Facial Landmarks Detection under Occlusions via Extended Restricted Boltzmann Machine

Bing Liu[1], Ping Fu[1], Tingwei Chen[1], Yongqiang Li[1], Jeng-Shyang Pan[2,3,4]

[1] Department of Test and Control Engineering, Harbin Institute of Technology, China
[2] Department of College of Computer Science and Engineering, Shandong University of Science and Technology, China
[3] Fujian Provincial Key Laboratory of Big Data Mining and Applications, Fujian University of Technology, China
[4] Chaoyang University of Technology, Taiwan
Liubing66@hit.edu.cn, fuping@hit.edu.cn, Ianisctw1212@outlook.com,
liyongqiang@hit.edu.cn, jengshyangpan@gmail.com

Abstract

Facial landmarks encode critical information about face, which plays an important role in human communications. Accurate detecting and tracking facial landmarks have great potential value in intelligent user interfaces such as human-computer interactions. However, for face images with sever occlusions which may happen in real life such as hand occlusion, gesture occlusion and etc, detecting the facial landmarks is still a challenging problem. In this paper, we present a robust facial landmark detection method for image with occlusions based on Restricted Boltzmann Machine (RBM). We first present a face shape prior model which is constructed based on RBM to model the spatial shape patterns of the face. The detection process is accomplished by combining the prior shape model with the image measurements of facial landmarks. The low accuracy image measurements can be refined by the shape information embedded in the prior model. For the landmarks with severe occlusions, we firstly evaluate and determine the facial landmark occlusions, and replace their image measurements. The new image measurements are then fed into the prior model as evidence to predict the true locations. Evaluation on 3 databases demonstrates that the proposed method can detect facial landmarks accurately under severe occlusion, and achieved significant improvement over the current state of the art methods.

Keywords: Facial landmark detection, Restricted Boltzmann Machine, Occlusion detection

1 Introduction

Face is a powerful and immediate means for human beings to communicate their emotions, intentions, and opinions [1], and hence face related analysis based on visual modality has been investigated to boost human computer interaction (HCI) experience. Facial landmarks detection is usually the initial step and groundwork for face related analysis. Because of the great potential application value in human computer interactions, facial landmarks detection has been explored for several decades. However, most current facial landmarks detection research works are based on controlled environment where ideal illumination, high resolution images and desirable viewpoint are available. When tested under various types of real scenes, especially when there are severe face occlusions, their performance drops severely. In this paper, we introduce a robust facial landmarks detection method which can deal with severe face occlusions.

Generally, facial landmarks detection technologies could be classified into two categories: model free and model-based algorithms. Model free approaches are general purpose point detector without the prior knowledge of the object. Each landmark is usually detected individually by performing a local search for the best matching position. In contrast to model free approaches, model based methods, such as Active Shape Model (ASM) [2], Active Appearance Model (AAM) [3], etc., focus on explicitly modeling the shape of objects. The model-based methods utilize much more prior knowledge on face to realize an effective detection and achieve great performance on face images with moderate facial expression changes under slight occlusions. However, they tend to fail when there are severe occlusions on the faces in the real-world applications.

The existing research on dealing with occlusions tries to reconstruct the occluded part by two kinds of information: temporal information and spatial information. The temporal model [4] capture the dynamic information of each frame in an image sequence. The limitation for this kind method is that the query image sequence must contain non-occluded face images in the beginning, and the expressions must change smoothly. The spatial models usually employ

*Corresponding Author: Jeng-Shyang Pan; E-mail: jengshyangpan@gmail.com
DOI: 10.3966/160792642020092105001

pre-trained shape model to captures the spatial relations, either local spatial relationship among sets of facial landmarks [5] or global spatial relationship among all landmarks [6-7].

The spatial model based methods usually combine the face shape patterns with the image measurements of facial landmarks. Image measurements are usually obtained by independent detector, and are fed into the prior model as evidence. The prior model can refine the image measurements with low accuracy by the spatial information embedded in the prior model. However, for facial landmarks with severe occlusions, their measurements may be far away from the true locations, and we refer this kind the landmarks as corruptions (outliers). When there are corruptions, the optimization process can refine the corrupted landmarks measurements, but at the same time the landmarks with high accuracy may be draw away from the true locations.

In this paper, we present a new strategy for facial landmarks detection under severe occlusions which is shown in Figure 1. We first construct a face shape prior model based on Restricted Boltzmann Machine (RBM) to capture the spatial patterns of face shape. Then the facial landmark measurements are obtained by an independent detector, i.e., LEAR in [6] is used in this work. We firstly evaluate and determine the corruption landmarks caused by occlusions. Then we replace the corruption measurements by sampling results of the pre-built RBM based prior model. Eventually, the new image measurements are fed to the prior model as evidence to predict the true locations of facial landmarks. We detect 19 facial landmarks in this work as shown in Figure 2.

