GENDER IDENTIFICATION IN UNCONSTRAINED SCENARIOS USING SELF-SIMILARITY OF GRADIENTS FEATURES

Hong Liu, Yuan Gao, Can Wang

Key Laboratory of Machine Perception
Shenzhen Graduate School, Peking University, China
E-mail: hongliu@pku.edu.cn, ygao@sz.pku.edu.cn, canwang@pku.edu.cn

ABSTRACT

Gender identification has been a hot research topic with wide application requirements from social life. In general, effective feature representation is the key to solving this problem. In this paper, a new feature named Self-Similarity of Gradients (GSS) is proposed, which captures pairwise statistics of localized gradient distributions. There are three contributions made by us to practical gender identification. First, GSS features are proposed for gender identification in the wild, which achieve good performance compared with baseline approaches. Second, we originally utilize 31-dimensional HOG for practical gender identification and its excellent results demonstrates that HOG with both contrast sensitive and insensitive information is a better fit for this topic than that with only contrast insensitive information. Last, feature combination and multi-classifier combination strategies are adopted and the best gender identification performance is achieved. Experimental results show that the combination of GSS, HOG and LBP using a linear SVM outperforms state-of-the-art on the LFW database, which meets the "wild" condition.

Index Terms— Histogram of Oriented Gradients, Self-Similarity of Gradients, AdaBoost, SVM, Labeled Faces in the Wild

1. INTRODUCTION

Gender identity is an important attribute for human living in social life, the identification of which has drawn a lot of attentions from the research area [1]. Several feature extraction methods have been utilized for solving this problem and raw pixels related ones are the most easy ways. Moghaddam and Yang utilized raw image pixels on the down-sampled images (12×21 pixels) from the FERET database with SVMs and achieved the accuracy of 96.6% [2]. Baluja and Rowley proposed a fast gender classification method by boosting only 500 pixel comparisons, which matched the performance of SVM [3]. Apart from pixel-based features, LBP has also been very popular for this task. Lian and Lu used LBP as a feature extracting method and SVM as a classifier, and their experiments on the CASPEAL database achieved the correct classification rate of 96.75% [4]. HOG is a popular shape feature descriptor, which is firstly designed for the pedestrian detection problem [5]. Guo et al. used HOG features for gender classification in consideration of the fact that HOG can characterize the shape and texture changes which are prominent and prevalent on faces [6]. In addtion, other common feature extraction methods, such as haar-like [7], SIFT [8], Gabor [9] etc., have been attempted for gender identification.

Studying gender identification in unconstrained scenarios is becoming a research hotspot. Shakhnarovich et al. collected 3,500 face pictures from the web under 30° frontal orientation, and Haar-like features using Adaboost achieved the performance of 79.0% [7]. Gao and Ai achieved gender classification accuracy of 95.51% using the probabilistic boosting tree with simple haar-like features on about 10,100 Mongoloid faces in real environment [10]. In the same year, Aghajanian et al. proposed a general Bayesian framework for within-object classification and achieved an 89% correct gender classification on 1,000 images (500 male, 500 female) from the web [11]. In 2012, Shan used Adaboost to learn the discriminative LBP features on part face images (7,443 pieces) from the Labeled Faces in the Wild (LFW [12]) database and his reported experimental results showed that the selected LBPH bins have strong gender separating capacity [13]. Since LFW is the only public available face database of the above four ([7, 10, 11, 13]) whose contents come from the web and meets the "wild" condition, it is utilized by us in our experiments. We reselect 5,660 images from LFW with each image has a near frontal face corresponding to a unique person.

Effective feature extraction is a vital step for gender identification. Recently, self-similarity on color channels (CSS) features have achieved improved performance on pedestrian detection tasks [14]. By visualizing the 31-dimensional HOG [15] of face images, we also observed some similarities be-

This work is supported by National Natural Science Foundation of China (NSFC, No.60875050, 60675025, 61340046), National High Technology Research and Development Program of China (863 Program, No.2006AA04Z247), Scientific and Technical Innovation Commission of Shenzhen Municipality (No.201005280682A, No.JCYJ20120614152234873, CXC201104210010A).).

