Fusion of Color SIFT Features for Image Classification with Applications to Biometrics

Abhishek Verma[1] and Chengjun Liu

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), the Color Grayscale SIFT Fusion (CGSF), and the CGSF+PHOG descriptors for image classification with special applications to biometrics. Classification is implemented using the Enhanced Fisher Model (EFM) classification technique. The effectiveness of the proposed descriptors and classification method are evaluated using 257 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 improves recognition performance upon other color SIFT descriptors; and (ii) the CSF, the CGSF, and the CGSF+PHOG descriptors perform better than the other color SIFT descriptors.*

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

1. INTRODUCTION

Color features contain significant discriminative information for biometric image classification and retrieval [1]-[3]. Another set of application where color based image classification can be very useful is in the identification of object and natural scene categories. The choice of a color space is important for many computer vision algorithms. Different color spaces display different color properties. 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]-[6].

Recently, there has been much emphasis on the detection and recognition of locally affine invariant regions [7]-[9]. 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 [7]-[9]. Affine region detectors when combined with the intensity Scale-Invariant Feature Transform (SIFT) descriptor [8] has been shown to outperform many alternatives [7].

In this paper, we extend the SIFT descriptor to different color spaces, including the recently proposed oRGB color space [10] and propose a new oRGB-SIFT feature representation, and then integrate it with other color SIFT features to produce the Color SIFT Fusion (CSF), the Color Grayscale SIFT Fusion (CGSF), and the CGSF combined with the Pyramid of Histograms of Orientation Gradients (PHOG) to obtain the CGSF+PHOG descriptors for image category classification with special applications to biometrics. Classification is implemented using the Enhanced Fisher Model (EFM) method [11], [12]. The effectiveness of the proposed descriptors and classification method will be evaluated using 257 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 [1], [3], [13] 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 [14]. More recent work on color based image classification appears in [1], [2], [15] that propose several new color spaces and methods for face classification and in [16] the HSV color space is used for the scene category recognition. Evaluation of local color invariant descriptors is performed in [4]. Fusion of color models, color region detection and color edge detection have been investigated for representation of color images [6]. Key contributions in color, texture, and shape abstraction have been discussed in Datta et al. [5].

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], [17], [18]. 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 [11], [19]. Recently, Support Vector Machine (SVM) classifier for multiple category recognition has gained popularity [16], [20] though it suffers from the drawback of being computationally too

[1] corresponding author

A. Verma and C. Liu are with the Department of Computer Science, New Jersey Institute of Technology, University Heights, Newark, NJ 07102 USA (e-mail: {av56, chengjun.liu}@njit.edu).

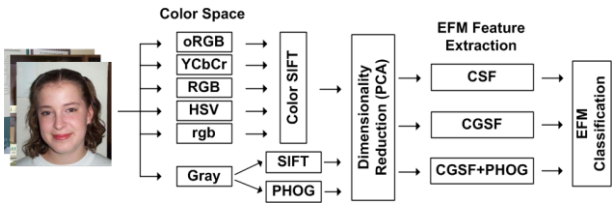

Fig.1 – Multiple feature fusion methodology using the EFM.

expensive on large scale image classification tasks. The EFM classifier has achieved good success for the task of image based recognition [12], [21], [22].

3. NEW COLOR DESCRIPTORS

We first review in this section five color spaces in which our new color 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 four new color SIFT descriptors: the oRGB-SIFT, the Color SIFT Fusion (CSF), the Color Grayscale SIFT Fusion (CGSF), and the CGSF combined with the Pyramid of Histograms of Orientation Gradients (CGSF+PHOG) 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. 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 [23]. 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.

The oRGB color space [10] has three channels L, C_1 and C_2. 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 C_1 and C_2. The value of C_1 lies within [-1, 1] and the value of C_2 lies within [-0.8660, 0.8660]. The L channel contains the luminance information and its values range between [0, 1].

$$\begin{bmatrix} L \\ C_1 \\ C_2 \end{bmatrix} = \begin{bmatrix} 0.2990 & 0.5870 & 0.1140 \\ 0.5000 & 0.5000 & -1.0000 \\ 0.8660 & -0.8660 & 0.0000 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

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 [8]. 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, the four color SIFT descriptors are defined: the RGB-SIFT, the YCbCr-SIFT, the HSV-SIFT, and the rgb-SIFT descriptors.

