SMILE DETECTION IN UNCONSTRAINED SCENARIOS USING SELF-SIMILARITY OF GRADIENTS FEATURES

Hong Liu, Yuan Gao, Pinging Wu

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

ABSTRACT

Smile detection in unconstrained scenarios is a hot research topic with many real-world applications. This paper presents a new approach to practical smile detection and the primary contributions are three-fold. (1) In the image registration procedure, an eyes-mouth alignment strategy is found to be more efficient than popular eyes alignment. (2) In the feature extraction procedure, a novel feature descriptor, Self-Similarity of Gradients (GSS), is proposed and achieved good performance in comparison with baseline approaches. (3) Feature combination and multi-classifier combination strategies are adopted in experiments and excellent results are obtained. Experimental results show that the combined features (HOG+GSS) using AdaBoost+SVM achieve improved performance over state-of-the-art in the GENKI4K benchmark.

Index Terms— Smile Detection, Self-Similarity of Gradients, AdaBoost, SVM

1. INTRODUCTION

Smile is one of the most common facial expressions in daily communication, and automatic smile detection in unconstrained scenarios is important for its wide applications such as interactive systems, video conferences, digital video cameras and patient monitoring.

In the past two decades, a considerable amount of researches about automatic facial expression recognition have been done [1, 2, 3]. However, most existing works of facial expression recognition are based on the data collected by asking subjects to pose deliberately the expression [4]. Spontaneous expressions differ from posed ones because they have different muscle movements. Besides, spontaneous expressions are more subtle and fleeting than posed ones [5]. Recently, research focus begins to transfer to the more realistic problem of analyzing spontaneous facial expressions [5, 6]. The work of Whitehill et al. was the foundation of automatic smile detection module in commercial digital cameras [7]. They also designed the GENKI database, which contains over 63,000 real-life images from the Web, for the challenging smile detection in unconstrained scenarios. Moreover, Shan proposed a novel smile detection approach by simply comparing the intensities of a few pixels in a image and achieved better performance than Gabor+SVM [8]. A deceit detection of posed smile and spontaneous smile was implemented by training AU6 and AU12 simultaneously [9].

In a smile detection system, efficient image registration and feature representation are both important [7]. For image registration, works in [10, 11, 12] present usual ways of image registration, however, there exist few works studying alignments using how many facial landmarks contribute to the best smile detector. For feature representation, there are some feature extraction methods commonly used in facial expression recognition such as PCA [13], LDA [14], Gabor [6], Haar [15] and LBP [16]. More recently, HOG features have become one of the most popular features for object detection [17]. Bai et al. proposed to use the pyramidal representation of HOG (PHOG) as the features extracted for smile detection and its performance is comparable to Gabor [18]. Felzenszwalb et al. used an analytic dimensionality reduction approach to obtain low-dimensional HOG features, including contrast sensitive, contrast insensitive and gradient energy information [19]. Their HOG is adopted by us due to its low dimensions and outstanding performance in smile detection tasks.

In this paper, we focus on practical smile detection of face images in unconstrained scenarios. Therefore, GENKI4K is chosen as the evaluation database, which exactly meets the real-world conditions. Since image registration is significant for smile detection, we compare two different alignment strategies, eyes-based and eyes-mouth based. When using the same baseline feature extraction approaches, the latter one achieves much better performance than the former, which shows the efficiency of eyes-mouth alignment strategy. Based on the analysis of HOG's visualization and the inspiration from self-similarity on color channels (CSS) [20], a new feature, Self-Similarity of Gradients (GSS), is proposed

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).).

978-1-4799-5751-4/14/$31.00 ©2014 IEEE

to capture pairwise statistics of localized gradient distributions. In combination of HOG, the feature achieves improved performance over state-of-the-art using AdaBoost+SVM, indicating the effectiveness of our proposed GSS features.

2. IMAGE REGISTRATION

Image registration is one of the vital procedures of developing a high-performance smile detector [7]. It consists of steps including rotating, cropping and scaling, which are based on the pre-processing stage where facial landmarks have been found. Although automatic facial landmarks detection has been studied a lot [21, 22, 23], this work mainly focuses on the evaluation of our proposed GSS features. Therefore, we adopted the manual way and all the facial landmarks were hand-labeled by two students in our lab. There are two face alignment strategies for comparison:

Eyes Alignment Strategy. Two centers of the eyes of all the faces in GENKI4K are aligned to fixed locations.

