Int. J. Biometrics, Vol. x, No. x, xxxx

New Color SIFT Descriptors for Image Classification with Applications to Biometrics

Abhishek Verma and Chengjun Liu

Department of Computer Science
New Jersey Institute of Technology
Newark, NJ 07102, USA
E-mail:{av56, chengjun.liu}@njit.edu

Jiancheng (Kevin) Jia

International Game Technology
Reno, NV 89521, USA
E-mail: kevin.jia@igt.com

Abstract: This paper first presents a new oRGB-SIFT descriptor, and then integrates it with other color SIFT features to produce the novel Color SIFT Fusion (CSF) and the Color Grayscale SIFT Fusion (CGSF) descriptors for image classification 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. The effectiveness of the proposed descriptors and classification method are evaluated using 20 image categories from two large scale, grand challenge datasets: the Caltech 256 database and the UPOL Iris database. The experimental results show that (i) the proposed oRGB-SIFT descriptor slightly improves recognition performance upon other color SIFT descriptors; and (ii) both the CSF and the CGSF descriptors perform better than the other color SIFT descriptors.

Keywords: oRGB-SIFT descriptor, Color SIFT Fusion (CSF), Color Grayscale SIFT Fusion (CGSF), EFM-KNN classifier, image classification, biometrics.

Reference to this paper should be made as follows: A. Verma, C. Liu, and J. Jia (xxxx) 'New Color SIFT Descriptors for Image Classification with Applications to Biometrics', *Int. J. Biometrics*, Vol. x, No. x, pp.xxx–xxx.

Biographical notes: Abhishek Verma received the Master of Computer Applications (MCA) degree from Bangalore University, Bangalore, India, in 2003, and the MS in Computer Science from New Jersey Institute of Technology, NJ, USA, in 2006. He is currently working towards the Ph.D. degree in Computer Science at New Jersey Institute of Technology. His research interests include object and scene classification, pattern recognition, content-based image retrieval systems, image processing, and iris recognition.

Chengjun Liu received the Ph.D. from George Mason University in 1999, and he is presently an Associate Professor of Computer Science and the director of the Face Recognition and Video Processing Lab at New Jersey Institute of Technology. His research interests are in Pattern Recognition (Face/Iris Recognition), Machine Learning (Statistical Learning, Kernel Methods, Similarity Measures), Computer Vision (Object/Face Detection, Video Processing), Security (Biometrics), and Image Processing (New Color Spaces, Gabor Image Representation). His recent research has been concerned with the development of novel and robust methods for image/video retrieval and object detection, tracking and recognition based upon statistical and machine learning concepts. The class of new methods he has developed includes the Bayesian Discriminating Features method (BDF), the Probabilistic Reasoning Models (PRM), the Enhanced Fisher Models (EFM), the Enhanced Independent Component Analysis (EICA), the Shape and Texture-based Fisher method (STF), the Gabor-Fisher Classifier (GFC), and the Independent Gabor Features (IGF) method. He has also pursued the development of novel evolutionary methods leading to the development of the Evolutionary Pursuit (EP) method for pattern recognition in general, and face recognition in particular.

Jiancheng (Kevin) Jia received his Ph.D. from Purdue University in 1991. He is currently a test engineer at International Game Technology, USA. His research interests include biometrics, data representation with neural networks, pattern recognition and machine vision.

1 Introduction

Color features provide powerful information for biometric image classification, indexing, and retrieval [19], [27], [23], as well as for identification of object and natural scene categories 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 the color space that produces the best result 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 [4], [3], [25].

Recently, there has been much emphasis on the detection and recognition of locally affine invariant regions [20], [22], [2]. 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 [20], [22], [2].

Affine region detectors when combined with the intensity Scale-Invariant Feature Transform (SIFT) descriptor [20] has been shown to outperform many alternatives [22].

In this paper, we extend the SIFT descriptor to different color spaces, including the recently proposed oRGB color space [2] and propose a new oRGB-SIFT feature representation, and then integrate it with other color SIFT features to produce the Color SIFT Fusion (CSF), and the Color Grayscale SIFT Fusion (CGSF) descriptors for image category classification with special applications to biometrics. Classification is implemented using a novel EFM-KNN classifier [17], [15], which combines the Enhanced Fisher Model (EFM) and the K Nearest Neighbor (KNN) decision rule [5]. The effectiveness of the proposed descriptors and classification method will be evaluated using 20 image categories from two large scale, grand challenge datasets: the Caltech 256 database and the UPOL Iris database.