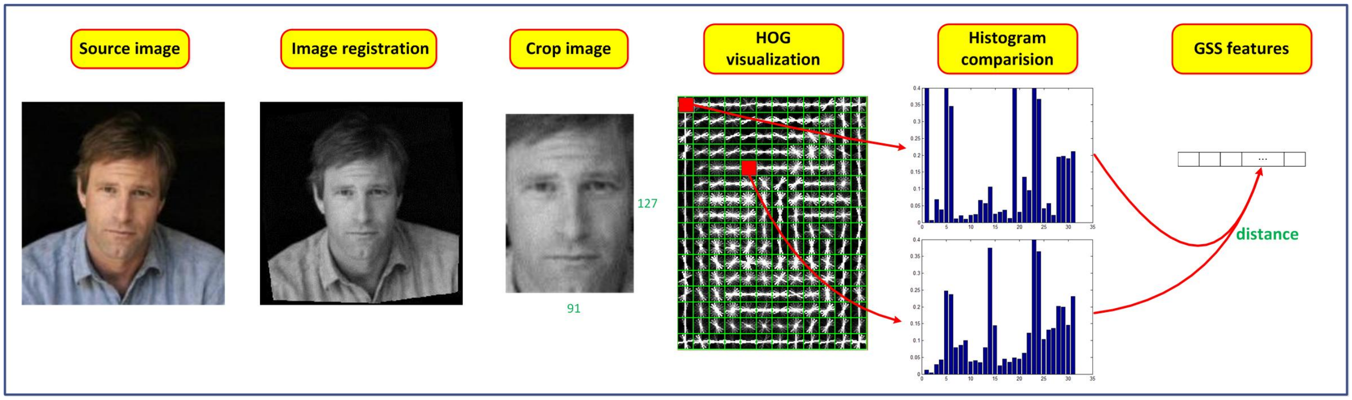

Fig. 1. Flowchart of extracting the GSS feature.

tween them. Inspired by this, a new feature, Self-Similarity of Gradients (GSS), is proposed, the performance of which on LFW is comparable to the raw pixels baseline. In addition, in baseline approaches, HOG31 performs the best, indicating that it is an efficient feature descriptor for gender identification in unconstrained scenarios. Through utilizing AdaBoost to select best features in a combined feature vector (GSS+HOG+LBP) and training a linear SVM classifier on them, the state-of-the-art performance can be achieved, which also implies the effectiveness of the proposed method.

2. FEATURE REPRESENTATION

Because our proposed GSS features are inspired by CSS [14] and, at the same time, based on analysis of HOG visualization, this section describes HOG visualization and representation of GSS sequently.

2.1. HOG Visualization

Visualizing features can help researchers gain a better understanding of the behaviours of detectors. In our gender identification work, two feature visualizing methods for HOG features are tried and illustrated in Fig. 2. In the first column, the two gray images stand for the "mean" faces of male and female images. HOG glyphs of (a)(b) are shown in (c)(d). And (e)(f) are results of using the HOG visualizing method of Vondrick et al. [16] corresponding to (a)(b).

When observing the visualized pictures (c)(d)(e)(f) in Fig. 2, we find that the main differences between male and female "mean" faces are distributed in the mouth, cheeks, eyes and the bridge of nose regions. In addition, the mouth region and eyes region in male "mean" face have a great similarity. However, in female "mean" face, the observation is not totally the same as previous. It is obvious that the center of the mouth is different from the other parts of the mouth but more like cheeks. Intuitively, these pairwise statistics of localized gradient distributions may contribute to a gender detector with good performance. Therefore, we encode self-similarities between cells of a HOG feature map and the detailed GSS feature is introduced in details here:

