Robust Automatic Facial Expression Detection Method

Yan Ouyang
Institute for Pattern Recognition and Artificial Intelligence/ Huazhong University of Science and Technology, Wuhan, China
Email:oyy_01@163.com

Nong Sang
Institute for Pattern Recognition and Artificial Intelligence/ Huazhong University of Science and Technology, Wuhan, China
Email: nsang@hust.edu.cn

***Abstract*—Recently, the recognition of occluded facial expressions attract more and more people's attention. Sparse representation based classification (SRC) method gives good performance on face recognition (FR) and facial expression recognition (FER), well-known for its robustness to occlusion. Histograms of Oriented Gradient (HOG) descriptors are very efficient to represent the shape information of different facial expressions and robust to various illumination. Since, this paper proposes a novel method by using HOG descriptors conjunction with SRC framework for FER. Experiment results show that the proposed method gives better performance than the existing state-of-the-art methods. Furthermore, the proposed method is not only robust to assigned occlusions, but also to random occlusions.**

***Index Terms*—facial expression recognition, local patch, HOG descriptors, Sparse representation based classification (SRC), assigned occlusions, random occlusions**

I. INTRODUCTION

Automatic facial expression recognition (FER) is very important in the area of computer vision. FER has been widely used in many applications of Human-computer interaction. For example, FER technology has been used to classify your expressions and make the computers or robots response to your expressions. In psychology, human facial expressions are categorized into six classes of expressions (happiness, sadness, disgust, surprise, anger and fear) [1]. Most of automatic facial expressions analysis work focuses on classifying an input face image as one of those six basic expressions [2-10] [23-25]. The common process of those methods consist of two steps: the first step is extracting features; the second is using learning method or classifying method to build classifiers based on those features. The results show good performance when solving FER problem upon clean face images.

Although most facial expression recognition methods show good performance on clean face images, occlusions still remain a challenge problem when solving FER problems. There're two strategies to solve occlusion problems. The first strategy is dividing face image into local patches and extracting features based on these patches [11].The other is use robust classifier such as Sparse Representation based Classification (SRC) [12-15]. The proposed methods in [11] and [15] are the latest methods to solve occlusion problems in FER. In paper [11], they first divide face image into local patches and extract Gabor feature in each patches, the final judgment is made by using NN classifier. There are two problems in this method. First, Haar feature is proved to give better performance than Gabor feature when recognizing face components [16]. The other is that SRC method is proved to show better robustness than NN when solving occlusion problems [17]. In paper [15], their success is strongly dependent on precisely extracting the regions of eyes and mouth. But accurate face component detection is very difficult [2].

As mentioned above, the suitable strategy for robust automatic FER needs to contain three processes, the first is dividing face image into local patches, the second is extracting illumination invariance features based on these patches, the third is using SRC framework as the final classifier. Although Haar feature is good for face component recognition, it is not robust to illumination variance. HOG descriptors can extract efficient shape information of different facial expressions and well-known for its illumination invariance [18]. Since, one advantage of this paper is proposing a novel method to classify facial expressions based on HOG descriptors and SRC framework. The experiment on Cohn-database [19] show that the proposed method is outperforms existing methods for FER and robust to occlusion. Our experiment also shows that the strategy of extracting HOG descriptors is quite affecting the success of proposed method. Since, the strategy of extracting HOG descriptors is another advantage of this paper. In our experiment, we test the proposed method on face image with eye occlusion, mouth occlusion and random occlusion. In practical FER, we don't know which part of the face is occluded. So, the robustness to random occlusion is very significant.

doi:10.4304/jsw.8.7.1759-1764

Briefly, the main contributions of the paper are:

(1)Facial expressions classification process is automatic and does not need selecting ROIs(Regions of interest) for each face image.

(2)We use HOG descriptors conjunction with SRC framework to solve FER problems. The proposed method gives better performance on clean and occluded face images than the existing method based on SRC such as RAW+SRC, PCA+SRC.

(3)We test our method on random occluded facial expression images. Random occlusion problem is more difficult than other occlusion problems for FER method to solve.

