April 29, 2017

**Responses to Review on the Revised Submission of**

## image4622

**Title: Illumination and Pose Variable Face Recognition via Adaptively Weighted ULBP\_MHOG and WSRC**

**Authors: Kang Wang, Zhenxue Chen, Jonathan Wu, Chengyun Liu**

Dear Editor,

Thank you very much for your letter dated April 7, 2017 and the attached comments from the technical editor and reviewers on the captioned paper.

We would like to give our great and sincere gratitude to the editor and reviewers for their valuable comments and constructive feedback on our previous submission.

In this new submission, each comment of the editor and reviewers has been carefully taken into consideration. The attached note explains the changes made in the revised version and our responses to each comment.

In addition to a careful incorporation of these comments, we went through the manuscript a few more times and improved the writing and presentation significantly, so that it is clearer and more succinct to readers.

Thank you again for processing our paper.

Sincerely yours,

|  |  |
| --- | --- |
| Kang Wang | Zhenxue Chen |
| *School of Control Science & Engineering* | *School of Control Science & Engineering* |
| Shandong University | Shandong University |
| *Jinan, P.R.China* | *Jinan, P.R.China* |
| [kangkang\_wanglucky@163.com](mailto:kangkang_wanglucky@163.com) | [chenzhenxue@sdu.edu.cn](mailto:chenzhenxue@sdu.edu.cn) |
| Jonathan Wu | Chenyun Liu |
| *Department of Electrical & Computer Engineering* | *School of Control Science & Engineering* |
| University of Windsor | Shandong University |
| *Windsor N9B 3P4, Canada* | *Jinan, P.R.China* |
| [wu@uwindsor.ca](mailto:wu@uwindsor.ca) | [liuchengyun@sdu.edu.cn](mailto:liuchengyun@sdu.edu.cn) |

**In the following, we simply list all the comments of the editor and each reviewer in *italic* font. It is immediately followed by how we revised the manuscript according to them (in regular font).**

**Response to Editor**

*Highlights consist of a short collection of bullet points that convey the core findings of the article and should be submitted in a separate file in the online submission system. Please use 'Highlights' in the file name and include 3 to 5* *bullet points (maximum 85 characters, including spaces, per bullet point).*

Thank you for processing our paper. We have written four bullet points in a separate file called Highlights, where per bullet point does not exceed 85 characters.

**Response to Reviewer 1**

*A challenging problem in face recognition i.e. illumination and pose variable face recognition has been discussed. An adaptive weighted ULBP\_MHOG and WSRC method is proposed to solve the problem. The literature review, method explanation and experimental section is valid. However, there are some points that should be considered in revised manuscript and those points are as follows:  
1. How the pose variation is minimized, that should be explained.*

Thank you for reviewing our paper. LBP and HOG features are not sensitive to illumination and poses. According to this characteristic, we extract the weighted ULBP\_MHOG features (WULBP\_MHOG) that are the fused features to describe face images with variant illumination and pose better. Meanwhile, WSRC has better performance as the classifier. The effective combination of WULBP\_MHOG and WSRC is used to minimize pose variation.

We also have added the explanation of minimizing pose variation in the first paragraph of Section 3 in the new version.

*2. Authors should explain the deficiency and shortages of previous methods in literature. Moreover, explaining how proposed method is better than others.*

The previous methods can be divided into 3D-based methods and 2D-based methods. The elapsed time is large for 3D-based methods, because they need large amount of 3D shape information for computation. As for 2D-based methods, most methods are used to solve the only illumination or pose variant face recognition. Some methods solve illumination and pose variant face recognition together, such as LBP-Gabor, LBP-HOG features, the deep learning, etc. The dimension of Gabor is large due to many Gabor wavelet kernels used in feature extraction. Our method has less computation than 3D-based and deep learning methods. In order to increase accuracy and solve illumination and pose variant together, the proposed method chooses the effective WSRC classifier and discriminating WULBP\_MHOG features with lower dimensions.

We have added the specific explanation of deficiency and shortages of previous methods in all paragraphs of Section 2 in the revised version. And the advantages of our method have been explained in the last paragraph of Section 2 in the revision.