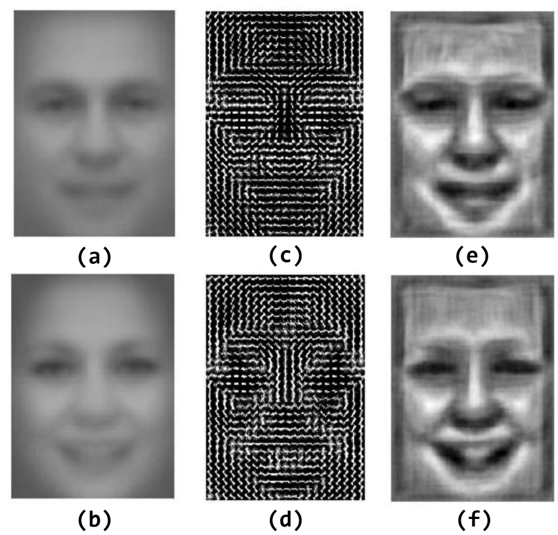

Fig. 2. The "mean" faces of males (top-left) and females (bottom-left) in the GFW database with their HOG visualization images by two different methods.

2.2. Self-Similarity of Gradients Features

The GSS feature is founded on the cell-based HOG feature map, so we define the size of a HOG feature map as $M \times N \times K$. Here, $M \times N$ stands for how many cells in a HOG feature map and K is the HOG dimensions. When computing the GSS feature, for each chosen cell sequently from the cell-based HOG feature map we calculate its distance to all the other cells in the feature map including itself. Therefore, how many calculations directly decide how many components in a GSS feature vector. If GSS is a D_{GSS}-dimensional vector, the D_{GSS} can be computed as:

$$D_{GSS} = A^2_{M\times N} = (M \times N)^2 \quad (1)$$

To obtain the distance between different cells, several functions for comparing histograms are tested and the correlation comparison method is eventually used:

$$d(H_1, H_2) = \frac{\sum_I (H_1(I) - \bar{H}_1)(H_2(I) - \bar{H}_2)}{\sqrt{\sum_I (H_1(I) - \bar{H}_1)^2 \sum_I (H_2(I) - \bar{H}_2)^2}}$$
$$where\ \bar{H}_k = \frac{1}{N} \sum_J H_k(J) \tag{2}$$

The flowchart in Fig. 1 provide the visual process of extracting GSS features from a source image.

3. CLASSIFIERS

Two classes of popular machine learning algorithms, SVM and AdaBoost, have been utilized in our work.

- For the first classifier, a linear SVM [17] is selected in consideration of its good performance, simplicity and, last but not least, the speed.
- For the second classifier, Gentle AdaBoost [18] is also chosen for that it has been the most practically efficient boosting algorithm.
- Apart of using these two algorithms individually, the combination of AdaBoost and SVM is also an effective way to improve classification performance. In this case, AdaBoost is firstly applied to only select features and then a SVM classifier is trained on the selected features.

These two algorithm implementations refer to LIBLINEAR[1] and GML AdaBoost Matlab Toolbox[2].

4. EXPERIMENTS AND DISCUSSIONS

Experiments here are composed of three steps. The first step briefly introduced which images were used in our experiments from LFW. The second is experimental settings which involve three parameters: baseline features, our proposed GSS feature and cross validation parameters. The last step is experiment design with results and exhaustive analysis.

4.1. Selection Strategy of LFW Database

LFW is a face database in unconstrained scenarios which is firstly designed for solving the problem of face recognition and exactly meets the requirement here. LFW has totally 5,749 different persons with 13,233 images from the Web. And there are 1,680 people have two or more images in this database.

Image registration is a usual process step for image classification problems. Fig. 1 exhibits the image registration process using a commercial face alignment software [19]. After this, a fixed 127 $\times$ 91 pixels size window in the center of each image is cropped for use. So here, for the 1,680 people with two or more images we manually select a well normalized picture for each one. For the rest 4,069 person who have only one image, we also choose these with good image registration. And all the selected pictures are manually classified into two different categories (male and female) by two students in our lab. At last, there are total 5,660 images (4,197 males and 1,463 females) and each image corresponds to a unique person. The complete categorized images used in our experiments can be accessed through our web site[3].