The rest of the paper is organized as follows. Section 2 introduces the strategy of HOG descriptor extraction. Section 3 introduces Sparse Representation based Classification. Section 4 introduces the proposed method in this paper. Section 5 shows the experiment results. Section 6 concludes the paper.

II. HOG Descriptors Extraction Strategy

As mentioned in [18], the general process of HOG descriptors can be described as Fig. 1.

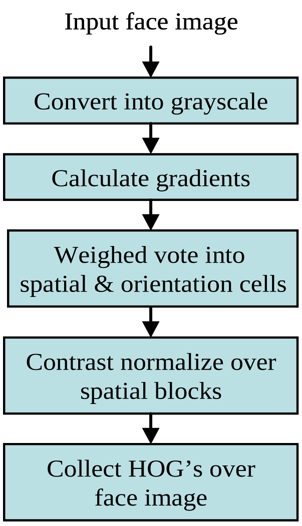

Figure 1. The general process of extracting HOG descriptors.

As shown in Fig.1, it can be seen that HOG descriptors extract shape information of face images. In the fourth step of process, HOG descriptor execute contrast normalize over spatial blocks. Since HOG descriptor is gray-level invariance. The main procedure which affecting the success of FER is step 3. In step 3, the partition of spatial cells and selection of orientation bins are the most important. In [20], face component region is dividing into eight cells with no overlap between neighboring cells. As mentioned in [2] and [11], two neighboring cells of face image are interrelated. This relationship will help classifier to give more efficient performance when solving FER problem. So, they divide face image into local patches and each patch is 50% overlapping with neighboring patch. It is hard to say which segmentation is better, so we choose six separation strategies in this paper. As shown in Fig.2, our experiment image is normalized to the size of 64×64. The six segmentations are: (1) face images are divided into 961 cells , cell size is 4×4 and neighboring cells are 50% overlapping; (2) face images are divided into 256 cells ,cell size is 4×4 and neighboring cells are no overlapping; (3) face images are divided into 225 cells, cell size is 8×8 and neighboring cells are 50% overlapping; (4) face images are divided into 64 cells, cell size is 8×8 and neighboring cells are no overlapping; (5) face images are divided into 49 cells, cell size is 16×16 and neighboring cells are 50% overlapping; (6) face images are divided into 16 cells, cell size is 16×16 and neighboring cells are no overlapping.

(1) (2) (3) (4) (5) (6)

Figure 2. Examples of six segmentation strategies

Another important parameter is orientation bin numbers. In the application of human detection, "fine orientation coding turns out to be essential for good performance" [18].But facial expression is a six-class classification problem, redundant information may influence the result of FER. So, we choose six bin numbers N=2, 6,8,9,12,18 with space range [0,360]. The rest parameters are described as below: (1) We calculate the orientation gradients by using 3×3 Sobel mask without Gaussian smoothing; (2) Each spatial block contains four cells; (3) we use L2-norm method to implement block normalization. The experiment result in section 5 shows that the best selections of parameters are: (1) cell size is 4×4 and neighboring cells are 50% overlapping; (2) The bin number N=9 which spaced over [0,360].

III. Sparse Representation Based Classification

Essentially, sparse representation is representing a signal y by using a dictionary $T=\{t_1,t_2,\dots t_n\}$ .The weights x corresponding to T is calculated by solving an l_1-minimization problem:

$$\min\|x\|_1 \quad s.t.\, Tx=y \tag{1}$$

As proposed in [12], examples of six expression classes are available and used to construct a training set when solving FER problems. Each example may be holistic face image features (eg. Eigenfaces[17]) or feature vector extracted from the facial expression images (e.g. Gabor features[12]). Assume that each expression class contains n_i(i=1,2,3,4,5,6) examples where i denote the number of class. So we can construct a dictionary T from the training set:

$$T=\{T_1,T_2,T_3,T_4,T_5,T_6\} \quad T_i=\{t_1^i,t_2^i,\cdots,t_{n_i}^i\} \tag{2}$$

Where T_i represents a set of examples belongs to the ith class and n = n1+n2+n3+n4+n5+n6. If there is a test sample y, according to the SRC framework, we represent y as a sparse linear combination of elements from dictionary T:

$$y = \alpha_1^1 t_1^1 + \cdots \alpha_{n_1}^1 t_{n_1}^1 + \cdots + \alpha_1^6 t_1^6 + \cdots \alpha_{n_6}^6 t_{n_6}^6 \quad (3)$$

