Gender Recognition from Facial Images using Local Gradient Feature Descriptors

Olarik Surinta
Multi-agent Inelligent Simulation Laboratory (MISL)
Faculty of Informatics, Mahasarakham University
Maha Sarakham, Thailand
olarik.s@msu.ac.th

Thananchai Khamket
Applied Informatics Group, Department of Information Technology, Faculty of Informatics, Mahasarakham University
Maha Sarakahm, Thailand
thananchai.k@msu.ac.th

***Abstract*—Local gradient feature descriptors have been proposed to calculate the invariant feature vector. These local gradient methods are very fast to compute the feature vector and achieved very high recognition accuracy when combined with the support vector machine (SVM) classifier. Hence, they have been proposed to solve many problems in image recognition, such as the human face, object, plant, and animal recognition. In this paper, we propose the use of the Haar-cascade classifier for the face detection and the local gradient feature descriptors combined with the SVM classifier to solve the gender recognition problem. We detected 4,624 face images from the ColorFERET dataset. The face images data used in gender recognition included 2,854 male and 1,770 female images, respectively. We divided the dataset into train and test set using 2-fold and 10-fold cross-validation. First, we experimented on 2-fold cross-validation, the results showed that the histogram of oriented gradient (HOG) descriptor outperforms the scale-invariant feature transform (SIFT) descriptor when combined with the support vector machine (SVM) algorithm. The accuracy of the HOG+SVM and the SIFT+SVM were 96.50% and 95.98%. Second, we experimented on 10-fold cross-validation and the SIFT+SVM showed high performance with an accuracy of 99.20%. We discovered that the SIFT+SVM method needed more training data when creating the model. On the other hand, the HOG+SVM method provided better accuracy when the training data was insufficient.**

Keywords—gender recognition, face detection, local gradient feature descriptor, support vector machine

I. Introduction

Gender recognition can be used to improve the efficiency of surveillance and security systems, authentication systems, and face recognition systems [1]. Moreover, it can also be developed into a variety of applications. Research in gender recognition involves with three major tasks; face detection, feature extraction (called *face encoding*) and recognition system [2]–[6].

Related work. In [7], the deep convolutional neural networks and support vector machines were proposed for gender recognition and tested on the ColorFERET dataset. The pre-processing step consists of detecting and cropping the face image. The face images after the detection stage consisted of 8,364 face images and stored at 256x256 pixel resolution. After that, the data augmentation technique is implemented to generate new face images. A pre-trained model of the AlexNet architecture was used to train the face images. The linear support vector machine is attached to the last fully connected layer. Using this method, the best accuracy was 97.3%.

In [3] a local feature descriptor called pyramid histogram of oriented gradients (PHOG) was proposed to represent a local gradient of the image. For the HOG descriptor, the feature vector is calculated according to Equation (1). Additionally, the PHOG descriptor allows dividing an image into a small block at several pyramid levels [8]. The gradient orientations in every level are stored into orientation bins. Then, all of the orientation bins in each pyramid levels are combined. The feature vector is then classified using the SVM classifier with the RBF kernel. The proposed method achieved an accuracy of 88.5% on the labeled faces in the wild (LFW) database.

Also, in [5] proposed multiscale facial fusion feature; however, the multiscale method is related to the pyramid technique [3], [8]. The fusion features used in the experimented, including local phase quantization (LPQ) and local binary pattern (LBP) descriptors. The combination of the feature vector is extracted from two descriptor methods and sent to the SVM classifier to classify the face image. The multiscale facial fusion method obtained an accuracy of 86.11% on the images of groups (IoG) dataset.

In this paper, we first applied a well-known Haar-cascade classifier, which was invented by Viola and Jones [9], [10] that proposed for object and pedestrian detection, to first find the exact location of a face from the complete image. Note that we focus only on the frontal face, and we ignore the profile face if the head of the people is turned to left or to right. Due to the challenge of the ColorFERET dataset [11], [12], we can extract only 4,624 face images from the 11,119 images. After that, all face images were resized to the same size. The face image resolution used in the experiments was 88x80 pixels.

Secondly, two local gradient feature descriptors called the histogram of oriented gradients (HOG) [13] and the scale-invariant feature transform (SIFT) [14] descriptors are proposed to extract the gradient feature from the face image. We experimented with the performance of the local gradient descriptors using several parameters. We set up the parameters of the HOG descriptor; orientations, pixels per cell, and cell per block and the SIFT descriptor; patch size.

