).

as:

$$G(f) = \exp\left( \frac{-(\log(f/f_0))^2}{2(\log(\sigma/f_0))^2} \right) \tag{5.4}$$

where $f_0$ represents the centre frequency, and $\sigma$ gives the bandwidth of the filter. Details of the Log-Gabor filter are given by Field (1987).

Hamming distance is used to measure the similarity of the two iris templates. The Hamming distance defines the similarity between two iriscodes, and the two iriscodes are a match when their Hamming distance is close to each other. In comparing the bit patterns $X$ and $Y$, the Hamming distance (HD) is defined as the sum of disagreeing bits (sum of the XOR between $X$ and $Y$) over $N$, which is the total number of bits in the bit pattern. Below is the formula:

$$HD = \frac{1}{N} \sum_{j=1}^{N} X_j \oplus Y_j \tag{5.5}$$

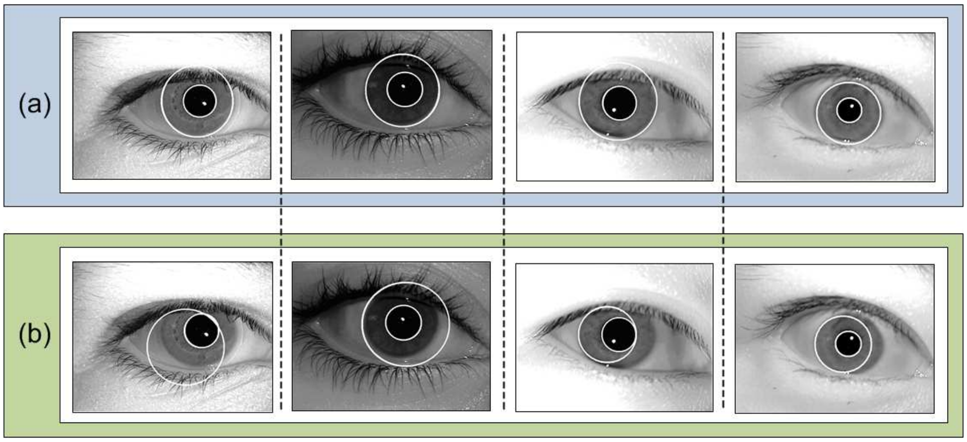


**Figure 5.11** Comparison of the limbic boundary segmentation performance of the proposed improved method with the irisBEE method (Phillips et al. 2008). (a) Examples of correct segmentation by proposed method and (b) examples of incorrect segmentation by the irisBEE method (Phillips et al. 2008).

Noise bits in the two templates are discarded. The iris template is shifted bit-wise from -15 degrees to +15 degrees with an increment of 1.5 degrees each time, and the Hamming distance is computed for two shift positions. The lowest Hamming distance is the best match between the two templates. As suggested by Daugman (2002) such shifting is necessary to take care of the misalignment in the normalized iris patterns caused by the rotational differences during imaging.

## 5.2 Experimental Results

Here, the details of the experimental evaluation of the proposed method are presented on the ICE dataset. In order to make a through comparative assessment of the performance of proposed method with other methods, three sets of experiments for the right eye and the left eye are performed. First the correctness of iris segmentation is assessed, followed by an assessment of the rank-one recognition performance and finally the verification performance for the right eye and the left eye is evaluated according to the experimental setup proposed by the ICE system. The rank-one recognition criterion and the verification criterion evaluate the performance of proposed method for recognition from two different viewpoints;

**Table 5.2** Comparison with the Results (%) from the irisBEE Method (Phillips et al. 2008) of the Correctness of Segmentation for the Pupil Region, Limbic Boundary and Iris Region

| Regions | Proposed Method | | irisBEE Method | |
|---|---|---|---|---|
| | Right Eye | Left Eye | Right Eye | Left Eye |
| Pupil Region | **99.7** | **99.9** | 95.4 | 95.7 |
| Limbic Boundary | **98.7** | **99.0** | 93.4 | 93.6 |
| Iris Region | **98.5** | **98.8** | 90.2 | 90.5 |

more details are provided later in this Section. For all experiments discussed here the input image is scaled to 0.4 of its original size, this significantly cuts down the processing time without compromising the correctness of the results.

### 5.2.1 Assessing the Correctness of Segmentation

The first set of experiments is designed to assess the correctness of segmentation for the pupil region, the limbic boundary and the iris region on the right eye and the left eye. Considering the nature of the ICE dataset, definition of the correctness of segmentation is provided here and it is based on the assumption that the pupil and iris can be modeled as a circle. The pupil region is said to be correctly segmented when the circle fully encloses the pupil region and does not include any area other than the dark pupil. Incorrect pupil segmentation may cover parts of the iris region and or only enclose the pupil region partially. Refer to Section 5.1.2 for the discussion on the method and Figure 5.10(c) and (d) for the results. The limbic boundary is said to be correctly segmented when the circle fully bounds the iris region from outside and does not include any area outside of the iris region other than the pupil or the eyelids that may sometimes occlude the iris. Incorrect limbic boundary segmentation may cover parts of the sclera region and or only enclose the iris region partially. Refer to Section 5.1.2 for the discussion on the method and Figure 5.11(a) and (b) for the results. The iris region is said to be correctly segmented when for any given eye image both the pupil and the limbic boundary are correctly detected.