In the ideal case, y will only represent by the elements of one class in the dictionary T which y belongs to. So formation (3) can be rewritten as:

$$y = Tx_0 \quad (4)$$

where $x_0 = [0,0,0,\dots,\alpha_1^i,\alpha_2^i,\dots\alpha_{n_i}^i,\dots 0,]$ a coefficient vector is whose entries are zero except those associated with the ith class. Since the linear representation is sparse and can be obtained by solving formation (1), it means that the x_0 can be calculated by solving an l_1-minimization problem as shown in formation (5):

$$\min\|x\|_1 \quad s.t. \quad Dx = y \quad (5)$$

In practical, y can't be ideally represented by the elements of ith class in dictionary T. So, we need to calculate six reconstruct value $T\delta_i(x)$ where $\delta_i(x) = [0,\dots,x_1^i,x_2^i,\cdots,x_{n_i}^i,\dots,0,]$ and compare the value of residual between each $T\delta_i(x)$ and y. Finally, the classification decision can be obtained by formation (6).

$$identify(y) = \arg\min_i \|y - T\delta_i(x)\|_2 \quad (6)$$

As mentioned in [17], SRC can solve two problems: occlusion and corruption. When a face image y' with corruption or occlusion, it can be seen as consist of a clean face image y_0 and a noise model e:

$$y' = y_0 + e \quad (7)$$

When use SRC to classify the image y', formation (7) can be rewritten with considering of formation (4):

$$y' = y_0 + e = Tx_0 + e = [T, I] \cdot \begin{bmatrix} x_0 \\ e_0 \end{bmatrix} = B\omega_0 \quad (8)$$

Where B = [T, I], so we can use formation (5) to solve formation (8):

$$\min\|\hat{\omega}\|_1 \quad s.t. \quad B\omega = y' \quad (9)$$

For $\hat{\omega} = [\hat{x},\hat{e}]$ is computed by formation (9), formation (6) can be rewritten as:

$$\begin{aligned} identify(y') &= \arg\min_i \|y' - T\cdot[\delta_i(\hat{x}),\delta_i(\hat{e})]\|_2 \\ &= \arg\min_i \|y' - e - T\cdot[\delta_i(\hat{x})]\|_2 \quad (10) \\ &= \arg\min_i \|y_0 - T\cdot[\delta_i(\hat{x})]\|_2 \end{aligned}$$

So, SRC will recover a clean image from the dictionary which compensates the noise model of corruption and occlusion.

IV HOG DESCRIPTORS CONJUNCTION WITH SRC

As mentioned above, there're three key points need to consider when designing a FER method for solving occlusion problems: (1) Segment face image into local patches; (2) Find a suitable feature transformation based on these patches; (3) Use SRC as classifier. In [11], Gabor feature was extracted based on local patches. But [16] demonstrates that haar-like feature is better than Gabor features for expression recognition. However haar-like features is not gray-level invariance, so it need preprocessing such as global histogram equalization [2].But local histogram equalization seems more suitable than global histogram equalization for texture recognition [22].As mentioned in section II, the process of extracting HOG descriptors has considered local histogram equalization. Since, this section we proposed a novel method based on HOG descriptors and SRC framework. The proposed method is consisting of three components. First, face images are divided into 961 local patches and the neighboring patches are designed to have 50% overlapping; Second, HOG descriptors features are extracted based on these local patches, finally SRC framework was used as classifier to recognize facial expressions. This algorithm is shown in Table I:

TABLE I.
THE ALGORITHM OF HOG+SRC METHOD

1. Input: 1. A test sample y 2. An dictionary D which constructed by training samples from six expression classes
Step 1: Calculate the gradients of test sample y. Step 2: Divide y into 961 local patches and extract HOG descriptors features based on these local patches to obtain a feature vector. The feature transformation can be seen as linear projection. We use P to represent this projection. Step 3: Appling P to both side of formation (4): $P\cdot y = P\cdot D\cdot x$ Step 4: Calculating the x by solving l_1-minimization problem: $\min\|x\|_1 \quad s.t. \quad PDx = Py$ Step 5: Judging the class where test sample y belongs to: $identify(Py) = \arg\min_i \|Py - PD\delta_i(x)\|_2$

V EXPERIMENT

Experiments are conducted on the Cohn-Kanade facial expression database [19]. We selected 339 image sequences from 94 subjects. Table II shows the sequence numbers of each expression used in the experiment. We select the last three frames of each sequence to organize our dataset and normalize the entire dataset to 64×64 pixels by using the same method as [2].