2 Related Work

This section briefly surveys the recent work on biometric image retrieval and object and scene recognition. In recent years, use of color as a means to biometric image recognition [19], [12], [23] and object and scene classification has gained popularity. Color features can capture discriminative information by means of the color invariants, color histogram, color texture, etc. 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 [26]. More recent work on color based image classification appears in [19], [27], [13] that propose several new color spaces and methods for face classification and in [1] the HSV color space is used for the scene category recognition. Evaluation of local color invariant descriptors is performed in [3]. Fusion of color models, color region detection and color edge detection have been investigated for representation of color images [25]. Key contributions in color, texture, and shape abstraction have been discussed in Datta et al. [4].

Efficient retrieval requires a robust feature extraction method that has the ability to learn meaningful low-dimensional patterns in spaces of very high dimensionality [9], [18], [14]. 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 [15], [11]. Recently, Support Vector Machine (SVM) classifier for multiple category recognition has gained popularity [28], [1] though it suffers from the drawback of being computationally too expensive on large scale image classification tasks. The EFM classifier has achieved good success for the task of image based recognition [17], [16], [10].

3 New Color SIFT Descriptors

We first review in this section five color spaces in which our new color SIFT descriptors are defined, and then discuss five conventional SIFT descriptors: the

RGB-SIFT, the rgb-SIFT, the HSV-SIFT, the YCbCr-SIFT, and the grayscale-SIFT descriptors. We finally present three new color SIFT descriptors: the oRGB-SIFT, the Color SIFT Fusion (CSF), and the Color Grayscale SIFT Fusion (CGSF) descriptors for image classification with applications to biometrics.

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:

$$\begin{aligned} r &= R/(R+G+B) \\ g &= G/(R+G+B) \\ b &= B/(R+G+B) \end{aligned} \tag{1}$$

Due to the normalization r and g are scale-invariant and thereby invariant to light intensity changes, shadows and shading [6].

The HSV color space is motivated by human vision system because human describes color by means of hue, saturation, and brightness. Hue and saturation define chrominance, while intensity or value specifies luminance [7]. The HSV color space is defined as follows [24]:

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

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

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 65.4810 & 128.5530 & 24.9660 \\ -37.7745 & -74.1592 & 111.9337 \\ 111.9581 & -93.7509 & -18.2072 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{3}$$

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

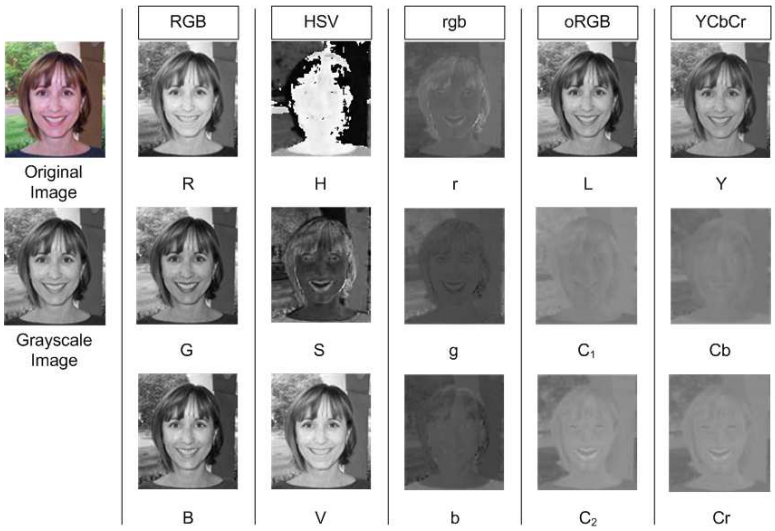

Figure 1 Color component images in the five color spaces: RGB, HSV, rgb, oRGB, and YCbCr. The color image is from the Caltech 256 dataset, whose grayscale image is displayed as well.

The oRGB color space [2] 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 L channel contains the luminance information and its values range between [0, 1]:

$$\begin{bmatrix} L \\ C1 \\ C2 \end{bmatrix} = \begin{bmatrix} 0.2990 & 0.5870 & 0.1140 \\ 0.5000 & 0.5000 & -1.0000 \\ 0.8660 & -0.8660 & 0.0000 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{4}$$

Fig. 1 shows the color component images in the five color spaces: RGB, HSV, rgb, oRGB, and YCbCr.