Finally, the support vector machine (SVM) [15] with the radial basis function (RBF) kernel is proposed to create a model of the gender feature vector from the training data. We implemented the grid-search method to discover the hyper-parameters (C and γ) [16] until obtaining the best optimize parameters were obtained. Also, the average accuracies and the standard deviation were used to compare the experimental results.

Contributions. In this paper, we proposed two well-known local gradient feature descriptors; the HOG and SIFT descriptors, to compute the invariant feature vector from the face images. These local gradient feature descriptors are designed to extract features from the gradient image for object detection purposes. The feature descriptor combined with the SVM with the RBF kernel is presented to address the gender recognition problems. The results show that our proposed method achieves very high recognition accuracy.

Paper Outline. The rest of the paper is presented as follows: In Section II, the gender recognition method, which is proposed is explained. In Section III, experimental settings and the results are presented. The conclusion and future work are given in Section IV.

II. Gender Recognition Methods

A. Face Detection

For face detection, the Haar-cascade classifier was proposed by Viola and Jones [10] in 2004. This method, the Haar features were used to compute the feature vector. The sub-window scans within the image to capture the small image. Then send the data of the sub-window to calculate the feature vector. Consequently, the feature vector was trained and predicted with the AdaBoost algorithm.

B. Local Gradient Feature Descriptors

To study the effectiveness of local gradient feature descriptors for gender recognition, we compare two well-known gradient features, called *the histogram of oriented gradients* and *the scale-invariant feature transform*. In this study, the face images are resized to 88x80 pixel resolutions.

1) Histogram of Oriented Gradients (HOG): The HOG descriptor was invented by Dalal and Triggs [13] in 2005 for detecting a pedestrian in an image. The basis of this technique is to compute the gradient orientation from small connected regions of an image. The features that calculated from this technique are robust to the light and geometric changes [17].

The notation of the HOG method [18] can be written as follows:

$$\Phi_f(\mathrm{X}) = Db * \left[(g_f * \mathrm{X}) \odot (g_f * \mathrm{X})\right] \tag{1}$$

where X is an input image and $\mathrm{X} \in \mathbb{R}^D$.

First, X is convolved with the simple convolution kernel g_f in the horizontal and vertical directions.

Second, blurred with b and the nonlinear transform ($\odot$) is applied to removes sensitivity to edge contrast and increases edge bandwidth. Third, the gradient orientations are weighted and stored into orientation bins Db.

Finally, The histograms from each block are describes as the feature descriptor. Then, the L2-Normalization is used to normalize the feature descriptors. The equation of L2-Normalization [16] can be written as follows:

$$V'_k = \frac{V_k}{\sqrt{\|V_k\|^2 + \varepsilon}} \tag{2}$$

where

V_k is the histograms from all block regions
ε is a very small value and close to zero
V'_k is the normalized feature descriptor.

2) Scale-Invariant Feature Transform (SIFT): The SIFT method was proposed by Lowe [14] in 2004 for extracting invariant features from images. The features are invariant to image scale and rotation. The complete process of the SIFT method consists of scale-space extrema detection, keypoint localization, orientation assignment, and the local image descriptor. The complete SIFT method is applied to localization of the object in the target image.

3) In this paper, we focus only on computing the invariant feature, called *the SIFT descriptor*. First, the Gaussian kernel is used to convolution the image, I.

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \tag{3}$$

where

$I(x, y)$ is the pixel at location x, y of image I
$G(x, y, \sigma)$ is the Gaussian kernel, and σ determines the width of the Gaussian kernel.

Second, the gradient orientation $\theta(x, y)$ and magnitude $m(x, y)$ are computed from the image, $L(x, y)$:

$$G_x = L(x+1, y, \sigma) - L(x-1, y, \sigma) \tag{4}$$
$$G_y = L(x, y+1, \sigma) - L(x, y-1, \sigma)$$

where G_x is the horizontal and G_y is the vertical components of the gradients.

Third, a sliding window method is used to slide through the whole image to capture a small region. Then all regions are sent to the SIFT descriptor to extract the gradient orientations and magnitudes.

Finally, each region is divided into 4x4 equal blocks. Then an orientation histogram is created for each block. Each histogram uses 8 bins to store the orientation values, which results in 128 dimensions for each region.