4.2. Experimental Settings

As baseline features to compare against, raw pixels and standard LBP are utilized in our experiments. For computing the raw pixel feature, each face image of 127 $\times$ 91 pixels is down-sampled to 64 $\times$ 46 pixels. Therefore the vector length of the raw pixel feature for a image is 2,944. When calculating the LBP feature, all the face images are firstly divided into 42 subregions uniformly (7 rows, 6 columns). Then the uniform LBP operator, LBP(8,2,u2) [20], is used to extract LBP features from all these subregions. At last, the standard LBP feature for a image is a histogram having 2,478 (42$\times$59) bins, which represents the concatenation of the LBP features of all 42 subregions.

HOG is a very popular feature in object detection tasks. There two types of HOG features used here, which are the original 36-dimensional HOG [5, 21] and Felzenszwalb's 31-dimensional HOG [15]. The spatial bin size is set to 8$\times$8 for both. The 36-dimensional HOG implementation refer to Piotr's Image & Video Matlab Toolbox[4] and the calculation of 31-dimensional HOG utilizes DPM code[5]. Here, we use HOG36 and HOG31 to represent them respectively. It is easy to get a feature map of size of 17 $\times$ 13 $\times$ 36 for HOG36 and 16 $\times$ 11 $\times$ 31 for HOG31. Next, we transform a feature map to a feature vector through concatenating elements in a feature map one by one sequently. The final dimension of HOG36 is 7,956 and for HOG31 it is 5,456.

For computational simplicity, GSS is extracted on the down-scaled image with size of 64 $\times$ 48 pixels. M and N are set to 8 and 6 respectively. According to Eq. (1)(2), a 2,304-dimensional GSS feature vector can be calculated.

At last, five-fold cross validation is adopted and can be briefly described below. All selected images are divided into five heaps with the same ratio between male and female. Each time select one distinct heap for testing and use the other four heaps for training. This procedure is then repeated four times.

4.3. Results and Analysis

Table I shows results of using above five features with AdaBoost and SVM. It can observed that HOG $>$ Standard LBP $>$ Raw pixels $\approx$ GSS. This shows that HOG features behave excellent in the gender identification problem. Besides, HOG31 performs the best and outperforms the second place

[1] http://www.csie.ntu.edu.tw/∼cjlin/liblinear/

[2] http://graphics.cs.msu.ru/ru/science/research/machinelearning/adaboosttoolbox/

[3] https://github.com/ygaopku

[4] http://vision.ucsd.edu/∼pdollar/toolbox/doc/

[5] http://cs.brown.edu/∼pff/latent-release4/

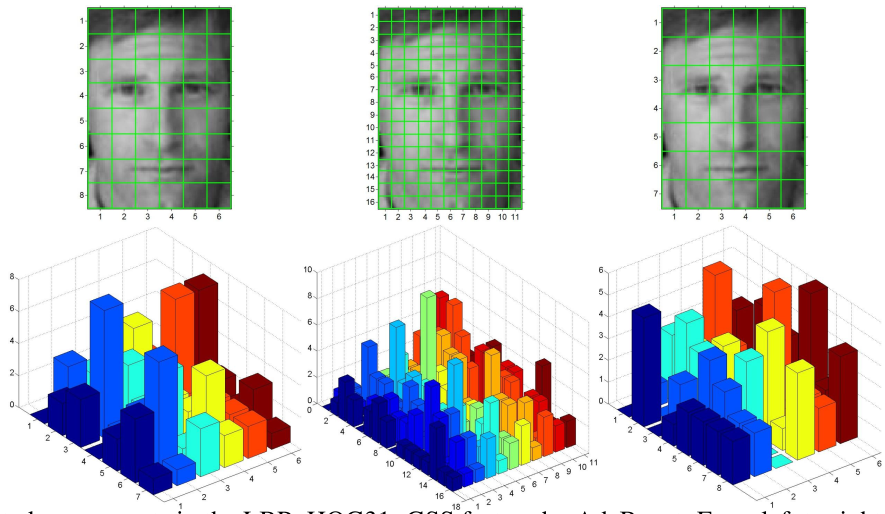

Fig. 3. Top 500 selected components in the LBP+HOG31+GSS feature by AdaBoost. From left to right, they are visualization pictures of LBP, HOG31 and GSS separately using these components.