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 [20]. 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 three 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 and the grayscale-SIFT descriptor.

4 The Novel EFM-KNN Classifier

Image classification using the new descriptors introduced in the preceding section is implemented using a novel EFM-KNN classifier [17], [15], which combines the Enhanced Fisher Model (EFM) and the K Nearest Neighbor (KNN) decision rule [5]. Let $\mathcal{X} \in \mathbb{R}^N$ be a random vector whose covariance matrix is $\Sigma_{\mathcal{X}}$:

$$\Sigma_{\mathcal{X}} = \mathcal{E}\{[\mathcal{X} - \mathcal{E}(\mathcal{X})][\mathcal{X} - \mathcal{E}(\mathcal{X})]^t\} \tag{5}$$

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

$$\Sigma_{\mathcal{X}} = \Phi \Lambda \Phi^t \tag{6}$$

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:

$$\mathcal{Y} = P^t \mathcal{X} \tag{7}$$

where $P = [\phi_1 \phi_2 \dots \phi_K]$, and $K < N$. $\mathcal{Y} \in \mathbb{R}^K$ 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 [5] 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 [5]:

$$S_w = \sum_{i=1}^{L} P(\omega_i) \mathcal{E}\{(\mathcal{Y} - M_i)(\mathcal{Y} - M_i)^t | \omega_i\} \tag{8}$$

$$S_b = \sum_{i=1}^{L} P(\omega_i)(M_i - M)(M_i - M)^t \tag{9}$$

where $P(\omega_i)$ is *a priori* probability, ω_i represent the classes, and M_i and M are the means of the classes and the grand mean, respectively. The criterion the FLD

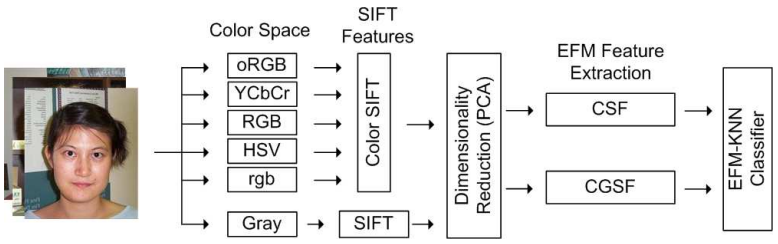

Figure 2 Multiple feature fusion methodology using the EFM.

method optimizes is $J_1 = tr(S_w^{-1} S_b)$, which is maximized when Ψ contains the eigenvectors of the matrix $S_w^{-1} S_b$ [5]:

$$S_w^{-1} S_b \Psi = \Psi \Delta \tag{10}$$

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

$$\mathcal{Z} = \Psi^t \mathcal{Y} \tag{11}$$

$\mathcal{Z}$ thus is more effective than the feature vector $\mathcal{Y}$ 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 [15]. A new method, the Enhanced Fisher Model (EFM), is capable of improving the generalization performance of the FLD method [15]. 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 [15]. The simultaneous diagonalization is stepwise equivalent to two operations as pointed out by Fukunaga [5]: 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 [15], 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) [15].

Image classification is implemented using the EFM-KNN classifier, and Fig. 2 shows the fusion methodology of multiple descriptors using the EFM-KNN classifier.

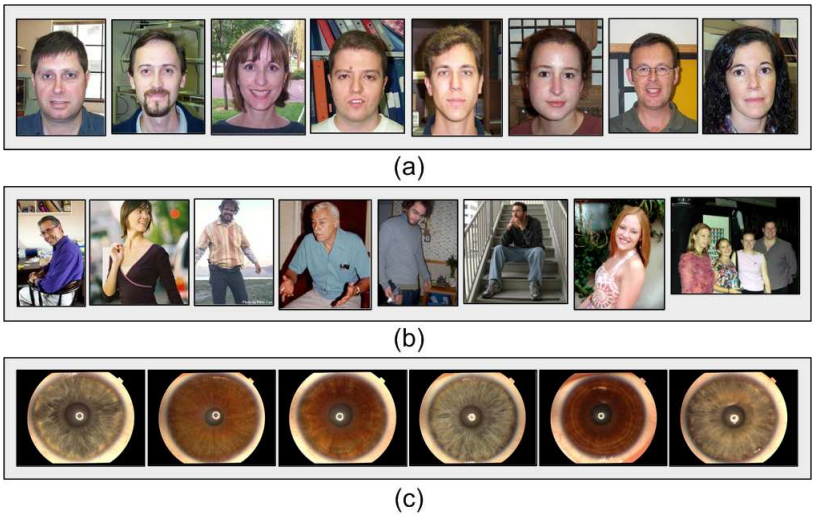

Figure 3 Example images from the following categories: (a) Faces category in the Caltech 256 dataset; (b) People category in the Caltech 256 dataset; (c) Iris category in the UPOL dataset.