*3. Investigating the recent advancement of deep learning (CNN or Alexnet) in classification problems, authors should explain the drawbacks of deep learning in the case of illumination and pose. Moreover, explaining the drawbacks of deep learning. The comparison between hand-crafted features like LBP and HOG used in your case against deep learning based methods would make this paper more solid.*

The drawbacks of deep learning are as follows: (1) It needs large scale training databases that are lacking in real world; (2) The deep nets are complex in training process, and it needs many experiments to verity and obtain a simpler and effective deep net for different databases; (3) It also needs many experiments to search for suitable parameters in deep models; (4) Because of the complex computing process and large scale training databases, the GPU is required. When the deep learning method is used to face recognition with illumination and pose changes, the drawbacks proposed above also exist.

The explanation has been added in the last paragraph of Section 2 in the new version too. And the comparison between hand-crafted features like LBP and HOG used in our case against deep learning based methods has been shown in Table 1, Table 2, Table 6 and Table 8 of Section 4 in the new version. In addition, we also have added some discussions in the last paragraph of Section 4.1, 4.2, 4.5 and 4.6, respectively. Here, we use CNN model as Fig. 1 to do experiments. As for different databases, we need to adjust CNN parameters many times. The elapsed time is large. Besides, our training sets are not large scale, which affects the accuracy in the part of experiments used CNN.



**Fig. 1** The used CNN architecture

(Note: There are 5 layers, where the first and the third layers are convolution layers, the second and the fourth layers are sub-sampling layers, and the last layer is fully-connected layer. The template size considered in convolution and sub-sampling layers are 5 × 5 and 2 × 2, respectively. The numbers of feature maps in the first and the third layers are 5 and 10, respectively. The second and the fourth layers have the same numbers of feature maps as the first and the third layers, respectively. In addition, the face images are cropped to 28 × 28 pixels and given as the input to CNN.)

*4. The evaluation protocol in the case of Yale B database is confusing. In Yale B database there are 5 different sets depending on illumination angle. There are extended Yale B database of 38 subjects. Author should test that also.*

We have added discussions and experiments on the five subsets of the Yale B database in Table 3, Table 4 and the last paragraph of Section 4.3 in the revised version.

Experimental results on the Extended Yale B database have been added in the Section 4.4 in the new version, where five different sets have also been tested respectively.

*5. In the case of illumination variation, author should compare the accuracy results with the state-of-the-art methods such as LG-face (Reference no.9).*

We have added comparison results with some state-of-the-art methods on five databases, which are shown in the Table 1, Table 2, Table 4, Table 6 and Table 8 of Section 4 in the new vision. As for the Extended Yale B database, we also have added comparison results with LG-face in Table 6, whose results are from Ref. 14. (Note: The original Ref. 9 has been changed to Ref. 14 in the revised version. We email to authors of Ref. 14 for paper codes, but we do not receive the reply. So we cannot run their paper codes on our databases. We just can reference their paper results to compare with us only on the Extended Yale B database.) Experimental results show that our method has greater performance.

**Response to Reviewer 2**

*Pose and Illumination are huge problem in Face recognition problem. This paper proposes a solution to these problem by using adaptive weighted features which combine ULBP and MHOG and WSRC.  
The paper clarifies the idea clearly, and it is easy to understand the procedure. However, I have some comments and questions about the paper.  
  
1. Well, the structure of this paper is clear. But, it seems like a report for the work. In my understanding, the MHOG and WSRC may be novel part. It is lack of theoretical analysis. Why these methods work well? There should be some explanations about these.*

Thank you for reviewing our paper.

As for MHOG features, the MHOG features include more gradient information, compared with the original HOG features. Mostly, the feature dimension of MHOG decreases greatly, for blocks are not divided into cells. The detailed analysis has been added in the last paragraph of the Section 3.4 in the new version.

As for the WSRC classifier, it combines both linearity and locality information for face recognition. SRC just uses the linearity structure of data, while locality information is more important than linearity under some conditions. So WSRC has better performance. The detailed analysis has been added in the second paragraph of the Section 3.7 in the revised version.

Besides, we have added some analysis and explanations to keep continuity of the context.