Table I. Experimental results of gender identification with baseline approaches.

Approach			Accuracy (%)
Feature	Dimension	Classifier	
Raw pixels	500	AdaBoost	88.23 ± 0.91
	2,944	SVM	88.99 ± 0.24
Standard LBP	500	AdaBoost	90.55 ± 0.53
	2,478	SVM	91.43 ± 0.64
HOG36	500	AdaBoost	91.22 ± 0.84
	7,956	SVM	92.05 ± 0.97
HOG31	500	AdaBoost	92.99 ± 0.85
	5,456	SVM	$\mathbf{94.38 \pm 0.39}$
GSS	500	AdaBoost	86.91 ± 1.13
	2,304	SVM	$\mathbf{88.96 \pm 0.38}$

(HOG36) by 2.33%, which suggests that HOG with both contrast sensitive and insensitive information is more suited for gender classification than that with only contrast insensitive information. Here, GSS and one of our baselines, raw pixels, perform comparably, but note that, the dimensionality of GSS is smaller than raw pixels. Thus, it can be concluded that the new proposed feature, GSS, is a good choice for real-world gender recognition task.

Because GSS is based on HOG and it describes local self-similarities of a HOG map, the combination of these two features may achieve better performance than using any of them alone. Therefore, comparison experiments are executed and the results are listed in Table II. Here, we firstly use AdaBoost to select 500 weak learners and each weak learner corresponds to one component of a feature vector. Then, we adopt linear SVM to perform the gender identification on these selected components. The overall best performance is achieved by HOG31+LBP+GSS using a linear SVM with recognition rate of 95.76%, which is 0.88% higher than that of HOG31+LBP. This means that GSS indeed helps improve the performance of the final gender classifier.

The distribution of the top 500 selected components in the LBP+HOG31+GSS feature by AdaBoost is illustrated in Fig. 3. As we can see, the left bar chart demonstrates that discriminative LBP bins are mainly distributed in the regions around the eyes, which has also been mentioned in Shan's work [13]. Middle bar chart stands for the distribution of selected components in HOG31. Different from LBP, the information with strong capability of distinguishing genders mainly concentrate on the forehead and the mouth regions. When drawing the right bar graph, it is a little different from previous two graphes. For each selected component in GSS, the exact positions corresponding to it can be traced and we plus one in both of these two positions. Overall the visualization picture for GSS shows that the mouth, forehead, checks regions are where the useful comparisons located.

Table II. Experimental results of the concatenating features of different features.

Approach			Accuracy (%)
Feature	Dimension	Classifier	
HOG31+LBP	500	AdaBoost	93.06 ± 0.79
		SVM	94.88 ± 0.81
HOG31+LBP+GSS	500	AdaBoost	93.23 ± 0.57
		SVM	$\mathbf{95.76 \pm 1.21}$

5. CONCLUSIONS

Although there exists a lot of urgent requirements from real-life applications for practical gender identification, it still remains challenging in research area. Effective feature extraction is a vital step for gender identification. In this paper, we first propose a brand new feature GSS, which has achieved good performance on the unconstrained LFW face database. Second, by combining GSS with baseline features (HOG31 and LBP) and using AdaBoost as the best features selector, a better performance of identifying gender can be achieved. Experimental evaluation on the public LFW demonstrates the effectiveness of the proposed method.