The Pyramid of Histograms of Orientation Gradients (PHOG) descriptor [24] 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.

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 and the grayscale-SIFT descriptor. The CGSF is combined with the PHOG to obtain the CGSF+PHOG.

4. THE EFM CLASSIFICATION

Image classification using the new descriptors introduced in the preceding section is implemented using

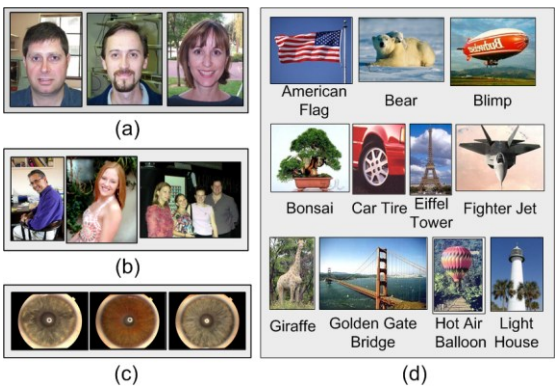

Fig.2 – Example images from the following categories: (a) Faces category in the Caltech 256 dataset; (b) People category in the Caltech 256 dataset; (c) Iris category in the UPOL dataset, and (d) several more example images from the Caltech 256 dataset.

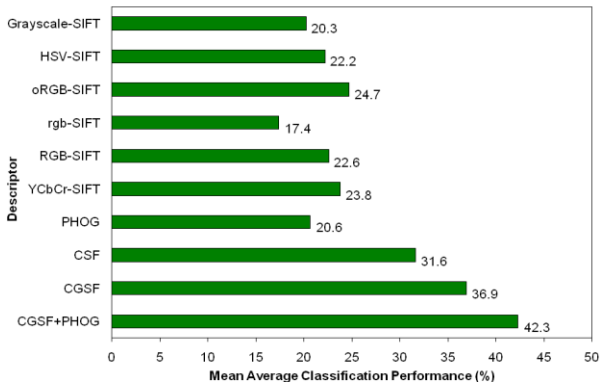

Fig.3 – The mean average classification performance of the ten descriptors.

the Enhanced Fisher Model (EFM) classification method [11], [12]. Let $X \in R^N$ be a random vector whose covariance matrix is Σ_X:

$$\Sigma_X = \varepsilon\{[X - \varepsilon(X)][X - \varepsilon(X)]^t\} \quad (2)$$

where $\varepsilon(.)$ is the expectation operator and t denotes the transpose operation. The eigenvectors of the covariance matrix Σ_X can be derived by PCA:

$$\Sigma_X = \Phi \Lambda \Phi^t \quad (3)$$

where $\Phi = [\phi_1 \phi_2 \phi_3 .. \phi_N]$ is an orthogonal eigenvector matrix and $\Lambda = diag\{\lambda_1, \lambda_2, ..., \lambda_N\}$ a diagonal eigenvalue matrix with diagonal elements in decreasing order. An important application of PCA is dimensionality reduction:

$$Y = P^t X \quad (4)$$

where $P = [\phi_1 \phi_2 \phi_3 .. \phi_K]$, and $K < N$. $X \in R^K$ 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 [25] 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 [25]:

$$S_w = \sum_{i=1}^{L} P(\omega_i) \varepsilon\{(Y - M_i)(Y - M_i)^t \mid \omega_i\} \quad (5)$$

$$S_b = \sum_{i=1}^{L} P(\omega_i)(M_i - M_0)(M_i - M_0)^t \quad (6)$$

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 method optimizes is $J_1 = tr(S_w^{-1} S_b)$, which is maximized when Ψ contains the eigenvectors of the matrix $S_w^{-1} S_b$ [25]:

$$S_w^{-1} S_b \Psi = \Psi \Delta \quad (7)$$

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 Y onto the eigenvectors of Ψ:

$$Z = \Psi^t Y \quad (8)$$

Z thus is more effective than the feature vector 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 [11]. A new method, the Enhanced Fisher Model (EFM), is capable of improving the generalization performance of the FLD method [11]. 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 [11]. The simultaneous diagonalization is stepwise equivalent to two operations as pointed out by Fukunaga [25]: 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 [11], 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 CA space) are not too small (for better generalization performance) [11].