5 Experiments

We apply the following two publicly accessible datasets to evaluate our proposed descriptors and classification method: the Caltech 256 object categories [8] and the UPOL iris dataset [21]. The Caltech 256 dataset [8] 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, back-grounds, image sizes, and camera systematics. The various categories represent a wide 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 351 x 351 pixels. See Fig. 3 (a) and (b) for some sample images from the Faces and People categories and Fig. 4 for some images from the object categories. The UPOL iris dataset [21] contains 128 unique eyes (or classes) belonging to 64 subjects with each class containing 3 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 576 x 768 pixels. See Fig. 3 (c) for some sample images from this dataset.

In order for us to make a thorough comparative assessment of our descriptors and methods; from the above two databases we generate the Biometric Dataset with 20 categories that includes the Iris category from the UPOL dataset, Faces and People categories and 17 randomly chosen categories from the Caltech 256 dataset.

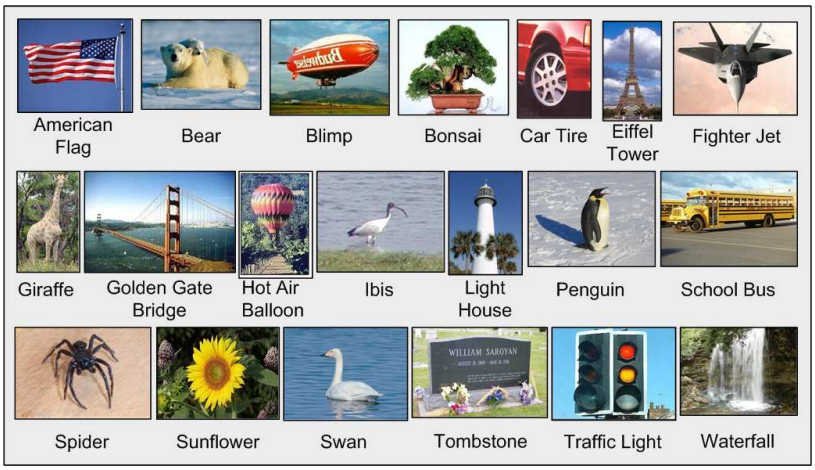

Figure 4 Example images from the Caltech 256 dataset.

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. The dataset is split randomly into two separate sets of images for training and testing. We randomly select from each class 60 images for training and 20 images for testing. 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 eight different descriptors: the oRGB-SIFT, the YCbCr-SIFT, the RGB-SIFT, the HSV-SIFT, the rgb-SIFT, the grayscale-SIFT, the CSF, and the CGSF 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.

The first set of experiments assesses the overall classification performance of the eight descriptors on the Biometric Dataset with 20 categories. Note that for each category we implement five-fold cross validation 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. Fig. 5 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 CSF, and the CGSF descriptors.

The best recognition rate that we obtain is 75.5% from the CGSF, 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.