The remainder of this paper is organized as follows. In Section 2, we present a brief review of the related works. Then, we present the RBM based face shape prior model in Section 3. Section 4 describes the corruptions prediction and replacement method and Section 5 presents the facial landmark true locations inference method. Experimental evaluations are presented in Section 6. Section 7 concludes the paper.

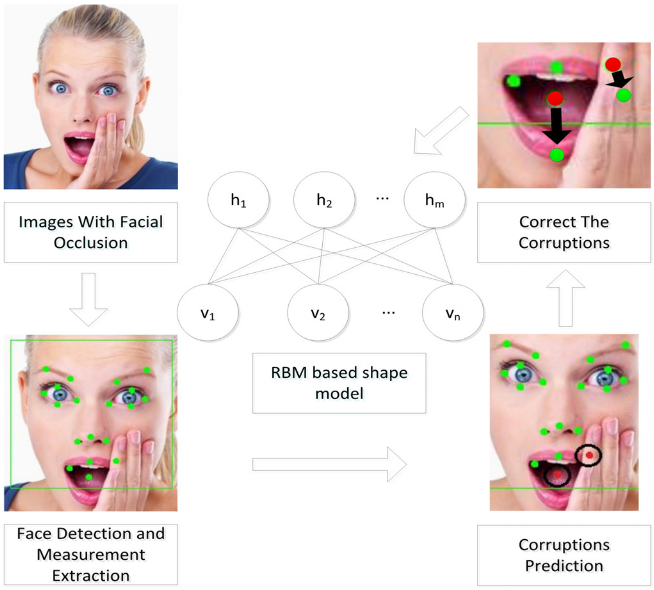


Figure 1. The flowchart of the proposed method

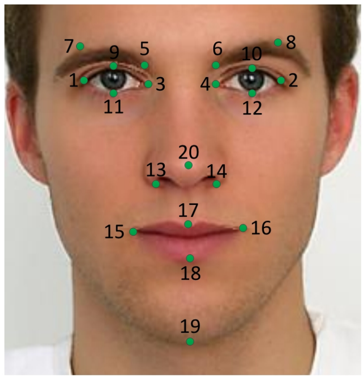


Figure 2. The 19 facial landmarks detected in this work

2 Related Work

Facial landmarks are prominent feature points surrounding facial components, which encode critical information about face shape and face shape deformation. Automatic facial landmarks detection and tracking is always a challenging problem, especially on face with occlusions [8-10]. Obviously, model free methods which search and detect each facial landmark individually fail to deal with occlusion problems. In contrast to model free approaches, statistical model-based methods, such as Active Shape Model (ASM) [2] and Active Appearance Model (AAM) [3], focus on explicitly modeling the shape of objects. The ASM model constrains the face shape to vary only in ways that have been learnt in a set of labeled training examples. The shape of a face is represented by a set of facial point coordinates. During training process, Principle Component Analysis (PCA) is employed to represent the shape variations in a linear subspace. AAM model further improves the ASM model by building linear generative models to capture both shape and appearance variations of faces.

However, with the strong assumption that the face shapes follow a Gaussian distribution, ASM can be constrained into a linear subspace. Similarly, AAM and other variations of ASM can also be modeled by a Gaussian distribution. Therefore, it is difficult for the ASM and its variations to model faces with various expressions.

In [11], Valstar and Martinez combine Support Vector Regression and Markov Random Fields to increase the accuracy and robustness of the facial landmarks detection. To handle the facial expression and pose variations, work [11] first learns mappings from the local area of a feature landmark to the true location, and then employs Markov Random Fields to constrain the search space. Nevertheless, it is still linear model that can only deal with moderate changes.

Different from holistic methods that predict the facial landmark locations from all the facial images, Huang and Liu [5] presented separate Gaussian models for every single shape component to model more detailed local shape deformations. Work [5] built a Gaussian process latent variable model in which the latent node controls the shape variations and models the nonlinear interrelationships over shape components. In addition, an illumination-robust feature and an efficient sub-window search technique make the system [5] can handle not only images with exaggerated expressions and slight shading variation but also images with occlusion and heavy shadows.

Recently, due to the nonlinear nature of Restricted Boltzmann Machines (RBM) and their variants, researchers find that this kind of methods are more suitable for capturing the large variations of objects shape. In [7], Deep Boltzmann Machines (DBM) based model is proposed to capture the face shape variations due to facial expressions for near-frontal view face. However, the face occlusions often result in gross errors or outliers which distort results detected by the shape model based methods. To handle this, a normal attempt is to employ the Bayesian inference [12] to optimize the detected landmarks. A generative model is built to maximize a posterior of observations [13]. Each landmark may have several candidate positions, and the shape is reconstructed based on their probability. [7] use the Bayesian inference to combine the face shape prior models with image measurements of facial feature points.