Image classification is implemented using the EFM classification technique, and Fig. 1 shows the fusion methodology of multiple descriptors.

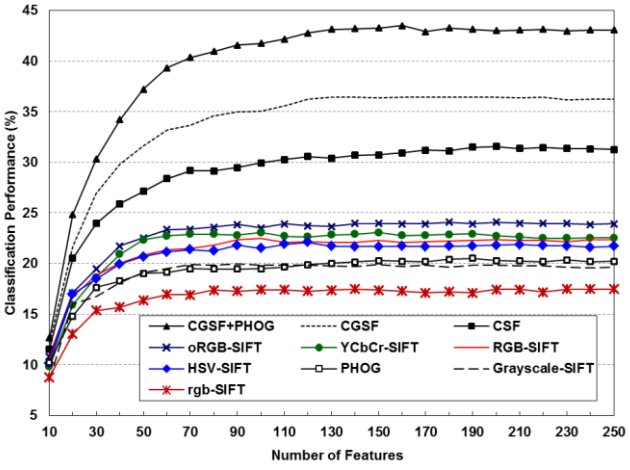

Fig.4 – Classification results using the EFM method across the ten descriptors with varying number of features on the Biometric dataset.

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

Category	CGSF+ PHOG	CGSF	CSF	oRGB SIFT	YCbCr SIFT	RGB SIFT	HSV SIFT	rgb SIFT	Gray SIFT	PHOG
iris	**100**	**100**	**100**	**100**	**100**	**100**	**100**	98	97	95
faces	**97**	**97**	**97**	92	60	92	**97**	85	73	95
people	**17**	14	13	10	8	10	9	7	7	11
car side	**100**	**100**	**100**	93	67	**100**	**100**	**100**	**100**	95
leopards	**100**	98	**100**	70	70	93	97	98	98	97
motorbikes	**98**	92	90	82	77	70	70	50	73	97
sunflower	**97**	**97**	93	88	68	95	92	73	90	53
trilobite	**95**	80	67	62	62	60	58	60	50	83
lawn mower	**93**	80	77	78	77	70	65	38	78	77
american flag	**90**	88	85	60	60	70	65	77	40	5
zebra	**90**	87	67	75	85	27	27	22	30	38
chess board	88	**92**	88	90	80	78	78	73	83	13
tower pisa	88	**92**	85	82	87	77	78	57	77	77
swiss army knife	**87**	82	77	60	60	75	75	70	68	8
airplanes	**85**	70	63	60	62	38	28	28	48	82
saturn	**85**	83	72	83	80	67	70	32	52	53
cereal box	83	**88**	72	73	52	52	57	47	47	17
french horn	83	**85**	78	72	67	68	67	73	70	32
ketch	**83**	62	57	28	30	32	32	28	37	65
pci card	**83**	80	78	68	50	58	57	52	58	10
hibiscus	82	**85**	80	65	72	72	73	67	62	48

5. EXPERIMENTS

We apply the following two publicly accessible datasets to evaluate our proposed descriptors and classification method: the Caltech 256 object categories [26] and the UPOL iris dataset [27]. The Caltech 256 dataset [26] 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, and image sizes. 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. 2(a) and (b) for some sample images from the Faces and People categories and Fig. 2(d) for some images from the object categories. The UPOL iris dataset [27] 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. 2(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 257 categories that includes the Iris category from the UPOL dataset and all categories 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 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 Enhanced Fisher Model (EFM) method.