The validation in our experiment is person-independent. We randomly select 66% subjects to organize the dictionary and the rest subjects are used as test samples. This validation method is the same as [2]. All the programs are implemented by using MATLAB and experiments are conducted on the computer with 3.00GHz Intel CoreTM Quad processor and 8GB memory. L1_ls Toolbox [21] is used to solve l_1-minimization problem and λ is 0.001.

TABLE II.
SEQUENCE NUMBERS OF EACH EXPRESSION USED IN THE EXPERIMENTS

Anger	Fear	Happiness	Disgust	Sadness	Surprise
38	54	75	37	60	75

A.HOG Descriptors Parameter Selection

As mentioned above, we need to find a suitable face image segmentation strategy and orientation bin numbers for HOG descriptors. We research six segmentation strategies (as shown in section II) and six bin numbers. Fig. 3 shows the results of different parameter selection. It can be seen that when face images are divided into 961 local patches and bin number set to 9, the experiment result is the best. The best accurate rate is 94.1%.

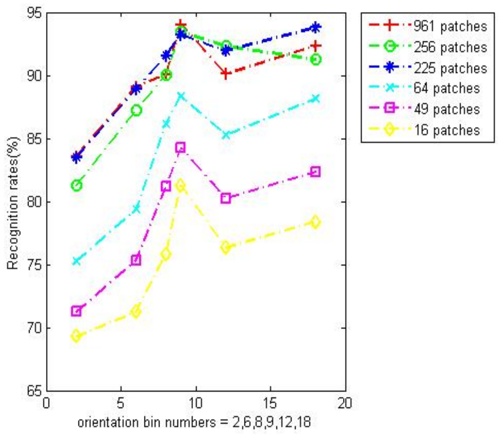

Figure 3. Recognition rates based on different parameters

B. Experiment on Clean Face Image

In this part, we test our proposed method (HOG+SRC) on our dataset without any occlusion on face image. Table III shows the confusion matrix of our proposed method.

There is an issue we need to focus on: Is the proposed method more efficient than the exist methods based on SRC framework? Since, we implement four automatic FER methods based on our dataset and the results are shown in Table IV. For example, RAW+SRC [12], Gabor+SRC[12], Eigen+SRC [12], LBP+SRC[13].

We also compared the proposed method to other automatic FER methods. The results are shown in Table V. But in table V, the results are not directly comparable due to different normalize methods, validation methods, and so on. However, those methods are person-independent automatic FER approaches. So, we can say that our proposed method is efficient when solving person-independent FER problem through the comparison.

C. Experiment on face image with assigned occlusion

In this paper, we occlude eyes region and mouth region of face image to test our proposed method. As shown in Fig.4, there're three degrees of occlusion such as small occlusion, medium occlusion and large occlusion.

We also implement four FER methods based on SRC on our dataset. As shown in table VI, it can be seen that our proposed method can still reach an accuracy rate of 85.12% when solving large occlusion of eye. And the performance of our proposed methods is better than those four methods based on SRC. From table VI, we can see that mouth occlusion problem is more difficult than eye occlusion problem for our method to solve.

(1) (2) (3) (4) (5) (6)

Figure 4. Examples of face image with assigned occlusion. (1) Eye occluded with small occlusion. (2) Eye occluded with medium occlusion. (3) Eye occluded with large occlusion. (4) Mouth occluded with small occlusion. (5) Mouth occluded with medium occlusion. (6) Mouth occluded with large occlusion.