**Table 5.3** Rank-one Recognition Performance (%) at Various Values of $\rho$

| $\rho$ | Right Eye | Left Eye |
|---|---|---|
| 0.7 | 95.4 | 95.9 |
| 1.0 | 97.6 | 98.1 |
| 1.3 | 98.3 | 98.5 |
| 1.6 | 98.7 | 98.8 |
| **1.9** | **99.0** | **99.0** |
| 2.2 | 98.9 | 99.0 |
| 2.5 | 98.9 | 98.9 |

Table 5.1 gives the results of the correctness of the pupil and iris region segmentation. The power-law transformations are performed for pupil detection on the right and left eye image at various values of $\rho$. At $\rho = 1$ and $c = 1$ the power-law transformation leaves the intensity values of the pixels in the input image unchanged. For values of $\rho > 1$, the power-law transformation enhances the contrast in the dark regions and thereby makes the pupil boundary easier to detect. This is confirmed by the percentage of correct pupil detection as $\rho$ goes higher. Also, for $\rho < 1$, the contrast between the pupil and the surrounding region decreases making it harder to detect the pupil. Best pupil detection results are obtained at $\rho = 1.9$ with close to 100% correct pupil detection for the left eye and 99.7% for the right eye. For the $\rho$ values higher than 1.9 no significant change in the segmentation performance is noticed. The best result for the iris region detection is 98.5% for the right eye and 98.8% for the left eye. The iris region detection is at its highest when the pupil region detection is at its maximum; this is largely due to the fact that for the proposed method the correct detection of iris region is to an extent dependent on the correct pupil region detection. Finally, the iris region detection rate at $\rho = 1.9$ is 1.1% higher for both the right and the left eye when compared with the rate at $\rho = 1$.

Figure 5.10(c) shows examples of correct segmentation of the pupil based on proposed improved pupil region detection method. Input images are shown in Figure 5.10(a) and the result of the power-law transformation can be seen from Figure 5.10(b). Compar-

**Table 5.4** Comparison of the Rank-one Recognition Performance (%) with the Other Methods

| Methods | Right Eye | Left Eye |
|---|---:|---:|
| Proposed Method | **99.0** | **99.0** |
| ND_IRIS (Liu et al. 2005) | - | 97.1 |
| irisBEE method (Phillips et al. 2008) | 95.5 | 96.3 |

ison of results to the incorrect segmentation results of the irisBEE method (Phillips et al. 2008) is shown in Figure 5.10(d). Figure 5.11(a) presents the results of proposed improved limbic boundary segmentation method and a comparison with the incorrect limbic boundary detections of the irisBEE method (Phillips et al. 2008) is shown in Figure 5.11(b).

From Table 5.2 it can be seen that proposed method improves upon the irisBEE method (Phillips et al. 2008) for pupil region segmentation by 4.3% and 4.2% for the right eye and the left eye respectively. Proposed limbic boundary detection rates are higher by 5.3% and 5.4% for the right and left eye respectively. Finally, proposed method leads to an improvement upon the irisBEE method (Phillips et al. 2008) by 8.3% for both the right and the left eye iris region detection.

### 5.2.2 Assessment of the Rank-one Recognition Performance

Here, in order to determine the effectiveness of the proposed method an evaluation is performed based on the rank-one recognition rate. This is a popular evaluation criterion for iris recognition. In order to obtain the recognition rate, first the Hamming distance between every pair of a query image and a target image is calculated, and then the nearest-neighbor classifier is employed for classifying all query images. If the query image and the target image belong to the same subject, then it is considered as a correct match. The recognition rate is the ratio of the number of correctly classified query images to the total number of query images. The rank-one recognition rate underlines the similarity of the samples that are close to one another within a class.

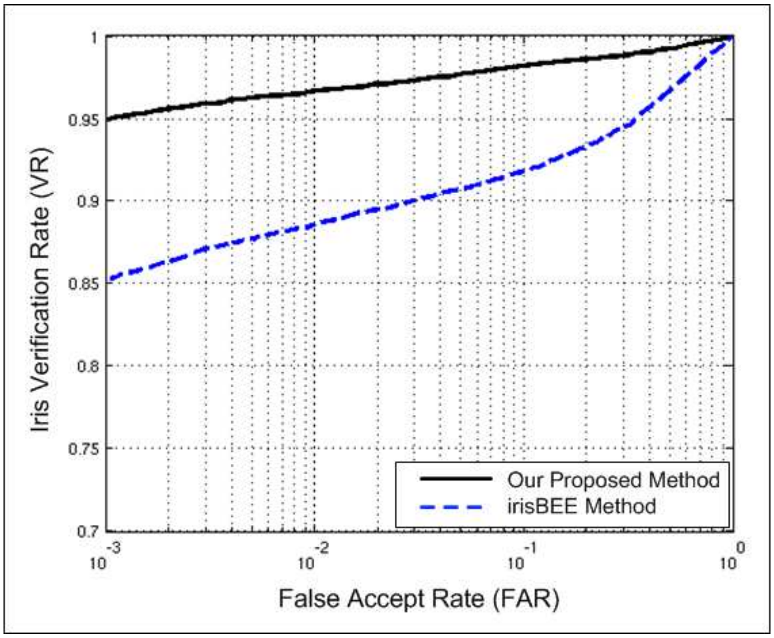


**Figure 5.12** Comparison of the iris verification performance (ROC curve for the right eye) of the irisBEE method (Phillips et al. 2008) with the proposed method.