The first set of experiments assesses the overall classification performance of the ten descriptors on the Biometric Dataset with 257 categories. Note that for each category we implement five-fold cross validation for each descriptor using the EFM classification technique to derive the average classification performance. As a result,

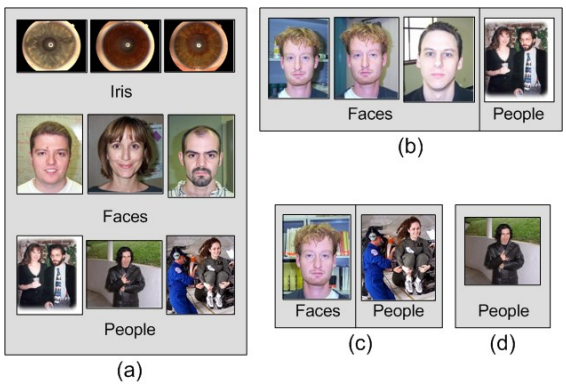

Fig.5 – Image recognition using the EFM classifier: (a) examples of the correctly classified images from the three biometric image categories; (b) images unrecognized using the grayscale-SIFT descriptor but recognized using the oRGB-SIFT descriptor; (c) images unrecognized using the oRGB-SIFT descriptor but recognized using the CSF descriptor; (d) images unrecognized using the CSF but recognized using the CGSF+PHOG.

Table 2. Classification Rate (%) Comparison on the Caltech 256 and the Biometric Datasets

Method	Number of training images per class							
	15		30		45		60	
*SPM [26]	-	-	34.10	± 0.20	-	-	-	-
*KC [28]	-	-	27.17	± 0.46	-	-	-	-
*KSPM [29]	23.34	± 0.42	29.51	± 0.52	-	-	-	-
*LSPM [29]	13.20	± 0.62	15.45	± 0.37	16.37	± 0.47	16.57	± 1.01
*ScSPM [29]	27.73	± 0.51	34.02	± 0.35	37.46	± 0.55	40.14	± 0.91
Our Method								
*CGSF+PHOG (PCA)	27.07	± 0.40	33.75	± 0.30	37.33	± 0.48	39.90	± 0.07
*CGSF+PHOG (EFM)	**29.97**	**± 0.52**	**36.49**	**± 0.51**	**39.74**	**± 0.41**	**42.06**	**± 0.45**
†CGSF+PHOG (PCA)	27.46	± 0.29	**34.10**	**± 0.35**	**37.60**	**± 0.44**	40.13	± 0.11
†CGSF+PHOG (EFM)	**30.39**	**± 0.46**	**36.87**	**± 0.55**	**40.04**	**± 0.43**	**42.28**	**± 0.41**

* Results on the Caltech 256 dataset; † Results on the Biometric 257 categories dataset.

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. Fig. 3 shows the mean average classification performance of various descriptors.

The best recognition rate that we obtain 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.

The second set of experiments evaluates the classification performance using the EFM method respectively by varying the number of features over the descriptors.

From Fig. 4 it can be seen that the success rate for the CGSF+PHOG stays consistently above that of the CGSF and CSF 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 show a similar increasing trend and decline only toward the later half and continue to perform better than the rest of the descriptors. The grayscale-SIFT maintains its higher performance over the rgb-SIFT.

The third set of experiments assesses the ten descriptors using the EFM classifier on individual image categories. Here we perform a detailed analysis of the performance of the descriptors with the EFM classifier over all the 257 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 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 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+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 we combine the PHOG with the CGSF.

The final set of experiments further assesses the performance of the descriptors based on the correctly recognized images. See Fig. 5(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. 5(b) shows some example images from the Faces class that are not recognized by the EFM 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. 5(c) shows images unrecognized using the oRGB-SIFT but recognized using the CSF and Fig. 5(d) shows images unrecognized using the CSF but recognized using the CGSF+PHOG.