C. Classifier Method

The *support vector machine (SVM)* algorithm is a supervised learning algorithm employed for recognizing the feature that is extracted from data. The SVM algorithm, invented by Vapnik [15], it has been successfully applied to many pattern recognition problems [16], [19] such as image classification, handwritten recognition, face detection, and face recognition.

The SVM algorithm is first created for binary classification problems [20]. This technique finds the function $g(\cdot)$ that is the best separation to the pattern data, called *hyperplane*.

The training set is (x_i, y_i) where $x_i \in \mathbb{R}^n$ and the output label are either +1 or -1, $y_i \in \{+1, -1\}$. It can be split by the following:

$$g(\mathbf{x}) = \boldsymbol{w}^T \cdot \mathbf{x} + b \tag{5}$$

where $\mathbf{x}$ is the weight vector and b is the bias value.

The largest distance between the nearest positives $\boldsymbol{w}^T \cdot \mathbf{x} + b = +1$ and negatives $\boldsymbol{w}^T \cdot \mathbf{x} + b = -1$ is the optimal separating hyperplane.

Moreover, the SVM can be extended to deal with non-linear data. Then, the soft constraint is proposed:

$$y_i(\boldsymbol{w}^T \cdot \mathbf{x} + b) \geq 1 - \xi_i \tag{6}$$

where ξ_i is the slack variable for data $\mathbf{x}_i$

The radial basis function (RBF) kernel is a non-linear similarity function and employed in the SVM classifier. In this kernel, the similarity value between the two input vectors are computed as follows:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2) \tag{7}$$

where γ is the parameter of the RBF kernel, note that the model can be overfitting when the γ parameter is too large because it increases the number of support vectors.

III. Experimental Settings and Results

In this section, we concisely describe the face image dataset used in the experiments. The experimental results consisting of the face detection results, parameter settings, grid search parameter estimation, and gender recognition results, are presented and discussed.

A. Face Image Dataset

In the experiments, we used a benchmark face image dataset, called the color face recognition technology (ColorFERET) dataset. Firstly, we used ColorFERET for face detection purpose. Secondly, we divided the ColorFERET dataset into training and test sets using 2-fold (50:50) and 10-fold (90:10) cross-validation, respectively, for gender recognition.

The ColorFERET was introduced by J. Phillips and P. Rauss [11], [12] for a facial recognition system. This dataset consists of 14,126 face images from 1,199 subjects. The resolution of images in the dataset is 384x256 pixels. An example from the ColorFERET dataset is shown in Fig. 1.

B. Experimental Results

1) Face Detection Result:

The evaluation method of the face detection is given by:

$$Acc_{fd} = Ac_{fd} - Er_{fd} \tag{5}$$

when

$$Ac_{fd} = \frac{c*100}{N}, \qquad Er_{fd} = \frac{e*100}{N} \tag{6}$$

where

c is the number of face images, after using a face detection technique.
e is the number of the error face images
N is the total number of the face images in the dataset.

In this paper, we proposed to use the Haar-cascade classifier for the face detection process.

Firstly, we used the Faced face detection method to detect face from the ColorFERET dataset. The result showed that the Haar-cascade classifier significantly outperformed the Faced method. The Faced method was a time-consuming when compared to the Haar-cascade classifier method.

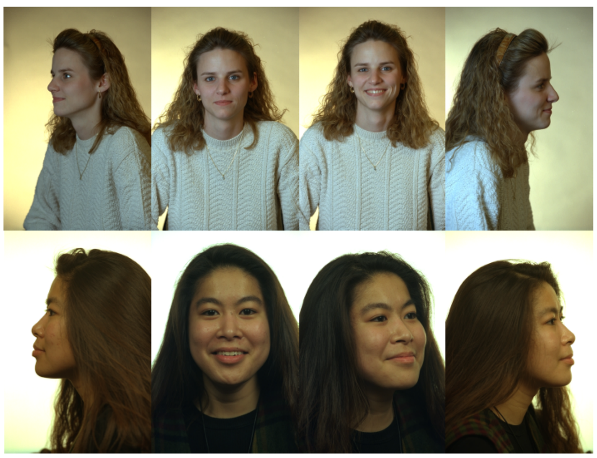
Secondly, we experimented with face detection using the Haar-cascade classifier. In the ColorFERET dataset, there are 13 different poses of each person, such as regular frontal image, profile left, half left, quarter left and also head turned to left and right between 15-75 degree as shown in Fig. 1. This method obtained an accuracy of 39.25% on ColorFERET dataset. The accuracy of the male and female faces was 36.87% and 41.63%, respectively.