Eyes-Mouth Alignment Strategy. Three fiducial points, centers of eyes and the mouth, are brought to fixed locations using an affine transform.

Eventually, after image registration, all the face images in GENKI4K possess the fixed size of 48×48.

3. FEATURE REPRESENTATION

As the GSS feature relies on HOG, in this section, the visualization of HOG and GSS features are described separately. The overall framework of our features extraction is presented in Fig. 1.

3.1. HOG Visualization

Visualizing features can help researchers gain a better understanding of the behaviours of detectors. In our smile detection work, two feature visualizing methods for HOG feature are tried and illustrated in Fig. 2. Two pictures in the first column stand for the "mean" faces of non-smile and smile images in GENKI4K. 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. corresponding to (a)(b).

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

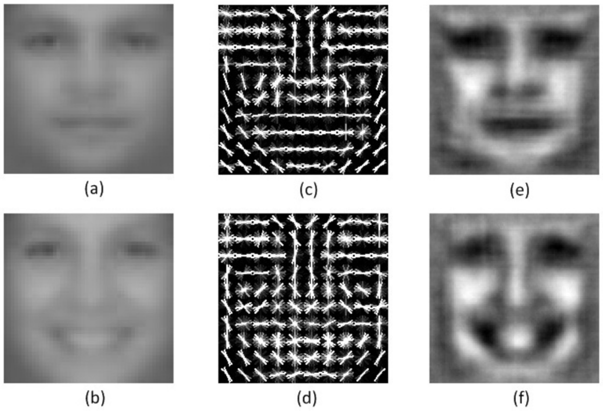

Fig. 2. The first column is the "mean" faces of non-smile (up) and smile (down). The second column is HOG visualization of the left "mean" faces [19]. The last column stands for the same HOG features as the second column but using the different visualizing method which has been described in great details [24].

3.2. Gradient of Self-Similarity Feature

It is well known that HOG features can express complex shapes of global distributions with gradient orientations. If adding local similarity information in HOG, a more discriminative classifier can be obtained. Motivated by this and the observation in previous subsection, a new feature named gradient of self-similarity (GSS) is proposed, which calculates local differences in HOG feature. The distances between "pixel" histograms can be regarded as the similarities in GSS. Several functions for comparing histograms are tested and the intersection comparison method is finally selected:

$$d(H_1, H_2) = \sum_{I} min(H_1(I), H_2(I)) \tag{1}$$

Here, H_1 and H_2 stand for two histograms of different cells in a HOG feature map.

As shown in Fig. 1, a cell is the basic computing unit of size $w_c \times h_c$. For an input image of size $w \times h$, we use replicate method to pad it with padsize value of $[h_c \ w_c]$. The output HOG feature map has $(w/w_c) \times (h/h_c) \times 31$ dimensions and the details are given in [19].

We only compare histograms in blocks of HOG. A block of $n \times n$ cells has $B_{compare}$ histogram comparisons, each of which corresponds to a value in the GSS feature vector.