From Table 5.3 it can be seen that the best recognition rate is 99% for both the right eye and the left eye at $\rho = 1.9$, when compared to the rate at $\rho = 1$, this is higher by 1.4% for the right eye and by 0.9% for the left eye. No significant change is noticed in the recognition performance for $\rho > 1.9$.

The rank-one recognition rate for proposed method as shown in Table 5.4 is 3.5% and 2.7% higher than that of the irisBEE method (Phillips et al. 2008) for the right eye and the left eye respectively. Furthermore, the proposed method improves upon the ND_IRIS (Liu et al. 2005) by a significant 1.9% for the left eye. Note that the authors in (Liu et al. 2005) do not report the recognition rate on the right eye. Additionally, proposed improved eyelid detection method as described in Section 5.1.2 contributes to a performance increase of 1% for both the right and the left eye.

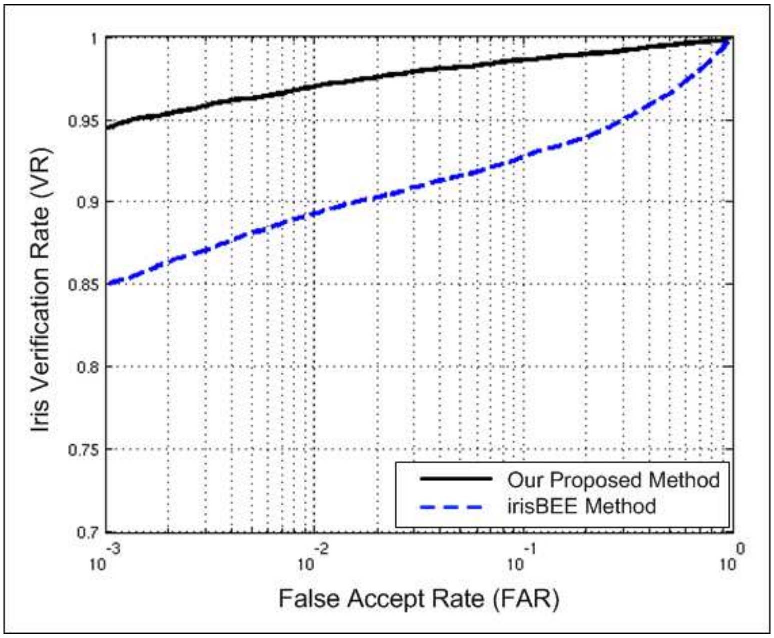


**Figure 5.13** Comparison of the iris verification performance (ROC curve for the left eye) of the irisBEE method (Phillips et al. 2008) with the proposed method.

### 5.2.3 Assessment of the Verification Performance and Equal Error Rate (EER)

For the final set of experiments an evaluation is performed on the verification performance and a comparison is made with the results from the irisBEE method (Phillips et al. 2008). The ICE protocol recommends using the receiver operating characteristic (ROC) curves, which plot the iris verification rate, i.e., the true accept rate versus the false accept rate (FAR), to report the iris recognition performance. The verification rate is the rate at which a matching algorithm correctly determines that a genuine sample matches an enrolled sample. The equal error rate (EER) is obtained when the FAR equals the false reject rate (FRR). Generally, the lower the EER value the higher will be the accuracy of the biometric system.

The ROC curves are automatically generated by the BEE system when a similarity matrix is input to the system. In particular, the BEE system generates two ROC curves, corresponding to the Experiment 1 for the right eye and Experiment 2 for the left eye images. The iris verification rate at the false accept rate of 0.1% is generally used as a standard for performance comparison (Yang et al. 2010).

**Table 5.5** Iris Verification Performance (%) at 0.1% False Accept Rate and EER at Various Values of $\rho$ (VR is the Verification Rate and EER is the Equal Error Rate)

| $\rho$ | Right Eye | | Left Eye | |
|---|---|---|---|---|
| | VR | EER | VR | EER |
| 0.7 | 85.1 | 8.3 | 84.7 | 7.7 |
| 1.0 | 91.3 | 5.2 | 90.9 | 4.6 |
| 1.3 | 92.8 | 4.9 | 92.2 | 4.2 |
| 1.6 | 94.2 | 3.9 | 93.3 | 3.1 |
| **1.9** | **95.1** | **2.8** | **94.4** | **2.3** |
| 2.2 | 95.1 | 2.8 | 94.4 | 2.3 |
| 2.5 | 95.0 | 2.8 | 94.3 | 2.3 |

It should be pointed out that the verification rate in the ICE Experiment 1 and 2 emphasizes the similarity of samples that are relatively distant from one another within a class because it needs to measure all similarity between samples, whereas the recognition rate discussed in Section 5.2.2 emphasizes the similarity of samples that are close to one another within a class since it applies a nearest-neighbor classifier. Therefore, these two criteria evaluate the performance of proposed method for recognition from two different viewpoints.