6. REFERENCES

[1] Choon Boon Ng, Yong Haur Tay, and Bok-Min Goi, "Recognizing human gender in computer vision: a survey," in *Pacific Rim International Conferences on Artificial Intelligence (PRICAI)*, vol. 7458, pp. 335–346. 2012.

[2] Baback Moghaddam and Ming-Hsuan Yang, "Learning gender with support faces," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 24, no. 5, pp. 707–711, 2002.

[3] Shumeet Baluja and Henry A Rowley, "Boosting sex identification performance," *International Journal of Computer Vision (IJCV)*, vol. 71, no. 1, pp. 111–119, 2007.

[4] Hui-Cheng Lian and Bao-Liang Lu, "Multi-view gender classification using local binary patterns and support vector machines," in *Advances in Neural Networks*, vol. 3972 of *Lecture Notes in Computer Science*, pp. 202–209. 2006.

[5] Navneet Dalal and Bill Triggs, "Histograms of oriented gradients for human detection," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, pp. 886–893. 2005.

[6] Guodong Guo, Charles R Dyer, Yun Fu, and Thomas S Huang, "Is gender recognition affected by age?," in *IEEE International Conference on Computer Vision Workshops (ICCVW)*, pp. 2032–2039. 2009.

[7] Gregory Shakhnarovich, Paul A Viola, and Baback Moghaddam, "A unified learning framework for real time face detection and classification," in *IEEE International Conference on Automatic Face and Gesture Recognition (FGR)*, pp. 14–21. 2002.

[8] Meltem Demirkus, Matthew Toews, James J Clark, and Tal Arbel, "Gender classification from unconstrained video sequences," in *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 55–62. 2010.

[9] XueMing Leng and YiDing Wang, "Improving generalization for gender classification," in *IEEE International Conference on Image Processing (ICIP)*, pp. 1656–1659. 2008.

[10] Wei Gao and Haizhou Ai, "Face gender classification on consumer images in a multiethnic environment," in *Advances in Biometrics*, pp. 169–178. 2009.

[11] Jania Aghajanian, Jonathan Warrell, Simon JD Prince, Peng Li, Jennifer L Rohn, and Buzz Baum, "Patch-based within-object classification," in *IEEE International Conference on Computer Vision (ICCV)*. IEEE, 2009, pp. 1125–1132.

[12] Gary B. Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," Tech. Rep. 07-49, University of Massachusetts, Amherst, October 2007.

[13] Caifeng Shan, "Learning local binary patterns for gender classification on real-world face images," *Pattern Recognition Letters (PRL)*, vol. 33, no. 4, pp. 431–437, 2012.

[14] Stefan Walk, Nikodem Majer, Konrad Schindler, and Bernt Schiele, "New features and insights for pedestrian detection," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1030–1037. 2010.

[15] Pedro F Felzenszwalb, Ross B Girshick, David McAllester, and Deva Ramanan, "Object detection with discriminatively trained part-based models," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 32, no. 9, pp. 1627–1645, 2010.

[16] Carl Vondrick, Aditya Khosla, Tomasz Malisiewicz, and Antonio Torralba, "HOGgles: Visualizing Object Detection Features," in *IEEE International Conference on Computer Vision (ICCV)*. 2013.

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

[18] Jerome Friedman, Trevor Hastie, and Robert Tibshirani, "Additive logistic regression: a statistical view of boosting," *The Annals of Statistics*, vol. 28, no. 2, pp. 337–407, 2000.

[19] Lior Wolf, Tal Hassner, and Yaniv Taigman, "Similarity scores based on background samples," in *Asian Conference on Computer Vision (ACCV)*, vol. 5995, pp. 88–97. 2010.

[20] Timo Ojala, Matti Pietikainen, and Topi Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 24, no. 7, pp. 971–987, 2002.

[21] Pedro Felzenszwalb, David McAllester, and Deva Ramanan, "A discriminatively trained, multiscale, deformable part model," in *IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, pp. 1–8. 2008.