See Fig. 6(a) for some examples of the images unrecognized by the EFM using the grayscale-SIFT but

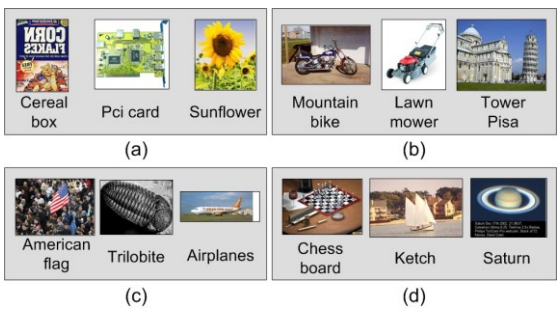

Fig.6 – Image recognition using the EFM classifier: (a) example images unrecognized using the grayscale-SIFT descriptor but recognized using the oRGB-SIFT descriptor; (b) example images unrecognized using the oRGB-SIFT but recognized using the CSF descriptor; (c) images unrecognized using the CSF but recognized using the CGSF+PHOG. (d) Images unrecognized using the PCA but recognized using the EFM on the CGSF+PHOG descriptor.

are correctly recognized by the oRGB-SIFT. Fig. 6(b) shows some images that were previously not recognized by the oRGB-SIFT but are correctly recognized by the CSF. In Fig. 6(c) we show some images unrecognized by the CSF but are correctly recognized by the CGSF+PHOG descriptor. Lastly in Fig. 6(d) images unrecognized by the PCA but recognized by the EFM classifier on the CGSF+PHOG descriptor.

Finally we present the results of our methods on varying number of training images per class over the Caltech 256 dataset and the Biometric dataset with 257 categories. We fix the number of test images per class at 20 for each of the experiments. We combine the Iris images from the UPOL dataset to the Caltech 256 to form the Biometric dataset. From Table 2 on the Biometric dataset it can be seen that on the 15 training images we achieve 30.39% success rate. This improves over the previous best result on the Caltech 256 by 2.66%. For the 30 and 45 training images we improve on the previous best on the Caltech 256 by 2.85% and 2.58% respectively. On the 60 training images we achieve a success rate of 42.28% and improve over the previous best by 2.14%.

From Table 2 on the Caltech 256 dataset it can be seen that on the 15 training images per class we achieve 29.97% success rate. This improves over the previous best result by 2.24%. For the set of 30 and 45 training images we improve on the previous best by 2.39% and 2.28% respectively. On the 60 training images we achieve a success rate of 42.06% and improve over the previous best by 1.92%.

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), the Color Grayscale SIFT Fusion (CGSF), and the CGSF+PHOG descriptors. Results of the experiments using 257 image categories from two large scale, grand challenge datasets show that our 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.

7. REFERENCES

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

[2] 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.

[3] 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.

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

[5] 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.

[6] 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.

[7] 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, nos. 1-2, pp. 43-72, 2005.

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

[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] 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.

[11] 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.

[12] 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.

[13] 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.

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

[15] 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.

[16] 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.

[17] 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.

[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, "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.

[20] 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.

[21] 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.

[22] 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.

[23] C.G. Gonzalez and R.E. Woods, *Digital Image Processing*, Upper Saddle River, NJ: Prentice-Hall, 2001.

[24] A. Bosch, A. Zisserman, and X. Munoz, "Representing shape with a spatial pyramid kernel," *in 2007 Int. Conf. on Image and Video Retrieval*, pp. 401-408.

[25] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, San Diego, CA: Academic Press, 2nd ed., 1990.

[26] G. Griffin, A. Holub, and P. Perona, "Caltech-256 object category dataset," California Institute of Technology, Pasadena, CA, Tech. Rep. CNS-TR-2007-001, Apr. 2007.

[27] M. Dobes, J. Martinek, D. Skoupil, Z. Dobesova, and J. Pospisil, "Human eye localization using the modified hough transform," *Optik*, vol 117, no. 10, pp. 468-473, 2006.

[28] J. Gemert, J-M. Geusebroek, C. Veenman, and A. Smeulders, "Kernel codebooks for scene categorization," *In 2008 European Conf. on Computer Vision*, pp. 696-709.

[29] J. Yang, K. Yu, Y. Gong, and T. Huang, "Linear spatial pyramid matching using sparse coding for image classification," *in 2009 IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 1794-1801.