From Table 5.5 it can be seen that the best verification rate and the lowest EER is achieved at $\rho = 1.9$. When compared to the performance at $\rho = 1$, the VR is higher by 3.8% at a low EER of 2.8% for the right eye and the VR is higher by 3.5% at the EER of 2.3% for the left eye. No significant change is noticed in the verification performance for $\rho > 1.9$.

See Figure 5.12 and Figure 5.13 for a comparison of the performance of proposed method with that of the irisBEE method (Phillips et al. 2008) in terms of the ROC curves. Figure 5.12 and Figure 5.13 show the ROC curves for the right eye experiment and the left eye experiment respectively. It can be observed that the proposed method improves the iris recognition performance significantly in comparison with the irisBEE method (Phillips et al. 2008).

**Table 5.6** Comparison with other Methods on the Iris Verification Performance (%) at 0.1% False Accept Rate and EER (VR is the Verification Rate and EER is the Equal Error Rate)

| Methods | Right Eye | | Left Eye | |
|---|---|---|---|---|
| | VR | EER | VR | EER |
| Proposed Method | 95.1 | **2.8** | 94.4 | **2.3** |
| SAGEM-Iridian† | above 99.0* | - | above 99.0* | - |
| irisBEE Method† | 85.2 | 8.5 | 84.9 | 7.8 |

†Phillips et al. (2008). *Result estimated from Figure 4 in Phillips et al. (2008).

From Table 5.6 it can be seen that proposed method improves upon the irisBEE method (Phillips et al. 2008) notably. For the right eye, the proposed method has a verification rate of 95.1%, which is about 10% higher than the irisBEE method (Phillips et al. 2008). The EER is 2.8%, which is much lower than the 8.5% for the irisBEE method (Phillips et al. 2008). For the left eye, the proposed method has a VR of 94.4%, which is again higher by 9.5% than the irisBEE method (Phillips et al. 2008). The EER is 2.3%, which is much lower than the 7.8% from the irisBEE method (Phillips et al. 2008); this emphasizes the higher accuracy of the proposed method.

# CHAPTER 6

# CONCLUSIONS AND FUTURE WORK

This dissertation investigates advanced techniques in color image search and retrieval. The SIFT descriptor is extended to different color spaces, including the recently proposed oRGB color space, a new oRGB-SIFT feature representation is proposed, furthermore it is integrated with other color SIFT features to produce the Color SIFT Fusion (CSF), the Color Grayscale SIFT Fusion (CGSF), and the CGSF+PHOG descriptors for image category classification. Applications to image search in object, scene, flower, and texture datasets are presented along with special applications to biometrics. Classification is implemented using a novel EFM-KNN classifier, which combines the Enhanced Fisher Model (EFM) and the K Nearest Neighbor (KNN) decision rule. Results of the experiments on the Caltech 256 dataset, the MIT Scene dataset, the Oxford Flower dataset, and the Upol Iris dataset show that the oRGB-SIFT descriptor improves recognition performance upon other color SIFT descriptors, the CSF, the CGSF, and the CGSF+PHOG descriptors perform better than the other color SIFT descriptors. The fusion of both Color SIFT descriptors (CSF) and Color Grayscale SIFT descriptor (CGSF) show significant improvement in the classification performance, which indicates that various color-SIFT descriptors and grayscale-SIFT descriptor are not redundant for image classification.

Four novel color Local Binary Pattern (LBP) descriptors are presented in this dissertation for scene image and image texture classification with applications to image search and retrieval. Specifically, the first color LBP descriptor, the oRGB-LBP descriptor, is derived by concatenating the LBP features of the component images in an opponent color space - the oRGB color space. The other three color LBP descriptors are obtained by the integration of the oRGB-LBP descriptor with some additional image features: the Color LBP Fusion (CLF) descriptor is constructed by integrating the RGB-LBP, the YCbCr-LBP, the HSV-LBP, the rgb-LBP, as well as the oRGB-LBP descriptor; the Color Grayscale LBP

Fusion (CGLF) descriptor is derived by integrating the grayscale-LBP descriptor and the CLF descriptor; and the CGLF+PHOG descriptor is obtained by integrating the Pyramid of Histograms of Orientation Gradients (PHOG) and the CGLF descriptor. Feature extraction applies the Enhanced Fisher Model (EFM) and image classification is based on the nearest neighbor classification rule (EFM-NN). The proposed image descriptors and the feature extraction and classification methods are evaluated using three grand challenge databases: the MIT scene database, the KTH-TIPS2-b database, and the KTH-TIPS materials database. The experimental results show that the proposed oRGB-LBP descriptor improves image classification performance upon other color LBP descriptors, and the CLF, the CGLF, and the CGLF+PHOG descriptors further improve upon the oRGB-LBP descriptor for scene image and image texture classification.