Another attempt is to combine the temporal information with the shape model. In [4], the facial features are tracked by incorporating temporal information through video streams and spatial information between feature points. The MRF based spatial relationship is used to bind the feature points and prevent possible drifts occurring due to occlusions. The linear temporal model captures the temporal behavior of each facial feature point. In addition, a Gabor feature based occlusion detector is developed and spatial constraints are utilized to prevent drifts because of occlusions.

As we known, for images with occlusions, face shape prior model can draw the corrupt measurements close to the true points. However, for severe occlusions, the measurements with high accuracy may be drawn away from the true locations because of optimization on the whole shape. Therefore, more recently, researchers try to qualify the measurements and just refine the corrupt measurements without changing the measurement with high accuracy. The work proposed in [6] presented a novel facial landmark detector algorithm that uses an estimation-based approach that employs Local Evidence Aggregated Regression (LEAR). Work [6] combines a regression based approach with a probabilistic graphical model-based face shape model. They propose to extend the regression based model to provide a quality measure of each prediction, and use the MRF based shape model to restrict and correct the sampling region.

3 Methods

3.1 RBM based Shape Prior Model

Facial landmark detection accuracy and robustness can be improved by incorporating the face shape prior model. Recent research shows that the Restricted Boltzmann Machines (RBM) and their variants are able to model high-order shape patterns. The RBM is a kind of undirected graphical model which contains a bipartite structure with two kinds of stochastic nodes: the visible and hidden layer nodes [14]. For RBM, there is no connection within the layer while nodes between layers are fully connected with undirected links.

In this paper, we construct a face shape prior model based on RBM (as shown in Figure 3) to explicitly capture the face shape patterns. The visible (observation) nodes $\{v_i\}$ are the coordinates of facial landmark locations, normalized according to the locations of eyes, denoted as $\mathrm{v}=\left[p_{1,x}, p_{1,y}, p_{2,x}, p_{2,y}, \ldots, p_{19,x}, p_{19,y}\right]^t$. The hidden nodes $\{h_j\}$ with binary states are connected to all visible nodes and therefore are used to model the face spatial patterns. Since the visible data are continuous, the Gaussian Bernoulli RBM (GB-RBM) [15] are employed for modeling continuous data. The total energy function of the proposed shape prior model is defined in (1):

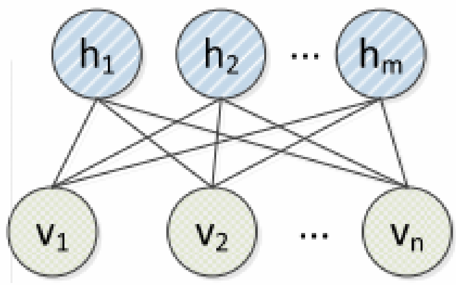


Figure 3. RBM based shape prior model

$$E\left(\mathrm{v},\mathrm{h}\middle|\theta\right)=\sum_i \frac{1}{2}\left(v_i-b_i\right)^2-\sum_{ij} W_{ij} v_i h_j-\sum_j c_j h_j \quad \textbf{(1)}$$

where $\theta=\{W,b,c\}$ are the parameters. W_{ij} measures the compatibility between visible node v_i and latent node h_j. b_i and c_j are the biases of the visible and latent units respectively.

The distribution of the facial landmarks (visible units) is calculated by marginalizing over all the hidden units with (2), where $Z(\theta)$ is the normalizing constant.

$$P(\mathrm{v}|\theta) = \frac{\sum_h e^{-E(\mathrm{v},\mathrm{h}|\theta)}}{Z(\theta)} \quad \textbf{(2)}$$

Given the training data $\{v_i\}_{i=1}^{N}$, the parameters are learned by maximizing the log likelihood with (3):

$$\theta^* = \arg\max_\theta L(\theta); L(\theta) = \frac{1}{N}\sum_{i=1}^{N}\log P(\mathrm{v}_i|\theta) \quad \textbf{(3)}$$

The gradient with respect to θ can be written in (4):

$$\frac{\partial L(\theta)}{\partial\theta} = \left\langle \frac{\partial E}{\partial\theta} \right\rangle_{p(\mathrm{h}|\mathrm{v},\theta)} - \left\langle \frac{\partial E}{\partial\theta} \right\rangle_{p(\mathrm{h},\mathrm{v}|\theta)} \quad \textbf{(4)}$$

where $\langle . \rangle_p$ represents expectation over distribution p. It is difficult to directly calculate equation (4) since it involves inferring $P(\mathrm{h},\mathrm{v})$ which is intractable. Hence we approximate it with contrastive divergency algorithm (CD) [15]. The basic idea is to approximate $P(\mathrm{h},\mathrm{v})$ with a one step Gibbs sampling from data.