$$B_{compare} = C^2_{(n \times n)} = n^2 \times (n^2 - 1)/2 \tag{2}$$

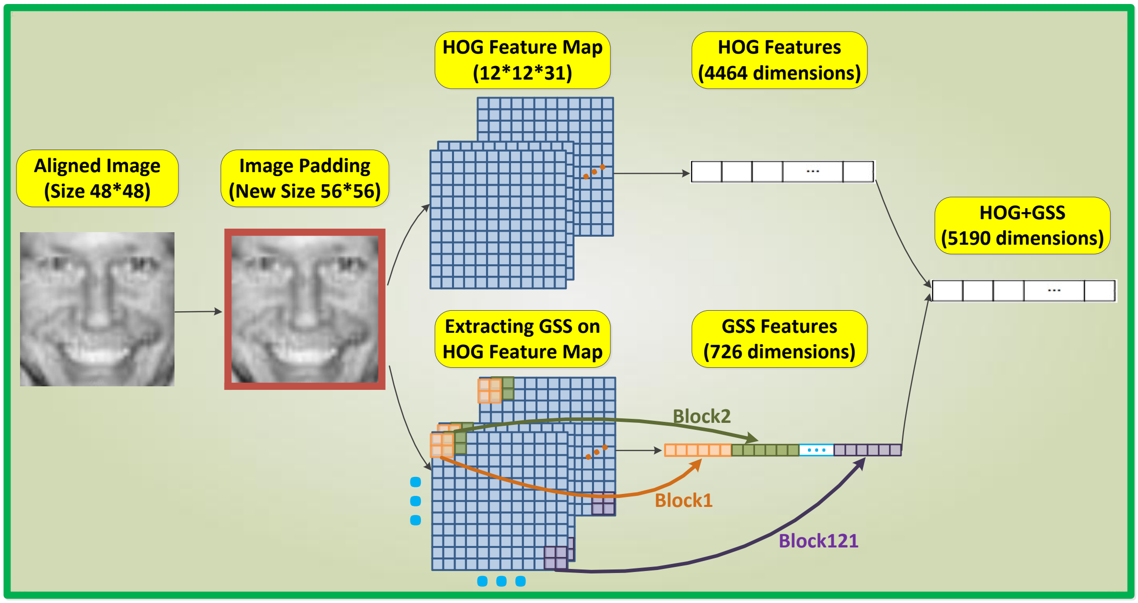
HOG Feature Map (12*12*31)
HOG Features (4464 dimensions)
Aligned Image (Size 48*48)
Image Padding (New Size 56*56)
HOG+GSS (5190 dimensions)
Extracting GSS on HOG Feature Map
GSS Features (726 dimensions)
Block2
Block1
Block121

Fig. 1. Framework of extracting HOG and GSS features.

With the block stride size of $k \cdot [h_c \, w_c]$, there are B_{num} blocks in a HOG map.

$$B_{num} = ((w/w_c - n)/k + 1) \times ((h/h_c - n)/k + 1) \quad (3)$$

Eventually, all the values of these histogram comparisons are combined into a $B_{num} \times B_{compare}$-dimensional vector, which is our proposed GSS features. The visualization of extracting GSS features is also exhibited in Fig. 1.

4. CLASSIFIERS

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

- For the first classifier, a linear SVM [25] is selected in consideration of its good performance, simplicity and, last but not least, the speed.
- For the second classifier, Gentle AdaBoost [26] 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].

5. EXPERIMENTS AND DISCUSSIONS

Experiments here are composed of three steps. The first step briefly introduced the “wild” smile database, GENKI4K. 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.

5.1. Smile Database

Our experiments are evaluated on the GENKI4K database. The reason why GENKI4K was chosen is that its contents come from real unconstrained scenarios and can be exploited to test the performances of feature extraction methods in real-world conditions. Here are some attributes of GENKI4K: (1) This database has 4,000 face images (1,838 “non-smile” and 2,162 “smile”). (2) The pose range (yaw, pitch, and roll of parameters of the head) of most images is within approximately ±20 degree of frontal. (3) GENKI4K also has various imaging conditions including, for instance, gender, age, ethnicity, glasses, facial hair, partial occlusion (very few) and so on.

5.2. Experimental Settings

The performance of our proposed GSS features are compared with Gabor, LBP and HOG. Their parameter settings are as following.

Gabor. The detailed parameters of Gabor feature used in our experiments are that 8 orientations and 5 spatial frequencies (9:36 pixels per cycle at 1/2 octave steps). We downsample the 40 Gabor Energy Filters by a factor of 4, so the Gabor

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

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

Table I. Smile detection accuracy of two different face alignment strategies using Gabor, LBP and HOG features.

Accuracy (%)	Image Registration Approach	
	Eyes-based	Eyes-mouth based
Gabor	91.18±0.43	93.81±0.36
LBP	88.64±1.08	90.48±0.86
HOG	91.78±0.39	93.83±0.39

feature vector has 23,040 dimensions.

LBP. For extracting LBP features, each face image of size of 48×48 is divided into 16 sub-regions of 12×12 pixels. Then we adopt 59-label LBP$(8, 2, u2)$ operator to compute LBP features for each sub-region. As a consequence, the LBP vector has 944 (16×59) dimensions.