Another major contribution of this dissertation lies in the area of iris image search. A robust iris recognition method with enhanced performance on the ICE dataset is presented. In particular, the power-law transformations for more accurate detection of the pupil region are implemented. Additionally, the proposed technique is able to considerably reduce the candidate limbic boundary search space, leading to a significant increase in the accuracy and speed of the segmentation. The segmentation performance is further increased with the application of the thresholding. Furthermore, for higher accuracy and speed, the limbic circle having a center within close range of the pupil center is selectively detected. Additionally, the proposed improved eyelid detection phase has shown to increase performance. From the experiments, it can be concluded that, using power-law transformations with $\rho = 1.9$ or above, the proposed method show constantly better performance for pupil and iris region segmentation for both left and right eyes using the ICE dataset, consequently comparison studies show improved segmentation performance comparing with the irisBEE method (Phillips et al. 2008) and on the rank-one recognition performance than other methods with improved VR and EER for both eyes.

Future work on color image search would include an evaluation of the color SIFT descriptors and classifier on the KTH-TIPS (Textures under varying Illumination, Pose and Scale) dataset (Hayman et al. 2004) and the KTH-TIPS2-b dataset (Caputo et al. 2005). Additionally, development of more discriminative color descriptors would be looked into in order to improve the classification performance on the more challenging datasets such as the Caltech 256 dataset. One way would be to develop descriptors from hybrid color spaces and normalized color spaces (Yang et al. 2010), along with the Gabor feature representation methods. Furthermore, exploration of advanced score level fusion methodology of the color descriptors for improved classification performance would be looked into.

Future work on iris image search aims to process the visible wavelength iris images, it will leverage previous research results in attenuating illumination variations (Liu 2006), (Yang and Liu 2007), applying novel color models (Liu and Liu 2008b), (Liu and Liu 2008a), (Yang and Liu 2008), (Liu 2008), as well as effective feature extraction and classification techniques (Liu and Wechsler 2001), (Liu and Wechsler 1998a), (Liu and Wechsler 1998b) to process the visible wavelength iris images, such as those in the UBIRIS database (Proenca et al. 2010), which display different characteristics from the IR or near IR iris images in the ICE database. On the one hand, the visible wavelength iris images possess additional classification cues, such as color. But on the other hand, they pose challenging issues as well, such as large variations in lighting conditions.

# REFERENCES

S. Agarwal and D. Roth. Learning a sparse representation for object detection. In *European Conference on Computer Vision*, volume 4, pages 113–130, Copenhagen, Denmark, May 27 - June 2 2002.

M. Aizerman, E. Braverman, and L. Rozonoer. Theoretical foundations of the potential function method in pattern recognition learning. *Automation and Remote Control*, 25:821–837, 1964.

S. Baker, A. Hentz, K. Bowyer, and P. Flynn. Degradation of iris recognition performance due to non-cosmetic prescription contact lenses. *Computer Vision and Image Understanding*, 114(9):1030–1044, 2010.

S. Banerji, A. Verma, and C. Liu. Novel color LBP descriptors for scene and image texture classification. In *15th Intl. Conf. on Image Processing, Computer Vision, and Pattern Recognition*, Las Vegas, Nevada, July 18-21 2011.

H. Bay, T. Tuytelaars, and L. Van Gool. SURF: Speeded up robust features. *Computer Vision and Image Understanding*, 110(3):346–359, 2008.

A. Bosch, A. Zisserman, and X. Munoz. Representing shape with a spatial pyramid kernel. In *Int. Conf. on Image and Video Retrieval*, pages 401–408, Amsterdam, The Netherlands, July 9-11 2007.

A. Bosch, A. Zisserman, and X. Munoz. Scene classification using a hybrid generative/discriminative approach. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 30(4):712–727, 2008.

B.E. Boser, I.M. Guyon, and V.N. Vapnik. A training algorithm for optimal margin classifiers. In ACM *Workshop on* COLT, pages 144–152, 1992.

K.W. Bowyer, K. Hollingsworth, and P.J. Flynn. Image understanding for iris biometrics: A survey. *Computer Vision Image Understanding*, 110(2):281–307, 2008.

M. Bratkova, S. Boulos, and P. Shirley. oRGB: A practical opponent color space for computer graphics. *IEEE Computer Graphics and Applications*, 29(1):42–55, 2009.

G. Burghouts and J.-M. Geusebroek. Performance evaluation of local color invariants. *Computer Vision and Image Understanding*, 113:48–62, 2009.

B. Caputo, E. Hayman, and P. Mallikarjuna. Class-specific material categorisation. In *IEEE Int. Conf. on Computer Vision*, pages 1597–1604, Beijing, China, October 17-20 2005.

C-C. Chang and C-J. Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2(3):1–27, 2011.

J. Chen, S. Shan, C. He, G. Zhao, M. Pietikainen, X. Chen, and W. Gao. WLD: A robust local image descriptor. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 32(9):1705–1720, 2010.

M. Crosier and L.D. Griffin. Texture classification with a dictionary of basic image features. In *Conf. on Computer Vision and Pattern Recognition*, pages 1–7, Anchorage, Alaska, June 24-26 2008.