3.2 Corruptions Prediction and Replacement

Given the pre-built shape prior model in the above section, we employ an independent feature point detector to get the image measurements, which can be then combined with the prior model to predict the true locations of the facial landmarks. The image measurements will be refined by the spatial patterns embedded in the shape model. However, for face images with severe occlusions, the image measurements for the occluded parts may be far away from the true locations, which are referred as corruptions. In this case, if we use the spatial relationships embedded in the pre-built face shape model to radically correct the measurements, it may draw the corrupted measurements close to the true position, but at the same time the measurements with high accuracy may be pulled away from the true positions. To overcome this problem, work [6] provided a quality measure of each prediction before correcting the measurements, and use a linear shape model (i.e., regression-based approaches) to just only restrict and correct the corruptions.

In this paper, we propose a new strategy to handle the corruption problem which can be summarized as four steps:

(1) The state of art facial landmark detector [6] is used to exact image measurements which are then used as evidence to predict the true positions. The measurements can be expressed as $\mathrm{v}_M = \left[p_{1,xm}, p_{1,ym}, p_{2,xm}, p_{2,ym}, \ldots, p_{19,xm}, p_{19,ym}\right]^t$.

(2) We propose a method to estimate the quality of the measurements, and predict the corruptions (outliers).

(3) Pre-built shape prior model is employed to re-produce measurements for corruptions found in step 2. In this way, we can get the new measurements v_{Mnew} without corruptions.

(4) The new measurements v_{Mnew} are then fed to the prior model as evidence to predict the true location: $\hat{\mathrm{v}} = \arg\max_{\mathrm{v}} P(\mathrm{v}|\mathrm{v}_{Mnew})$.

To predict the corruption, we first collect images without occlusions, and manually label the 19 facial landmarks we are going to detect. The normalized training data can be expressed as $\{\mathrm{v}_i\}_{i=1}^{N}$, which are used to train the shape prior model constructed in Sec. 3.1. Then we collect images with severe occlusions, and get the image measurements $\{\mathrm{v}_{Mi}\}_{i=1}^{M}$ for these images by the independent detector. The visible nodes of the shape prior model are initialized as v_{Mi}. Then the MCMC-based Gibbs-sampling method is used to reconstruct the measurements that fit the model. It's worth mentioning that we just do one time Gibbs-sampling to make sure the samples of visible nodes are slight deformations, but the differences for the corrupted feature points are obvious. In this way, we get the one time Gibbs-sampling results for the occluded images $\{\mathrm{v}_{Ri}\}_{i=1}^{M}$. We compute the differences between v_{Mi} and v_{Ri} which are then used to train a linear classifier to predict the occluded parts.

Once we have predicted the occluded parts of a query image, we employ the pre-built shape prior model to produce new samples to replace the image measurements for the corruptions. The samples are generated via the pre-built RBM based model by several times MCMC-based Gibbs-sampling. For each time reconstruction, we only update the visible nodes for occluded parts and the visible nodes for other parts are kept the same as the original image measurements. In this way, we replace the image measurements for the corruptions and get the new image measurements v_{Mnew}. For facial landmarks that are not occluded, the elements of v_M and v_{Mnew} are same. But for the occluded parts, we replace their measurements as the corresponding elements in v_{Mnew}. Hence there is no outliers in v_{Mnew}, and we can combine v_{Mnew} with the prior model to predict the true locations of facial landmarks.

3.3 Facial Landmark Detection Based on Prior Model

The accuracy and robustness of facial landmark detection can be improved by incorporating the face shape prior model. Given the new image measurements v_{Mnew}, facial landmark detection can be expressed as an optimization problem:

$$\hat{\mathrm{v}} = \arg\max_{\mathrm{v}} P\left(\mathrm{v} \middle| \mathrm{v}_{Mnew}\right) = \arg\max_{\mathrm{v}} P\left(\mathrm{v}_{Mnes} \middle| \mathrm{v}\right) P\left(\mathrm{v}\right) \quad \textbf{(5)}$$

$P\left(\mathrm{v}_{Mnes} \middle| \mathrm{v}\right)$ is the likelihood of the image measurements, and normally can be modeled by multivariate Gaussian distribution:

$$P\left(\mathrm{v}_{Mnes} \middle| \mathrm{v}\right) = \frac{1}{\left(2\pi\right)^{\frac{k}{2}} \left|\Sigma_l\right|^{-\frac{1}{2}}} e^{-\frac{1}{2}\left(\mathrm{v}-\mathrm{v}_{Mnew}\right)^t \Sigma_l^{-1} \left(\mathrm{v}-\mathrm{v}_{Mnew}\right)} \quad \textbf{(6)}$$

Where Σ_l is the covariance matrix that can be estimated from the training data.