Based on our experiments, the Haar-cascade classifier performed not very well when the face images turn to left or right, so the detection rate was quite low. On the other hand, this detection technique performed quite very well and fast when the regular frontal image were used. The results of the Haar-cascade face detection are shown in Table I.

TABLE I. Performance of the Face Detection using Haar-cascade Classifier on the ColorFERET Dataset

Gender	Number of male images	Number of face detected	Number of error detected
Male	7,139	2,854	222
Female	3,980	1,770	113

(a)



(b)

Fig. 1. Example face images of (a) male and (b) female from the ColorFERET dataset.

1) *Parameter Settings:*

We evaluated the performance of the HOG and SIFT descriptors using several parameters. The parameters of the HOG descriptor included orientations, pixels per cell, and cell per block [17]. The parameter of the SIFT descriptor comprised only patch size.

Note that the pixels per cell and cells per block of the HOG parameters and the patch size of the SIFT parameter are defined as a square. We use the SVM with the RBF kernel as a classifier and using the default C,γ parameters to find the best parameters of the HOG and SIFT descriptors. Also, 10-fold cross-validation over the training set was applied.

The best parameters of the HOG descriptor uses as 9 orientations, 8 pixels per cell, and 3 cells per block. The SIFT descriptor used patch size = 25 pixels. The accuracy results of the HOG and SIFT descriptors are shown in Table II and III.

(a)

(b)

Fig. 2. Sample face images of (a) male and (b) female after applying Haar-Cascade classifier from the ColorFERET dataset.

TABLE II. THE PERFORMANCE OF DIFFERENT HOG DESCRIPTOR PARAMETERS

HOG Descriptor Parameters			
Orientations	**Pixels per cell**	**Cells per block**	**Accuracy (%)**
4	8	1	94.6
8	16	1	92.2
8	16	2	92.8
9	8	1	94.8
9	**8**	**3**	**95.8**
9	16	1	93.3
24	16	1	92.0

2) *Grid Search Parameter Estimation:*

From the parameter settings section, the best feature descriptor parameter values were selected. Consequently, we have optimized the hyper-parameters of the SVM classifier with the RBF kernel. The grid search parameter method is suggested. We searched the hyper-parameter C and gamma between the number of 2^{-7} and 2^{7}. The best hyper-parameters found for our experiments are shown in Table IV.

TABLE III. THE PERFORMANCE OF THE SIFT DESCRIPTOR USING DIFFERENCE PATCH SIZES

SIFT Descriptor Parameters	
Patch sizes	**Accuracy (%)**
10	97.8
20	98.2
25	**98.4**
30	97.1
40	97.1
45	97.8
50	96.9

TABLE IV. THE BEST HYPER-PARAMETER VALUES FOR THE SVM CLASSIFIER WITH THE RBF KERNEL

Methods	C	γ
HOG	2^3	2^0
SIFT	2^3	2^{-5}

3) Gender Recognition Results:

The calculation of gender recognition accuracy is computed by multiply 100, with the total number of correct prediction and divided by the total number of face images in the dataset.

From the face detection result, we divided 4,624 face images into train and test sets with the ratio of 50:50 (2-cv) and 90:10 (10-cv). On this face images dataset, 2-fold cross-validation achieved accuracy of 96.50% when using the HOG descriptor. On the other hand, when performing the system with 10-fold cross-validation, the SIFT descriptor outperforms the HOG method with the accuracy of 99.20, which is the highest result based on our experiments. The accuracy results of gender recognition are shown in Table V.

IV. CONCLUSION

The main objective of this paper is to recognize gender (male and female) from facial images. First, the Haar-cascade Classifier was used to find the face from the whole image. Second, the face images were then assigned to the local gradient feature descriptors; the histogram of oriented gradients (HOG) and scale-invariant feature transform (SIFT) descriptors, to compute the feature vector.

TABLE V. THE ACCURACY (%) OF THE SVM CLASSIFIER OBTAINED WITH 2-FOLD AND 10-FOLD CROSS-VALIDATIONS

Methods	Accuracy (%)	
	2-cv	**10-cv**
HOG	96.50 ± 1.8	98.75 ± 2.5
SIFT	95.98 ± 0.4	$\mathbf{99.20 \pm 0.8}$

Finally, for gender recognition, finally, the invariant feature vector was classified using the support vector machine (SVM) with the radial basis function (RBF) kernel. From the experimental results, the SIFT descriptor outperformed the HOG descriptor when combined with SVM with RBF kernel. This method obtained very high recognition accuracy.