G. Csurka, C. Bray, C. Dance, and L. Fan. Visual categorization with bags of keypoints. In *Proc. Workshop Statistical Learning in Computer Vision*, pages 1–22, 2004.

R. Datta, D. Joshi, J. Li, and J. Wang. Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys*, 40(2):509–522, 2008.

J. Daugman. How iris recognition works. In *IEEE Int. Conf. on Image Processing*, pages 33–36, Rochester, New York, September 22-25 2002.

J. Daugman. Probing the uniqueness and randomness of iriscodes: Results from 200 billion iris pair comparisons. *Proc. of the IEEE*, 94(11):1927–1935, 2006.

J. Daugman. New methods in iris recognition. *IEEE Trans. on Systems, Man and Cybernetics B*, 37(5):1167–1175, 2007.

M. Dobes, J. Martinek, D. Skoupil, Z. Dobesova, and J. Pospisil. Human eye localization using the modified Hough transform. *Optik*, 117(10):468–473, 2006.

Y. Du, R.W. Ives, and D.M. Etter. Iris recognition: A chapter on biometrics. In *The Electrical Engineering Handbook*. CRC Press, Florida, 2004.

R. Fergus, P. Perona, and A. Zisserman. Object class recognition by unsupervised scale-invariant learning. In *IEEE Conf. on Computer Vision and Pattern Recognition*, volume 2, pages 264–271, Madison, Wisconsin, June 16-22 2003.

D.J. Field. Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America*, 12(4):2379–2394, 1987.

L. Flom and A. Safir. Iris recognition system. *U.S. Patent 4,641,349*, 1987.

D. Forsyth and J. Ponce. *Computer Vision: A Modern Approach*. Prentice Hall, NJ, 2003.

K. Fukunaga. *Introduction to Statistical Pattern Recognition*. Academic Press, San Diego, California, USA, 2nd edition, 1990.

J. Gemert, J-M. Geusebroek, C. Veenman, and A. Smeulders. Kernel codebooks for scene categorization. In *European Conf. on Computer Vision*, pages 696–709, Singapore, December 4-6 2008.

J.E. Gentile, N. Ratha, and J. Connell. SLIC: Short length iris code. In *Int. Conf. on Biometrics: Theory, Applications and Systems*, pages 1–7, Washington, D.C., September 28-30 2009.

T. Gevers, J. van de Weijer, and H. Stokman. Color feature detection: An overview. In R. Lukac and K.N. Plataniotis, editors, *Color Image Processing: Methods and Applications*. CRC Press, University of Toronto, Ontario, Canada, 2006.

C.G. Gonzalez and R.E. Woods. *Digital Image Processing*. Prentice Hall, Upper Saddle River, NJ, USA, 2001.

K. Grauman and T. Darrell. Pyramid match kernels: Discriminative classification with sets of image features. In *Int. Conference on Computer Vision*, volume 2, pages 1458–1465, Beijing, October 17-20 2005.

G. Griffin, A. Holub, and P. Perona. Caltech-256 object category dataset. Technical report, California Institute of Technology, 2007.

F. Hao, J. Daugman, and P. Zielinski. A fast search algorithm for a large fuzzy database. *IEEE Trans. on Information Forensics and Security*, 2(3):203–212, 2008.

C. Harris and M. Stephens. A combined corner and edge detector. In *Alvey Vision Conference*, pages 147–151, 1988.

E. Hayman, B. Caputo, M. Fritz, and J-O. Eklundh. On the significance of real-world conditions for material classification. In *European Conference on Computer Vision*, pages 253–266, Prague, Czech Republic, May 11-14 2004.

Z. He, T. Tan, Z. Sun, and X. Qiu. Toward accurate and fast iris segmentation for iris biometrics. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 31(9):1670–1684, 2009.

F. Jurie and B. Triggs. Creating efficient codebooks for visual recognition. In *Int. Conference on Computer Vision*, pages 604–610, Beijing, October 17-20 2005.

Y. Ke and R. Sukthankar. PCA-SIFT: A more distinctive representation for local image descriptors. In *IEEE Conf. on Computer Vision and Pattern Recognition*, volume 2, pages 506–513, Washington, D.C., June 27 - July 2 2004.

S. Kondra and V. Torre. Texture classification using three circular filters. In *IEEE Indian Conf. on Computer Vision, Graphics and Image Processing*, pages 429–434, Bhubaneshwar, India, December 16-19 2008.

S. Lazebnik, C. Schmid, and J. Ponce. A sparse texture representation using affine-invariant regions. In *IEEE Conf. on Computer Vision and Pattern Recognition*, volume 2, pages 319–324, Madison, Wisconsin, June 16-22 2003.

S. Lazebnik, C. Schmid, and J. Ponce. Semi-local affine parts for object recognition. In *British Machine Vision Conference*, volume 2, pages 959–968, London, September 7-9 2004.

T.K. Leung and J. Malik. Representing and recognizing the visual appearance of materials using three-dimensional textons. *Int. Journal of Computer Vision*, 43(1):29–44, 2001.

T. Lindeberg and J. Garding. Shape-adapted smoothing in estimation of 3D shape cues from affine deformations of local 2D brightness structure. *Image and Vision Computing*, 15(6):415–434, 1997.