$P(\mathrm{v})$ is the prior distribution of the facial landmarks modeled by the pre-built shape model based on RBM. However, it is difficult to analytically formulate $P(\mathrm{v})$ from the learned model. Hence we propose to estimate this prior probability numerically via sampling. $P(\mathrm{v})$ is also assumed as multivariate Gaussian distribution. Based on the sampled data from the shape prior model, we estimate the prior probability by calculating the mean vector μ_p and covariance matrix Σ_p. Since the likelihood probability is also multivariate Gaussian, the true locations of facial landmarks can be estimated by:

$$\mathrm{v} = \left(\Sigma_l^{-1} + \Sigma_p^{-1}\right)^{-1} \left(\Sigma_p^{-1}\mu_p + \Sigma_l^{-1}\mathrm{v}_{Mnew}^T\right) \quad \textbf{(7)}$$

4 Results and Discussion

In this section, we evaluate the proposed method on three databases, i.e., FERET database [16], extended Cohn-Kanade (CK+) [17] database and American Sign Language (ASL) database [18]. FERET database consists of gray images of human heads with frontal illumination, and with little or no expression. CK+ database contains 593 posed facial expression videos from 210 adults, among which 9% are female, 81% are Euro-American, 13% are Afro-American and 6% are from other groups. Experiments on CK+ database can evaluate the robustness of the proposed method to expression changes. However, neither FERET nor CK+ database contain face occlusions. Hence, we simulate face occlusions by adding 50×40 black masks for both FERET and CK+ databases. The mask is added to different face regions randomly for every image. To evaluate the robustness of the proposed method to spontaneous face occlusions, we also evaluate the proposed method on ASL database, which contains lots of spontaneous human gestures than cover parts of face.

Evaluation metric: For evaluation, we calculate the error as the distance between detected facial landmarks and the ground truth locations normalized by the interocular distance:

$$Error_i = \frac{\left\|P_i - \hat{P}_i\right\|_2}{D_I} \quad \textbf{(8)}$$

where P_i is the *ith* detected facial landmark, $\hat{P}_i$ is the corresponding true location, and D_I is the interocular distance for the corresponding frame.

Comparison: We are going to make a comparison between three methods: LEAR [6], a RBM based prior shape model method [7], and the propsoed method. LEAR is an independent feature point detector which use no prior information, and we denote LEAR as "Baseline method" in the experiments. Work [7] constructed a prior shape model based on RBM which models the spatial patterns of the face. We denote the method presented in [7] as "RBM Prior Model" in this paper. We implement the Baseline method and the RBM Prior Model method ourselves, and try our best to tune the parameters. And also, we fed RBM Prior Model and the proposed method with the same image measurements got by the Baseline method.

4.1 Evaluation on FERET Database

We collect 900 images of 600 subjects with frontal face from FERET database. We randomly select 600 images as training data and test on the remaining 300 images. All facial points were manually annotated. We added 50×40 black mask to 200 images of the training data which is used to train the linear classifier for occlusion prediction. All testing images are added with black mask to simulate occlusions.

The evaluation results are shown in Table 1. From Table 1 we can see that, for every part of the face, the RBM Prior model outperforms the baseline method, and the proposed method achieves the best performance. The proposed method reduce the over all detection error by 52.5% and 34.5% respectively compared to the baseline method [6] and RBM Prior Model [7]. Since there are severe occlusions for the testing images, the baseline method cannot extract effective features for occluded parts, and the ability to model spatial patterns are also limited. Hence, the baseline method only achieves an overall detection error of 6.78. Compared to the baseline method, RBM Prior model combines the image measurements with the shape prior model, which systematically models the spatial patters of the face, and therefore reduces the overall detection error to 4.63. Through the optimization process of the RBM Prior model, the landmarks of outliers can be drawn back to the true locations significantly, but at the same time, the image measurements with high accuracy may be pulled away from the true locations. For the proposed method, we first predict and replace the corruptions (outliers) caused by occlusions, and new image measurements with no outliers are fed to the prior model. In this way, the overall detection error are further reduced to 3.22 on FERET database.

Table 1. Evaluation results under occlusions on FERET database

Method	Eyebrow	Eye	Nose	Mouth	Average
Baseline method [15]	5.28	5.94	7.11	8.70	6.78
RBM Prior model [16]	3.26	3.55	5.20	6.57	4.63
Proposed method	2.28	1.97	3.37	5.25	3.22

Figure 4 demonstrates the cumulative error distribution for compared methods on FERET database. From Figure 4 we can see that the performance of the proposed method is significantly better than that the the baseline method and RBM Prior model. We also list some testing images with detected landmarks in Figure 5. From Figure 5 we can see that, for the occluded parts, the detection accuracy of the proposed method is better than that of the compared methods.

4.2 Evaluation on Extended CK Database

We also test the proposed method on CK+ database to evaluate the robustness of our model under varying facial expressions. CK+ database includes 593 sequences from 123 subjects that cover 7 basic facial expressions including anger, disgust, fear, happiness, sadness, surprise, and contempt. The image sequence varies in duration (i.e. 10 to 60 frames) and expression develops from onset (neutral frame) to peak intensity. We manually label the last five frames of each sequence to train the proposed model. We adopt leave-one-subject-out cross validation strategy, and for testing, we manually add 50×40 black mask to every query image.