HOG. The parameters of HOG are Extracting HOG and GSS features is shown in Fig. 1. When the cell is of size 4×4 ($w_c = h_c = 4$), the HOG feature is a 12×12×31 map. We concatenate each "pixel" in the map one by one sequentially. Therefore, the HOG feature is a 4,464-dimensional vector.

GSS. The details of GSS have been described in section 3 and GSS features used here are based on the above HOG. So $w_c = h_c = 4, n = 2, k = 1$ and the GSS has 726 dimensions.

HOG+GSS. Feature combination is a common strategy in previous smile detection work (PHOG+Gabor, see [18]). In addition, there is an important fact that HOG and GSS in our experiments have the same parameters, which means that the combination of these two features takes approximately the same time as computing GSS or HOG alone. However, PHOG and Gabor are two types of different features with nothing in common, so there are few techniques to employ to reduce the computation time.

At last, four-fold cross validation is adopted and can be briefly described below. All images in GENKI4K are divided into four heaps with the same ratio between non-smile and smile faces. Each time select one distinct heap for testing and use the other four heaps for training. This procedure is then repeated three times.

5.3. Results and Analysis

Detailed experimental results are shown in Table I and Table II. Exhaustive analysis are made based on the observations of the results.

Table I illustrates the results of baseline feature extraction approaches using eyes alignment and eyes-mouth alignment strategies. The common point is that the performances of baseline methods have significantly improved when using eyes-mouth alignment strategy. This indicates that information in the region of the mouth is very important for constructing an efficient smile detector. In addition, in image registration methods, the eyes-mouth alignment one is better than eyes alignment one. Therefore, in latter experiments, we adopted the eyes-mouth registration method.

Table II. Experimental results of smile detection compared with baseline approaches.

	Approach		Accuracy (%)
Feature	Dimension	Classifier	
Gabor	500	AdaBoost	92.11±0.48
	23,040	SVM	93.81±0.36
LBP	500	AdaBoost	89.94±0.43
	944	SVM	90.48±0.86
GSS	500	AdaBoost	84.69±0.61
	726	SVM	**87.74±0.57**
HOG	500	AdaBoost	92.48±0.74
	4,464	SVM	93.83±0.39
	500	AdaBoost+SVM	94.58±0.55
HOG+GSS	500	AdaBoost	92.88±0.53
	5,190	SVM	94.58±0.62
	500	AdaBoost+SVM	**95.13±0.95**

Table II demonstrates the smile recognition rates of our proposed GSS features with baseline methods. It's easy to find that, specified to any type of features, SVM performed better than AdaBoost. This suggests that SVM classifier is more suited for practical smile detection tasks than AdaBoost. When using the same classification method, SVM or AdaBoost, the order according to performances is: HOG+GSS $>$ HOG $\approx$ Gabor $>$ LBP $>$ GSS. We can conclude that (1) feature combination strategy is beneficial for improving the performance of a smile detector; (2) HOG and Gabor perform comparably, but in both dimensionality and computational complexity, HOG owns superior performances; (3) GSS achieves the smile recognition rate of 87.74%, which implies its efficiency for real-world smile detection tasks. In the end, together two traditional and famous classification methods contribute to the performance of the final smile detector, which indicates that multi-classifier combination is also very important for better performance. What's more, AdaBoost's ability of selecting discriminative features is powerful.

6. CONCLUSIONS

Smile detection has been widely applied in real-world applications, the performance of which directly decide the effects these applications. This paper firstly compares two different image registration approaches, the results of which show that mouth alignment is important for a smile detector. In addition, a new feature named GSS is proposed for practical smile detection. The combination of GSS and HOG achieve improved smile recognition rate when compared with other baseline feature extraction approaches, which implies the effectiveness of GSS features. Further more, to construct a more efficient smile detector, a multi-classifier strategy is adopted. The final smile detector using GSS+HOG with AdaBoost+SVM outperformed state-of-the-art on the unconstrained GENKI4K database.

7. REFERENCES

[1] Zhihong Zeng, Maja Pantic, Glenn I Roisman, and Thomas S Huang, “A survey of affect recognition methods: Audio, visual, and spontaneous expressions,” *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 31, no. 1, pp. 39–58, 2009.