TABLE III.
THE CONFUSION MATRIX (%) OF OUR PROPOSED METHOD BASED ON OUR DATASET

	Anger	Disgust	Fear	Happi-ness	Sadness	surprise
Anger	84.62	0	0	0	15.38	0
Disgust	0	93.33	0	0	6.67	0
Fear	0	13.33	86.67	0	0	0
Happi-ness	0	0	0	100	0	0
Sadness	0	0	0	0	100	0
Surprise	0	0	0	0	0	100

TABLE IV.
PERFORMANCE ON OUR DATASET

Methods	Accuracy(%)
RAW+SRC	85.70
Gabor+SRC	81.34
Eigen+SRC	80.32
LBP+SRC	59.34
Ours	94.1

TABLE V.
COMPARISON WITH DIFFERENT APPROACHES ON THE CK DATABASE

	Subjects	Classes	Measures	Rate(%)
[2]	96	6	—	92.3
[3]	96	6	10-fold	92.1
[4]	90	7	—	93.3
[5]	97	6	—	93.8
[6]	97	6	5-fold	90.9
[7]	90	6	—	93.66
[11]	94	7	10-fold	91.51
ours	94	6	—	94.1

TABLE VI
RECOGNITION RATES (%) ON THE OCCLUSION FACE IMAGE

	Eye small	Eye Medium	Eye Large	Mouth Small	Mouth Medium	Mouth Large
Ours	92.18	87.54	85.12	82.15	76.35	71.45
RAW+ SRC	83.54	75.45	60.12	75.35	69.25	51.56
Gabor+ SRC	76.32	68.23	60.23	72.12	63.12	52.12
Eigen+ SRC	71.23	62.23	51.12	70.23	64.12	50.23
LBP+ SRC	56.23	49.32	41.02	50.23	45.12	32.12

D. Experiment on Face Image with Random Occlusion

In practical FER, random occlusion must be considered. In this section, we conduct four experiments to study the performance of proposed method when solving random occlusion problem. In each experiment, we use a block with different size to occlude all our test samples and the position of occluded part is random distributed on the face image. These four block sizes are 10×10, 15×15, 20×20 and 25×25. Fig.5 shows some examples of random occlusion with different block.

(1) (2) (3) (4)

Figure 5. Examples of face image with random occlusion. (1) Random occluded with 10×10 block, (2) Random occluded with 15×15 block, (3) Random occluded with 20×20 block, (4) Random occluded with 25×25 block,

As shown in Table VII, it can be seen that the accuracy rate of proposed method can still reach 80.12% when using 25×25 block random occlusion. Since, the proposed method is robust to random occlusion.

TABLE VII
RECOGNITION RATES (%) ON THE RANDOM OCCLUSION FACE IMAGE

Random occlusion	10×10	15×15	20×20	25×25
Ours	92.21	91.35	89.17	80.12

VI CONCLUSION

In this paper, we find a suitable feature for SRC framework to solve facial expression recognition. HOG descriptors can not only solve the various illuminations, but also extract efficient shape information for FER. So we propose a new method: HOG+SRC. Experiments on the CK database show that the proposed method is efficient and robust to assigned and random occlusions.

REFERENCES

[1] C.E. Izard. The face of emotion. New York: AppletonCentury-Crofts.1971.1.

[2] P. Yang, Q.S. Liu, and D. N. Metaxas. Exploring facial expressions with compositional features.IEEE International Conference on Computer Vision and Pattern Recognition, 2010, pp:2638-2644.

[3] C. Shan, S. Gong, P.W. McOwan. Robust facial expression recognition using local binary patterns. IEEE International Conference on Image Processing.2005.pp:370-373.

[4] G. Littlewort, M. Bartlerr, I. Fasel, J. Susskind, and J. Movellan, Dynamics of facial expression extracted automatically from video. IEEE Workshop Face Processing in Video,2004.

[5] Y. Tian, Evaluation of face resolution for expression analysis, IEEE Workshop Face Processing in Video, 2004.

[6] M. Yeasin, R. Builot, and R. Sharma. From facial expression to level of interest: a spatio-temporal approach, Conference on Computer Vision and Pattern Recognition,2004, pp.922-927.

[7] P.S. Aleksic, and A.K. Katsaggelos. Automatic facial expression recognition using facial animation parameters and multi-stream HMMS,IEEE Transactions on Information Forensics and Security 1(2006) pp:: 3-11.