*2. Using information entropy as the weighted looks right. Is there other methods to generate weight? Adding weight to feature is a small idea in my opinion. Comparison can make the whole idea more convincing.*

There are two types of fixed and adaptive weights. As for fixed weights, they are obtained by experience. Due the main parts which include brows, eyes, nose and mouth have much more useful information for face recognition, so these areas can be given bigger weights. In experiments, the weights of mouth and nose are three, and the weights of brows and eyes are two. Besides, the weights of other areas are one. The detailed explanation has been added in the first paragraph of Section 3.5. As for adaptive weights, the information entropy method is commonly used in most papers. The comparisons about above two methods are added in Table 1, Table 2, Table 4, Table 6 and Table 8 of Section 4 in the new version, where Fixed\_WULBP \_MHOG+WSRC, Fixed\_WULBP\_MHOG+SVM, Fixed\_WULBP\_MHOG+RF and Fixed\_WULBP\_MHOG+NN use fixed weights, and ours (WULBP\_MHOG+WSRC), WULBP\_MHOG+SVM, WULBP\_MHOG+RF and WULBP\_MHOG+NN use the information entropy method. The comparison results show that the adaptively weighted method is superior to the fixedly weighted method with different classifiers in most cases. Due the contribution of each area has been influenced by internal and external conditions, such as different people, illumination, pose, etc. Information entropy method can adjust weights automatically for above changes, compared with the fixedly weighted method. And we also have added the discussion in the fifth paragraph of Section 4.6 in the revision.

*3. I think the proposed methods like adaptively weighted features, WSRC or MHOG cannot deal with Illumination and Pose problems. In other word, the relation between then problem and proposed methods is weak. The Illumination problem is solved by Tan’s method which is existing method. And I cannot see the relationship between pose problem and proposed method too.*

As for illumination and pose problems in face recognition, there are some existing methods using fused features to solve illumination and pose variant face recognition together, such as LBP-Gabor, LBP-HOG features, etc. The dimension of Gabor is large due to many Gabor wavelet kernels used in feature extraction. LBP and HOG features are not sensitive to illumination and poses. According to characteristics proposed above, we extract the weighted ULBP\_MHOG features (WULBP\_MHOG) which fuse features to describe face images with variant illumination and pose better. Meanwhile, WSRC has better performance as the classifier, which is robust to pose and illumination. The effective combination of WULBP\_MHOG and WSRC can improve the accuracy in face recognition with illumination and pose variant. In Section 2 in the new version, we have added the specific explanation about disadvantages of existing methods and advantages of our method.

In addition, we have added the analysis of MHOG in the last paragraph of Section 3.4 and the analysis of WSRC in the second paragraph of Section 3.7 in the revised version. In our paper, we propose the method that combines Tan’s method with our proposed WULBP\_MHOG and WSRC, which can improve performance in illumination and pose variant face recognition together. In order to test our proposed method, extensive experiments have been conducted on five databases in the revised version, which include the ORL, the Yale, the Yale B, the Extended Yale B and CMU-PIE databases with illumination and pose variants. Comparison results have been shown in Table 1, Table 2, Table 4, Table 6 and Table 8 of Section 4 in the new version.

*4. Many experiment results are shown in Section IV. And I think it clear enough to show the good performance of proposed methods. Although, the size of test datasets seems quite small. There are many larger datasets for testing. I just suggest that it is better to show some experiment results on larger or harder test datasets.*

We have added some experiments on other two larger databases, which are the Extended Yale B database and CMU-PIE database. And the detailed experimental results are shown on Section 4.4 and Section 4.5 in the revision.

**Response to Reviewer 3**

*1. The datasets used to evaluate the algorithms are very old and small datasets, and the authors are suggested to include the results on either PIE or MULTIPIE datasets.*

Thank you for reviewing our paper. Experimental results on the CMU-PIE database have been added in the Section 4.5 in the new version.

*2. Nowadays, since deep convolutional neural network-based (DCNN) methods have achieved very impressive performances for unconstrained face recognition and other computer vision tasks, the author should discuss the advantages and the scenarios when the proposed method is better or more suitable than those state-of-the-art CNN methods to use (e.g. require small training datasets or good for mobile devices because of not requiring GPU computation.) The LBP and HOG features have a large performance gap than the DCNN-based method. Besides, it would be interesting to see the fusion method applied to the features from DCNN methods (e.g.,* *VGGFace, OpenFace) with LBP and HOG-based features used in the paper. There are no papers for related work in 2016.*

Nowadays, deep convolutional neural network-based methods have achieved very impressive performance for unconstrained face recognition and other vision tasks. Compared with deep convolutional neural network-based methods, the advantages and scenarios of our method are as follows: (1) It does not need large scale training databases, and small training databases can be used to verify; (2) GPU computation is not required; (3) It is a complex process for deep learning methods to search for suitable parameters and simpler architectures by doing many experiments. But the proposed method has not this process. So the elapsed time is reduced.