C. Liu. A Bayesian discriminating features method for face detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25(6):725–740, 2003.

C. Liu. Enhanced independent component analysis and its application to content based face image retrieval. *IEEE Trans. Systems, Man, and Cybernetics, Part B: Cybernetics*, 34(2):1117–1127, 2004a.

C. Liu. Gabor-based kernel with fractional power polynomial models for face recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 26(5):572–581, 2004b.

C. Liu. Capitalize on dimensionality increasing techniques for improving face recognition grand challenge performance. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 28(5):725–737, 2006.

C. Liu. The Bayes decision rule induced similarity measures. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 6(29):1116–1117, 2007.

C. Liu. Learning the uncorrelated, independent, and discriminating color spaces for face recognition. *IEEE Trans. on Information Forensics and Security*, 3(2):213–222, 2008.

C. Liu and H. Wechsler. Probabilistic reasoning models for face recognition. In *Proc. Computer Vision and Pattern Recognition*, pages 827–832, Santa Barbara, CA, June 23-25 1998a.

C. Liu and H. Wechsler. Unified Bayesian framework for face recognition. In *Proc. IEEE Int. Conf. on Image Processing*, pages 151–155, Chicago, IL, October 4-7 1998b.

C. Liu and H. Wechsler. Evolutionary pursuit and its application to face recognition. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22(6):570–582, 2000a.

C. Liu and H. Wechsler. Robust coding schemes for indexing and retrieval from large face databases. *IEEE Trans. on Image Processing*, 9(1):132–137, 2000b.

C. Liu and H. Wechsler. A shape and texture based enhanced Fisher classifier for face recognition. *IEEE Trans. on Image Processing*, 10(4):598–608, 2001.

C. Liu and H. Wechsler. Gabor feature based classification using the enhanced Fisher linear discriminant model for face recognition. *IEEE Trans. on Image Processing*, 11(4): 467–476, 2002.

C. Liu and H. Wechsler. Independent component analysis of Gabor features for face recognition. *IEEE Trans. on Neural Networks*, 14(4):919–928, 2003.

C. Liu and J. Yang. ICA color space for pattern recognition. *IEEE Trans. on Neural Networks*, 2(20):248–257, 2009.

X. Liu, K.W. Bowyer, and P.J. Flynn. Experiments with an improved iris segmentation algorithm. In *IEEE Workshop on Automatic Identification Advanced Technologies*, pages 118–123, New York, October 16-18 2005.

Z. Liu and C. 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, 2008a.

Z. Liu and C. Liu. A hybrid color and frequency features method for face recognition. *IEEE Trans. on Image Processing*, 17(10):1975–1980, 2008b.

D.G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. Journal of Computer Vision*, 60(2):91–110, 2004.

L. Ma, T. Tan, Y. Wang, and D. Zhang. Efficient iris recognition by characterizing key local variations. *IEEE Trans. Image Processing*, 13(6):739–750, 2004.

L. Masek. Recognition of human iris patterns for biometric identification. B. Eng. thesis, The University of Western Australia, Western Australia, 2003.

K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 27(10):1615–1630, 2005.

K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool. A comparison of affine region detectors. *Int. Journal of Computer Vision*, 65(1-2):43–72, 2005.

W. Niblack, R. Barber, and W. Equitz. The QBIC project: Querying images by content using color, texture and shape. In *SPIE Conference on Geometric Methods in Computer Vision II*, pages 173–187, 1993.

M-E. Nilsback and A. Zisserman. A visual vocabulary for flower classification. In *Int. Conf. on Computer Vision and Pattern Recognition*, pages 1447–1454, Washington, DC, June 17-22 2006.

T. Ojala, M. Pietikainen, and D. Harwood. Performance evaluation of texture measures with classification based on Kullback discrimination of distributions. In *Int. Conf. on Pattern Recognition*, pages 582–585, Jerusalem, Israel, 1994.

T. Ojala, M. Pietikainen, and D. Harwood. A comparative study of texture measures with classification based on feature distributions. *Pattern Recognition*, 29(1):51–59, 1996.

A. Oliva and A. Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *Int. Journal of Computer Vision*, 42(3):145–175, 2001.

P.J. Phillips. FRGC and ICE workshop. Technical report, National Institute of Standards and Technology (NIST), USA, 2006.

P.J. Phillips, K.W. Bowyer, P.J. Flynn, X. Liu, and W.T. Scruggs. The iris challenge evaluation 2005. In *IEEE Second Int. Conf. on Biometrics: Theory, Applications and Systems*, Washington, D.C., September 29 - October 1 2008.

M. Pontil and A. Verri. Support vector machines for 3D object recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 20(6):637–646, 1998.

H. Proenca. *Towards Non-Cooperative Biometric Iris Recognition*. PhD thesis, University of Beira Interior, Covilha, Portugal, October 2006.

H. Proenca. Iris recognition: On the segmentation of degraded images acquired in the visible wavelength. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 32 (8):1502–1516, 2010.