[8] I. Kotsia, S. Zafeiriou, and I. Pitas. "Texture and shape information fusion for facial expression and facial action unit recognition", Pattern Recognition, 41,3,2008. pp.833-851.

[9] M. Pantic, and L.J.M. Rothkrantz. Automatic analysis of facial expressions: The state of the art.IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000.1

[10] J. Ou, X.B. Bai, Y. Pei, L. Ma, and W. Liu. Automatic facial expression recognition using Gabor filter and expression analysis. International Conference on Computer Modeling and Simulation 2010.pp:215-218.

[11] G. Wenfei, C. Xiang, and Y.V. Venkatesh. Huang. D.,& Lin. H., Facial expression recognition using radial encoding of local Gabor features and classifier synthesis. Pattern Recognition.,2012.45(1). pp:80-91.

[12] S.F. Cotter. Sparse representation for accurate classification of corrupted and occluded facial expressions.Proc.ICASSP., 2010, pp: 838-841,.

[13] M.W. Huang, Z.W. Wang, and Z.L. Ying. A new method for facial expression recognition based on Sparse representation plus LBP. International Congress on Image and Signal Processing (CISP2010), pp:1750-1754.

[14] M.W. Huang, and Z.L. Ming. The performance study of facial expression recognition via sparse representation. International conference on Machine Learning and Cybernetics(ICMLC).

[15] S.F. Cotter. Recognition of occluded facial expressions using a fusion of localized sparse representation classifiers. Digital Signal Processing Workshop and IEEE Signal Processing Education Workshop(DSP/SPE), 2011 .pp:4-7.

[16] J. Whitehill, and C.W. Omlin. Haar features for facs au recognition.FG,2006.2

[17] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, and Ma Y. Robust face recognition via sparse representation. IEEE Trans. on PAMI,31(2) ,2009, pp:210-227,.

[18] N. Dalal, and B. Triggs. Histograms of oriented gradients for human detection. IEEE Conf. on Computer Vision and Pattern Recognition 2005, pp.886-893.

[19] T. Kanade, J.F. Cohn, And Y. Tian. Comprehensive database for facial expression analysis in: International Conference on Automatic Face and Gesture Recognition. 2000, pp.46-53.

[20] Z.S. Li, J. Imai, and M. Kaneko. Facial expression recognition using facial-component-based bage of words and PHOG descriptors. Information and Media Technologies 5(3),2010, pp:1003-1009.

[21] S.J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky. A method for large-scale l1-regularized least squares. IEEE

Journal on Selected Topics in Signal Processing, 1(4),2007, pp:606–617.

[22] T. Ojala, M. Pietikaninen, and,T. Maenpaa. Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns.IEEE Trans on PAMI.2002(24) pp:971-987

[23] M.Z. Yang, K.q Wang and L.Zhang. Realistic Real-time Facial Expressions Animation via 3D Morphing Target. Journal of Software, Vol 8, No2(2013),pp 418-425.

[24] H.Madokoro, K.Sato. Orientation Selectivity for Representing Dynamic Diversity of Facial Expressions. Journal of Computers, Vol 7, No 9(2012),pp 2107-2113.

[25] Y.J.Fu, H.Yan. J.W. Li and R.x. Xiang. Robust Facial Features Localization on Rotation Arbitrary Multi-View face in Complex Background. Journal of Computers Vol 6, No 2(2011),pp 337-342.

Yan Ouyang received B.E. and M.E. degrees in Software engineering and Computer science from Wuhan University of technology, China, in 2006 and 2009, respectively. He is currently a doctoral student in Institute for Pattern Recognition and Artificial Intelligence, the Huazhong University of Science and Technology, Wuhan, China. His research interests include facial expression analysis, computer vision and pattern recognition.

Nong Sang received B.E., M.E. and D.E. degrees from the Huazhong University of Science and Technology, China in 1990, 1993, and 2001, respectively. He is currently a professor of Institute for Pattern Recognition and Artificial Intelligence. His research interests include the Biological perceptual model and its application in computer vision and pattern recognition, visual cognition, medical image analysis, video surveillance.