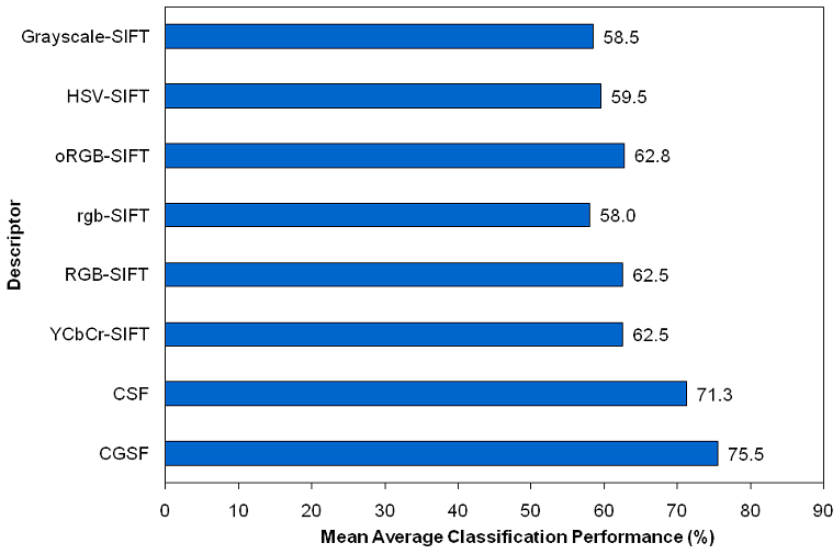

Figure 5 The mean average classification performance of the eight descriptors: the oRGB-SIFT, the YCbCr-SIFT, the RGB-SIFT, the HSV-SIFT, the rgb-SIFT, the grayscale-SIFT, the CSF, and the CGSF descriptors.

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 eight descriptors: CGSF, CSF, YCbCr-SIFT, oRGB-SIFT, RGB-SIFT, HSV-SIFT, Grayscale-SIFT, and rgb-SIFT. We compute classification performance for up to 780 features with the PCA method.

From Fig. 6 it can be seen that the success rate for the CGSF stays consistently above that of the CSF over varying number of features. These two 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 later 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, we compute the success rates for up to 19 features. From Fig. 7 it can be seen that the success rate for the CGSF stays consistently above that of 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 performance over the rgb-SIFT for the varying number of features.

The third set of experiments assesses the eight descriptors using the EFM-KNN classifier on individual image categories. Here we perform a detailed analysis of the performance of the descriptors with the EFM-KNN classifier over all the

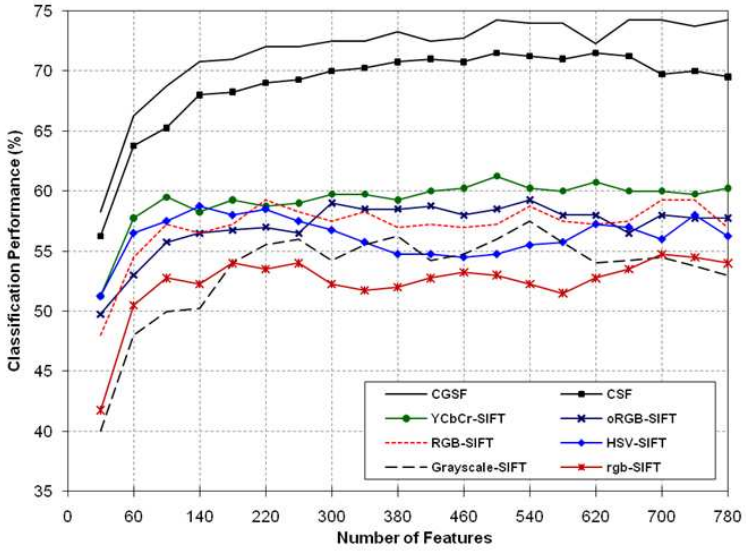

Figure 6 Classification results using the PCA method across the eight descriptors with varying number of features on the Biometric dataset.

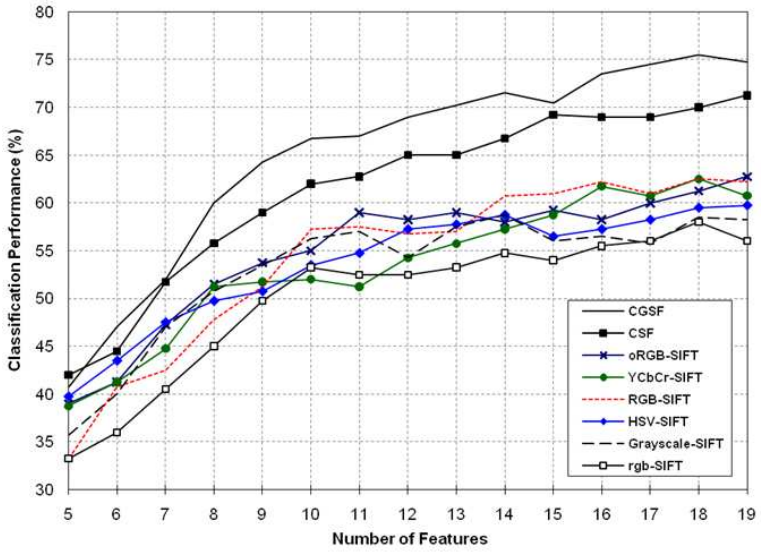

Figure 7 Classification results using the EFM-KNN method across the eight descriptors with varying number of features on the Biometric dataset.