H. Proenca, S. Filipe, R. Santos, J. Oliveira, and L.A. Alexandre. The UBIRIS.v2: A database of visible wavelength images captured on-the-move and at-a-distance. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 32(8):1529–1535, 2010.

H. Proenca and A. Luis. Toward noncooperative iris recognition: A classification approach using multiple signatures. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 29:607–612, 2007.

N.B. Puhan, N. Sudha, and S.K. Anirudh. Efficient segmentation technique for noisy frontal view iris images using Fourier spectral density. *Signal Image and Video Processing*, 5(1):105–119, 2011.

B. Schiele and J. Crowley. Recognition without correspondence using multidimensional receptive field histograms. *Int. Journal of Computer Vision*, 36(1):31–50, 2000.

P. Shih and C. Liu. Comparative assessment of content-based face image retrieval in different color spaces. *Int. Journal of Pattern Recognition and Artificial Intelligence*, 19(7):873–893, 2005.

P. Shih and C. Liu. An effective color feature extraction method using evolutionary computation for face recognition. *Int. Journal of Biometrics*, 3(3):206–227, 2011.

C. Shuo and C. Liu. Eye detection using color information and a new efficient SVM. In *Int. Conf. on Biometrics Theory, Applications and Systems*, Washington, D.C., September 27-29 2010.

A.R. Smith. Color gamut transform pairs. *Computer Graphics*, 12(3):12–19, 1978.

H. Stokman and T. Gevers. Selection and fusion of color models for image feature detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 29(3):371–381, 2007.

M.J. Swain and D.H. Ballard. Color indexing. *Int. Journal of Computer Vision*, 7(1):11–32, 1991.

J. Thornton, M. Savvides, and B.V.K. Vijayakumar. A Bayesian approach to deformed pattern matching of iris images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 4(29):596–606, 2007.

M. Varma and D. Ray. Learning the discriminative power-invariance trade-off. In *IEEE Int. Conf. on Computer Vision*, pages 1–8, Rio de Janeiro, October 14-20 2007.

M. Vatsa, R. Singh, and A. Noore. Improving iris recognition performance using segmentation, quality enhancement, match score fusion, and indexing. *IEEE Trans. on Systems, Man, and Cybernetics - Part B: Cybernetics*, 38(4):1021–1035, 2008.

A. Verma, S. Banerji, and C. Liu. A new color SIFT descriptor and methods for image category classification. In *International Congress on Computer Applications and Computational Science*, pages 819–822, Singapore, December 4-6 2010.

A. Verma and C. Liu. Efficient iris identification with improved segmentation techniques. In V.K. Mago and N. Bhatia, editors, *Cross-Disciplinary Applications of Artificial Intelligence and Pattern Recognition: Advancing Technologies*. IGI Global, Hershey, PA, USA, 2011a. (to appear).

A. Verma and C. Liu. Fusion of color SIFT features for image classification with applications to biometrics. In *11th IAPR International Conference on Pattern Recognition and Information Processing*, Minsk, Belarus, May 18-20 2011b.

A. Verma and C. Liu. Novel EFM-KNN classifier and a new color descriptor for image classification. In *20th IEEE Wireless and Optical Communications Conference (Multimedia Services and Applications)*, Newark, New Jersey, USA, April 15-16 2011c.

A. Verma, C. Liu, and J. Jia. Iris recognition based on robust iris segmentation and image enhancement. *Int. Journal of Biometrics*, 2011a. (to appear).

A. Verma, C. Liu, and J. Jia. New color SIFT descriptors for image classification with applications to biometrics. *Int. Journal of Biometrics*, 1(3):56–75, 2011b.

M. Weber, M. Welling, and P. Perona. Towards automatic discovery of object categories. In *IEEE Conf. on Computer Vision and Pattern Recognition*, volume 2, pages 2101–2109, Hilton Head, SC, June 13-15 2000.

R.P. Wildes. Iris recognition: An emerging biometric technology. *Proc. of the IEEE*, 85 (9):1348–1363, 1997.

J. Yang and C. Liu. Horizontal and vertical 2D PCA-based discriminant analysis for face verification on a large-scale database. *IEEE Trans. on Information Forensics and Security*, 2(4):781–792, 2007.

J. Yang and C. Liu. Color image discriminant models and algorithms for face recognition. *IEEE Trans. on Neural Networks*, 19(12):2088–2098, 2008.

J. Yang, C. Liu, and L. Zhang. Color space normalization: Enhancing the discriminating power of color spaces for face recognition. *Pattern Recognition*, 4(43):1454–1466, 2010.

J. Yang, K. Yu, Y. Gong, and T. Huang. Linear spatial pyramid matching using sparse coding for image classification. In *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 1794–1801, Singapore, December 4-6 2009.

J. Zhang, M. Marszalek, S. Lazebnik, and C. Schmid. Local features and kernels for classification of texture and object categories: A comprehensive study. *Int. Journal of Computer Vision*, 73(2):213–238, 2007.

C. Zhu, C. Bichot, and L. Chen. Multi-scale color local binary patterns for visual object classes recognition. In *Int. Conf. on Pattern Recognition*, pages 3065–3068, Istanbul, Turkey, August 23-26 2010.