The evaluation results are shown in Table 2 and Figure 6. From Table 2 we can see that, for every face region, the proposed method achieves the best performance. RBM Prior model decreases the overall detection error by 22.3% because of modeling the spatial patterns of face. In addition, the proposed method further improves the detection performance by 16.6% compared to RBM Prior model. Figure 6 shows the detection error for different facial expressions, and we can see the proposed method can decrease the detection error for all facial expressions.

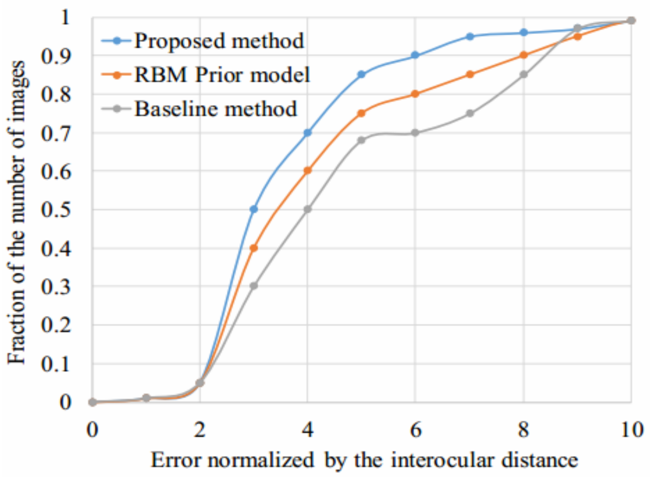
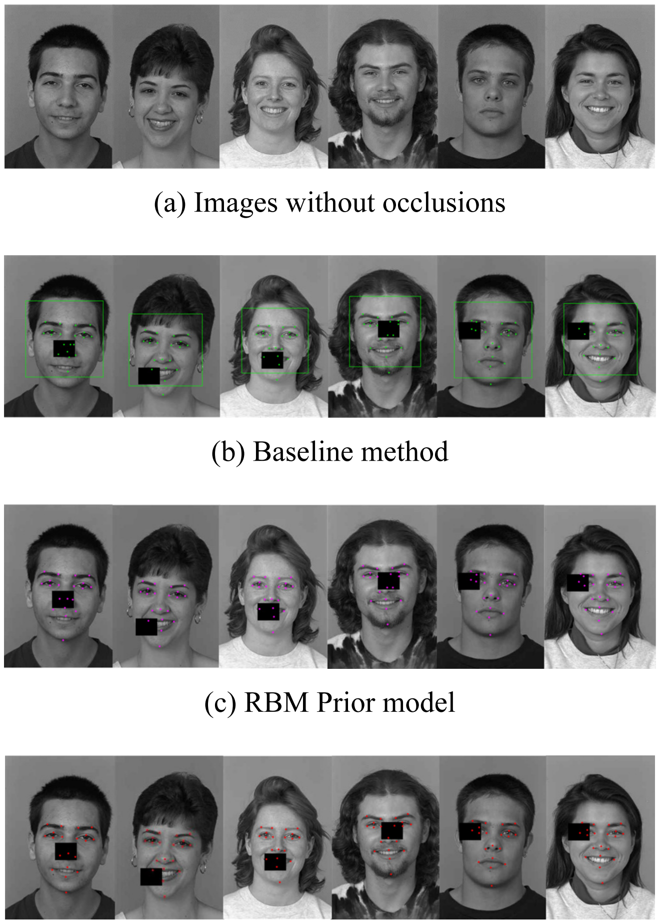


Figure 4. Cumulative error distribution for compared methods on FERET database



(a) Images without occlusions

(b) Baseline method

(c) RBM Prior model

(d) The proposed method

Figure 5. Testing image with detected landmarks on FERET database

Table 2. Evaluation results under occlusions on CK+ database

Method	Eyebrow	Eye	Nose	Mouth	Average
Baseline method [15]	8.04	7.93	8.85	9.31	8.43
RBM Prior model [16]	6.21	6.02	6.94	7.45	6.55
Proposed method	5.13	5.28	5.34	6.04	5.46