In future work, we plan to work on the deep convolutional neural network to detect the face (even from different poses) and compute the invariant feature vector. We also want to study the effect of the data augmentation to generate the new face images.

ACKNOWLEDGMENT

This research was supported by the Faculty of Informatics, Mahasarakham University, Thailand.

REFERENCES

[1] S. C. Nistor, A.-C. Marina, A. S. Darabant, and D. Borza, "Automatic gender recognition for 'in the wild' facial images using convolutional neural networks," in *Intelligent Computer Communication and Processing (ICCP), 13th IEEE International Conference on*, 2017, pp. 287–291.

[2] S. Yousefi and M. Zahedi, "Gender recognition based on SIFT features," *Int. J. Artif. Intell. Appl.*, vol. 2, no. 3, pp. 87–94, 2011.

[3] O. A. Arigbabu, S. M. S. Ahmad, W. A. W. Adnan, S. Yussof, and S. Mahmood, "Soft biometrics: Gender recognition from unconstrained face images using local feature descriptor," *J. Inf. Commun. Technol.*, vol. 14, no. 1, pp. 111–122, 2015.

[4] G. Azzopardi, A. Greco, and M. Vento, "Gender recognition from face images with trainable COSFIRE filters," *2016 13th IEEE Int. Conf. Adv. Video Signal Based Surveillance, AVSS 2016*, no. November 2017, pp. 235–241, 2016.

[5] C. Zhang, H. Ding, Y. Shang, Z. Shao, and X. Fu, "Gender classification based on multiscale facial fusion feature," *Math. Probl. Eng.*, vol. 2018, pp. 1–6, 2018.

[6] Z. Stawska, "Support Vector Machine in Gender Recognition," *Inf. Syst. Manag.*, vol. 6, no. 4, pp. 318–329, 2017.

[7] J. van de Wolfshaar, M. F. Karaaba, and M. A. Wiering, "Deep Convolutional Neural Networks and Support Vector Machines for Gender Recognition Deep Convolutional Neural Networks and Support Vector Machines for Gender Recognition," in *IEEE Symposium Series on Computational Intelligence*, 2015, pp. 188–195.

[8] P. P. Saragni, B. S. P. Mishra, and S. Dehuri, "Pyramid histogram of oriented gradients based human ear identification," *Int. J. Control Theory Appl.*, vol. 10, no. 15, pp. 125–133, 2017.

[9] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Computer Vision and Pattern Recognition (CVPR), IEEE Computer Society Conference on*, 2001, vol. 1, pp. 511–518.

[10] P. Viola and M. Jones, "Robust real-time face detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, 2004.

[11] P. J. Phillips, H. Wechsler, J. Huang, and P. J. Rauss, "The FERET database and evaluation procedure for face-

recognition algorithms," *Image Vis. Comput.*, vol. 16, no. 5, pp. 295–306, 1998.

[12] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 10, pp. 1090–1104, 2000.

[13] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition (CVPR), IEEE Computer Society Conference on*, 2005, pp. 886–893.

[14] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[15] V. N. Vapnik, *Statistical Learning Theory*. Wiley, 1998.

[16] O. Surinta, M. Karaaba, L. Schomaker, and M. Wiering, "Recognition of handwritten characters using local gradient feature descriptors," *Eng. Appl. Artif. Intell.*, vol. 45, pp. 405–414, 2015.

[17] X. Li and Z. Lin, "Face Recognition Based on HOG and Fast PCA Algorithm," in *Intelligent Data Analysis and Applications, the Fourth Euro-China Conference on*, 2017, vol. 682, pp. 10–21.

[18] H. Bristow and S. Lucey, "Why do linear SVMs trained on HOG features perform so well?," 2014.

[19] V. Codreanu *et al.*, "Evaluating automatically parallelized versions of the support vector machine," *Concurr. Comput. Pract. Exp.*, vol. 28, no. 7, pp. 2274–2294, 2014.

[20] O. Surinta, L. R. B. Schomaker, and M. A. Wiering, "A Comparison of Feature Extraction and Pixel-Based Methods for Recognizing Handwritten {B}angla Digits," in *Document Analysis and Recognition (ICDAR), The 12th International Conference on*, 2013, pp. 165–169.