The discussion has been added in the last paragraph of Section 2 in the new version too. And the comparisons with CNN are shown in Table 1, Table 2, Table 6 and Table 8 of Section 4 in the revised version. Besides, the fusion method applied to the features from CNN with WULBP\_HOG features used in the paper is shown in Table 1, Table 2 and Table 8 of Section 4 in the revised version. Here, the CNN model we used is shown as Fig. 1. From the experiments of Section 4.1, 4.2 and 4.5, we can see that it is complex to train the CNN model for suitable parameters.



**Fig. 1** The used CNN architecture

(Note: There are 5 layers, where the first and the third layers are convolution layers, the second and the fourth layers are sub-sampling layers, and the last layer is fully-connected layer. The template size considered in convolution and sub-sampling layers are 5 × 5 and 2 × 2, respectively. The numbers of feature maps in the first and the third layers are 5 and 10, respectively. The second and the fourth layers have the same numbers of feature maps as the first and the third layers, respectively. In addition, the face images are cropped to 28 × 28 pixels and given as the input to CNN.)

Besides, we have added some papers for related work in 2016 as references.

*3. Some typos and grammatical errors  
recognize identifies -> recognize identities  
weighted sparse represent -> weighted sparse representation  
a lot of work -> a lot of works  
In term of -> In terms of*

We have gone through the paper much more times to clear the writing bugs. We have corrected these typos and grammatical errors in the new version.

**Response to Reviewer 4**

*This paper proposes to use ULBP\_MHOG features and weighted sparse representation to handle face recognition in different illumination and pose variations. My major concerns are in the following.  
  
1. It is important to discuss and compare with deep-based face recognition methods as these methods can achieve state-of-the-art performances. Please provide detailed discussion with deep-based methods and provide some comparisons with them to demonstrate the superiorness of your approach when handling illumination and pose variations.*

Thank you for reviewing our paper. The detailed discussion with deep-based face recognition has been added in the last paragraph of Section 2 in the new version. Besides, the detailed discussion and comparisons with the deep method have been shown in Table 1, Table 2, Table 6 and Table 8 of Section 4 and the last paragraph of Section 4.6 in the revised version. Here, the deep learning method we used is the CNN model, which is shown in Fig. 1. From the experiments, the advantages and scenarios of our method are shown as follows: (1) It does not need large scale training databases, and small training databases can be used to verify; (2) GPU computation is not required; (3) It is a complex process for deep learning methods to search for suitable parameters and simpler architectures by doing many experiments. But the proposed method has not this process. So the elapsed time is reduced.



**Fig. 1** The used CNN architecture

(Note: There are 5 layers, where the first and the third layers are convolution layers, the second and the fourth layers are sub-sampling layers, and the last layer is fully-connected layer. The template size considered in convolution and sub-sampling layers are 5 × 5 and 2 × 2, respectively. The numbers of feature maps in the first and the third layers are 5 and 10, respectively. The second and the fourth layers have the same numbers of feature maps as the first and the third layers, respectively. In addition, the face images are cropped to 28 × 28 pixels and given as the input to CNN.) *2. Please cite and discuss the following works. Specifcially, please perform discussions on [1-3] which are representative deep-based methods and related survey works for face recognition. Besides, since many face image modification works ([4-5]) are used which can also change the condition of faces, please also discuss whether the proposed illumination and pose variable approach can handle these modified faces (or can be extended to handle these situations).    
[1] Sun, Yi, et al. "Deepid3: Face recognition with very deep neural networks." arXiv preprint arXiv:1502.00873 (2015).  
[2] Parkhi, Omkar M., Andrea Vedaldi, and Andrew Zisserman. "Deep Face Recognition." BMVC. Vol. 1. No. 3. 2015.  
[3] Ding, Changxing, and Dacheng Tao. "A comprehensive survey on pose-invariant face recognition." ACM Transactions on intelligent systems and technology (TIST) 7.3 (2016): 37.  
[4] Y. Zhang, W. Lin, B. Zhou, Z. Chen, B. Sheng, J. Wu, "Facial expression cloning with elastic and muscle models," Journal of Visual Communication and Image Representation, vol. 25, no. 5, pp. 916-927, 2014.  
[5]* *Wang, Nannan, et al. "A comprehensive survey to face hallucination." International journal of computer vision 106.1 (2014): 9-30.*

We have cited and discussed the works ([1-3]) in the second and third paragraph of the Section 2 and the last paragraph of the Section 4.6 in the new version.

And we have cited and discussed the works ([4-5]) in the third paragraph of the Section 1 in the new version.