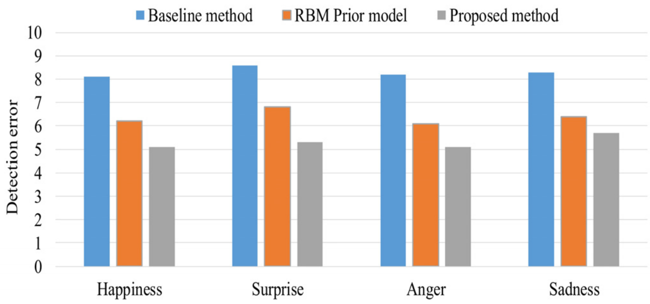


Figure 6. Detection errors for different facial expressions on CK+ database

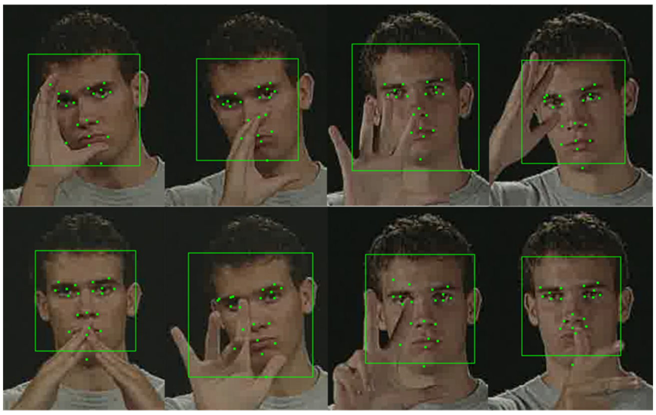
4.3 Evaluation on ASL Database

We also evaluate the proposed method on ASL database, which contains spontaneous expressions and sign language of human. These sing language accompanied with gesture often causes face occlusions. To test the generalization ability, we trained the proposed model and the compared methods on images collected from FEERA database. The experimental results are shown in Table 3.

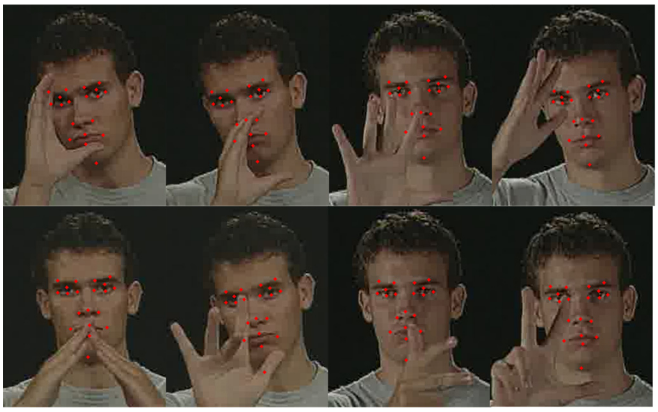
Table 3. Evaluation results on ASL database

Method	Eyebrow	Eye	Nose	Mouth	Average
Baseline method [15]	7.82	6.83	7.64	8.70	7.85
RBM Prior model [16]	5.15	4.32	5.98	7.29	5.63
Proposed method	4.53	3.39	5.39	5.87	4.82

From Table 3 we can see that, similar to cases on FEERA and CK+ database, RBM Prior model outperforms the baseline method, and the proposed method achieves the best performance on every local part of the face. RBM Prior model reduces the average detection error of baseline model from 7.85 to 5.63, and the proposed method further reduces the average detection error to 4.82. We also list some images with detected facial landmarks from ASL database in Figure 7. From Figure 7 we can see that the proposed method improves the detection results of the baseline method. And also, we expect more improvement if there are severe occlusions.



(a) Baseline method



(b) The proposed method

Figure 7. Evaluation results on ASL Database

The improvements compared to the baseline method mainly comes from two aspects: firstly, the corruptions of the image measurements are replaced which eliminates the ill effects to the optimization process. Secondly, the face spatial patters systematically embedded in the prior model can help refine the erroneous image measurements.

5 Conclusion

In this paper, we propose a prior shape model based method for facial landmark detection under occlusion. The prior shape model is constructed based on Restricted Boltzmann Machines (RBM) to model the spatial patters of the face. We propose a liner classifier to predict the corruptions (outliers) first, and then replace the corruptions based on sampling from the shape model. The new image measurements with no corruptions are fed to the prior shape model to predict the true locations of the facial landmarks. Experimental results on three databases demonstrate the superiority of the proposed method.

References

[1] M. S. Bartlett, G. Littlewort, I. Fasel, J. R. Movellan, Real Time Face Detection and Facial Expression Recognition: Development and Applications to Human Computer Interaction, *2003 Conference on Computer Vision and Pattern Recognition Workshop*, Madison, Wisconsin, USA, 2003, pp. 53-53.

[2] T. F. Cootes, C. J. Taylor, D. H. Cooper, J. Graham, Active Shape Models- Their Training and Application, *Computer Vision and Image Understanding*, Vol. 61, No. 1, pp. 38-59, January, 1995.

[3] I. Matthews, S. Baker, Active Appearance Models Revisited, *International Journal of Computer Vision*, Vol. 60, No. 2, pp. 135-164, November, 2004.

[4] S. Cosar, M. Cetin, A Graphical Model Based Solution to the Facial Feature Point Tracking Problem, *Image and Vision Computing*, Vol. 29, No. 5, pp. 335-350, April, 2011.