twenty image categories. First we present the classification results on the three biometric categories. Table 1 shows that the Iris category has a 100% recognition rate across all the descriptors. For the Faces category the color SIFT descriptors

Table 1 Category Wise Descriptor Performance (in percentage %) Split-out with the EFM-KNN Classifier on the Biometric Dataset. Note that the categories are sorted on the CGSF results.

Category	CGSF	CSF	oRGB SIFT	YCbCr SIFT	RGB SIFT	HSV SIFT	rgb SIFT	Gray SIFT
iris	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
faces	**100**	**100**	95	90	95	95	95	75
people	**70**	60	40	40	35	40	50	45
cartman	**100**	95	90	**100**	95	85	80	90
grand piano	**100**	95	85	85	70	95	65	90
roulette wheel	**95**	**95**	90	75	85	75	85	75
grapes	90	90	70	**95**	80	70	50	60
waterfall	90	**95**	80	75	85	70	**95**	75
human skeleton	**90**	80	70	60	75	65	65	60
rainbow	**85**	80	55	35	60	65	80	75
laptop	85	80	75	**90**	70	70	60	65
mountain bike	80	80	75	70	80	70	75	**85**
rotary phone	**80**	**80**	60	75	45	70	35	45
cockroach	**75**	70	50	50	60	55	55	55
centipede	**75**	65	55	60	55	55	55	45
owl	**60**	45	40	45	30	25	25	25
buddha	50	40	40	**65**	45	20	40	45
jesus christ	**40**	30	35	10	30	25	20	20
wheelbarrow	**25**	20	**25**	10	**25**	20	10	**25**
snake	20	25	25	20	**30**	20	20	15
Mean	**75.5**	**71.25**	**62.75**	**62.5**	**62.5**	**59.5**	**58**	**58.5**

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 70% with the CGSF, which is a respectable recognition rate when we consider very high intra-class variabilities 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 not redundant for recognition of the People category.

The average success rate for the CGSF descriptor over the top 15 categories is 87.7% with only five categories below the 70% 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.

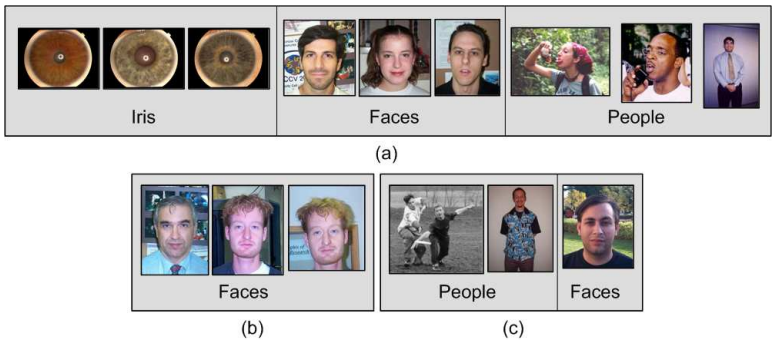

Figure 8 Image recognition using the EFM-KNN classifier: (a) examples of the correctly classified images from the three biometric image categories; (b) images unrecognized using the grayscale-SIFT descriptor but recognized using the oRGB-SIFT descriptor; (c) images unrecognized using the oRGB-SIFT descriptor but recognized using the CSF descriptor.

Figure 9 Image recognition using the EFM-KNN classifier: (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.

The final set of experiments further assesses the performance of the descriptors based on the correctly recognized images. Fig. 8 (a) for some examples of the correctly classified images from the Iris, Faces, and People categories. Notice the high intra-class variabilities for the Faces and People classes. Fig. 8 (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. Fig. 8 (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.

Fig. 9 (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. Fig. 9 (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.

6 Conclusion

We have proposed a new oRGB-SIFT feature descriptor, and then integrated it with other color SIFT features to produce the Color SIFT Fusion (CSF) and the Color Grayscale SIFT Fusion (CGSF) descriptors. Results of the experiments using 20 image categories from two large scale, grand challenge datasets show that our oRGB-SIFT descriptor improves recognition performance upon other color SIFT descriptors, and both the CSF and the CGSF 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.