[2] Beat Fasel and Juergen Luettin, “Automatic facial expression analysis: a survey,” *Pattern Recognition (PR)*, vol. 36, no. 1, pp. 259–275, 2003.

[3] Maja Pantic and Leon J. M. Rothkrantz, “Automatic analysis of facial expressions: the state of the art,” *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 22, no. 12, pp. 1424–1445, 2000.

[4] Takeo Kanade, Jeffrey F Cohn, and Yingli Tian, “Comprehensive database for facial expression analysis,” in *IEEE International Conference on Automatic Face and Gesture Recognition (FGR)*, pp. 46–53. 2000.

[5] Jeffrey F Cohn and Karen L Schmidt, “The timing of facial motion in posed and spontaneous smiles,” *International Journal of Wavelets, Multiresolution and Information Processing*, vol. 2, no. 2, pp. 121–132, 2004.

[6] Marian Bartlett, Gwen Littlewort, Mark Frank, Claudia Lainscsek, Ian Fasel, and Javier Movellan, “Automatic recognition of facial actions in spontaneous expressions,” *Journal of Multimedia*, vol. 1, no. 6, pp. 22–35, 2006.

[7] Jacob Whitehill, Gwen Littlewort, Ian Fasel, Marian Bartlett, and Javier Movellan, “Toward practical smile detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI)*, vol. 31, no. 11, pp. 2106–2111, 2009.

[8] Caifeng Shan, “Smile detection by boosting pixel differences,” *IEEE Transactions on Image Processing (TIP)*, vol. 21, no. 1, pp. 431–436, 2012.

[9] Hong Liu and Pingping Wu, “Comparison of methods for smile deceit detection by training AU6 and AU12 simultaneously,” in *IEEE International Conference on Image Processing (ICIP)*, pp. 1805–1808. 2012.

[10] Lior Wolf, Tal Hassner, and Yaniv Taigman, “Similarity scores based on background samples,” in *Asian Conference of Computer Vision (ACCV)*, pp. 88–97. 2010.

[11] Erno Makinen and Roope Raisamo, “Evaluation of gender classification methods with automatically detected and aligned faces,” *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 30, no. 3, pp. 541–547, 2008.

[12] Irfan A. Essa and Alex Paul Pentland, “Coding, analysis, interpretation, and recognition of facial expressions,” *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 19, no. 7, pp. 757–763, 1997.

[13] Matthew Turk and Alex Pentland, “Eigenfaces for recognition,” *Journal of cognitive neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[14] Peter N. Belhumeur, João P Hespanha, and David J. Kriegman, “Eigenfaces vs. fisherfaces: Recognition using class specific linear projection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 19, no. 7, pp. 711–720, 1997.

[15] Jacob Whitehill and Christian W Omlin, “Haar features for facs au recognition,” in *IEEE International Conference on Automatic Face and Gesture Recognition (FGR)*, pp. 5–101. 2006.

[16] Abdenour Hadid, Matti Pietikainen, and Timo Ahonen, “A discriminative feature space for detecting and recognizing faces,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2, pp. 797–804. 2004.

[17] 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.

[18] Yang Bai, Lihua Guo, Lianwen Jin, and Qinghua Huang, “A novel feature extraction method using pyramid histogram of orientation gradients for smile recognition,” in *IEEE International Conference on Image Processing (ICIP)*, pp. 3305–3308. 2009.

[19] 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.

[20] 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.

[21] Michal Uřičář, Vojtěch Franc, and Václav Hlaváč, “Detector of facial landmarks learned by the structured output svm,” *International Conference on Computer Vision Theory and Applications (VISAPP)*, pp. 547–556, 2012.

[22] Mark Everingham, Josef Sivic, and Andrew Zisserman, ““Hello! My name is... Buffy” – automatic naming of characters in TV video,” in *British Machine Vision Conference (BMVC)*. 2006.

[23] Timothy F. Cootes, Gareth J. Edwards, and Christopher J. Taylor, “Active appearance models,” *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 23, no. 6, pp. 681–685, 2001.

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

[25] 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.

[26] 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.