References

[1] A. Bosch, A. Zisserman, and X. Munoz, "Scene classification using a hybrid generative/discriminative approach," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 30, no. 4, pp. 712–727, 2008.

[2] M. Bratkova, S. Boulos, and P. Shirley, "oRGB: A practical opponent color space for computer graphics," *IEEE Computer Graphics and Applications*, vol. 29, no. 1, pp. 42–55, 2009.

[3] G. Burghouts and J.M. Geusebroek, "Performance evaluation of local color invariants," *Computer Vision and Image Understanding*, vol. 113, pp. 48–62, 2009.

[4] R. Datta, D. Joshi, J. Li, and J. Wang, "Image retrieval: Ideas, influences, and trends of the new age," *ACM Computing Surveys*, vol. 40, no. 2, pp. 509–522, 2008.

[5] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic Press, second edition, 1990.

[6] T. Gevers, J. van de Weijer, and H. Stokman, "Color feature detection: An overview," in *Color Image Processing: Methods and Applications*, R. Lukac and K.N. Plataniotis, Eds. CRC Press, 2006.

[7] C.G. Gonzalez and R.E. Woods, *Digital Image Procession*, Prentice Hall, 2001.

[8] G. Griffin, A. Holub, and P. Perona, "Caltech-256 object category dataset," Tech. Rep., California Institute of Technology, 2007.

[9] C. Liu, "A Bayesian discriminating features method for face detection," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 25, no. 6, pp. 725–740, 2003.

[10] C. Liu, "Enhanced independent component analysis and its application to content based face image retrieval," *IEEE Trans. Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 34, no. 2, pp. 1117–1127, 2004.

[11] C. Liu, "Gabor-based kernel with fractional power polynomial models for face recognition," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 26, no. 5, pp. 572–581, 2004.

[12] C. Liu, "Capitalize on dimensionality increasing techniques for improving face recognition grand challenge performance," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 28, no. 5, pp. 725–737, 2006.

[13] C. Liu, "Learning the uncorrelated, independent, and discriminating color spaces for face recognition," *IEEE Trans. on Information Forensics and Security*, vol. 3, no. 2, pp. 213–222, 2008.

[14] C. Liu and H. Wechsler, "Evolutionary pursuit and its application to face recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 6, pp. 570–582, 2000.

[15] C. Liu and H. Wechsler, "Robust coding schemes for indexing and retrieval from large face databases," *IEEE Trans. on Image Processing*, vol. 9, no. 1, pp. 132–137, 2000.

[16] C. Liu and H. Wechsler, "A shape and texture based enhanced fisher classifier for face recognition," *IEEE Trans. on Image Processing*, vol. 10, no. 4, pp. 598–608, 2001.

[17] C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced Fisher linear discriminant model for face recognition," *IEEE Trans. on Image Processing*, vol. 11, no. 4, pp. 467–476, 2002.

[18] C. Liu and H. Wechsler, "Independent component analysis of gabor features for face recognition," *IEEE Trans. on Neural Networks*, vol. 14, no. 4, pp. 919–928, 2003.

[19] C. Liu and J. Yang, "ICA color space for pattern recognition," *IEEE Trans. on Neural Networks*, vol. 20, no. 2, pp. 248–257, 2009.

[20] D.G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[21] Dobes M. and L. Machala, "Iris database," *http://www.inf.upol.cz/iris/*.

[22] 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*, vol. 65, no. 1-2, pp. 43–72, 2005.

[23] 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*, vol. 19, no. 7, pp. 873–893, 2005.

[24] A. R. Smith, "Color gamut transform pairs," in *Int. Conference on Computer Graphics and Interactive Techniques*, 1978.

[25] H. Stokman and T. Gevers, "Selection and fusion of color models for image feature detection," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 29, no. 3, pp. 371–381, 2007.

[26] M. J. Swain and D. H. Ballard, "Color indexing," *Int. Journal of Computer Vision*, vol. 7, no. 1, pp. 11–32, 1991.

[27] J. Yang and C. Liu, "Color image discriminant models and algorithms for face recognition," *IEEE Transactions on Neural Networks*, vol. 19, no. 12, pp. 2088–2098, 2008.

[28] 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*, vol. 73, no. 2, pp. 213–238, 2007.