[5] Y. Huang, Q. Liu, D. N. Metaxas, A Component-based Framework for Generalized Face Alignment, *IEEE Transactions on Systems, Man, and Cybernetics*, Part B (Cybernetics), Vol. 41, no. 1, pp. 287-298, February, 2011.

[6] B. Martinez, M. F. Valstar, X. Binefa, M. Pantic, Local Evidence Aggregation for Regression-based Facial Point Detection, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 35, No. 5, pp. 1149-1163, May, 2013.

[7] Y. Wu, Z. Wang, Q. Ji, Facial Feature Tracking under Varying Facial Expressions and Face Poses Based on Restricted Boltzmann Machines, *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Portland, OR, USA, 2013, pp. 3452-3459

[8] Y. Yu, S. Zhang, F. Yang, D. Metaxas, Localization of Multi-pose and Occluded Facial Features Via Sparse Shape Representation, *International Symposium on Visual Computing Advances in Visual Computing*, Rethymnon, Crete, Greece, 2013, pp. 486-495.

[9] M. Xu, B. Liu, P. Fu, J. Li, Y. H. Hu, Video Saliency Detection Via Graph Clustering with Motion Energy and Spatiotemporal Objectness, *IEEE Transactions on Multimedia* (Early Access), pp. 1-1, May, 2019.

[10] M. Xu, B. Liu, P. Fu, J. Li, Y. H. Hu, S. Feng, Video Salient Object Detection Via Robust Seeds Extraction and Multi-graphs Manifold Propagation, *IEEE Transactions on Circuits and Systems for Video Technology* (Early Access), pp. 1-1, June, 2019.

[11] M. Valstar, B. Martinez, X. Binefa, M. Pantic, Facial Point Detection Using Boosted Regression and Graph Models, *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Francisco, CA, USA, 2010, pp. 2729-2736.

[12] Y. Zhou, L. Gu, H. J. Zhang, Bayesian Tangent Shape Model: Estimating Shape and Pose Parameters Via Bayesian Inference, *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Madison, WI, USA, 2003, pp. 109-116.

[13] L. Gu, T. Kanade, A Generative Shape Regularization Model for Robust Face Alignment, *European Conference on Computer Vision*, Marseille, France, 2008, pp. 413-426.

[14] P. Smolensky, Information Processing in Dynamical Systems: Foundations of Harmony Theory, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Vol. 1, MIT Press Cambridge, 1986, pp. 194-281.

[15] G. E. Hinton, Training Products of Experts by Minimizing Contrastive Divergence, *Neural Computation*, Vol. 14, No. 8, pp. 1771-1800, August, 2002.

[16] P. J. Phillips, H. Moon, S. A. Rizvi, P. J. Rauss, The Feret Evaluation Methodology for Face-recognition Algorithms, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, no. 10, pp. 1090-1104, October, 2000.

[17] J.-S. Pan, C.-Y. Lee, A. Sghaier, M. Zeghid, J. Xie, Novel Systolization of Subquadratic Space Complexity Multipliers Based on Toeplitz Matrix-Vector Product Approach, *IEEE Transactions on Very Large Scale Integration Systems*, Vol. 27, No. 7, pp. 1614-1622, July, 2019.

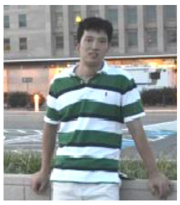
[18] A. M. Martinez, R. B. Wilbur, R. Shay, A. C. Kak, Purdue RVL-SLLL ASL Database for Automatic Recognition of American Sign Language, *IEEE International Conference on Multimodal Interfaces*, Pittsburgh, PA, USA, 2002, pp. 167-172.

Biographies

Bing Liu received his B.S., M.S. and Ph.D. degree in electrical engineering from Harbin Institute of Technology, Harbin, China. He is currently with Department of Test and Control Engineering, Harbin Institute of Technology, China. His current research interests include computer vision and high-performance computing.

Ping Fu receieved the B.S. at University of Science and Technology of China, Hefei, China, the M.S. and Ph.D. at Harbin Institute of Technology, Harbin, China. He is currently with Department of Test and Control Engineering, Harbin Institute of Technology, China. His research interests span Digital Image Processing, Computer Vision.

Tingwei Chen is currently pursuing his M.S degree at Harbin Institute of Technology, Harbin, China. His research area lies in Computer vision.



Yongqiang Li receieved his B.S., M.S., and Ph.D. from Harbin Institute of Technology, Harbin, China. He is currently with School of Electrical Engineering and Automation, Harbin Institute of Technology, China. His research interests lie in machine learning, Environmental awareness and automatic measurement and control technology.

Jeng-shyang Pan received the B.S. degree in electronic engineering from the National Taiwan University of Science and Technology, Taipei, Taiwan, the M.S. degree in communication engineering from the National Chiao Tung University, Hsinchu, Taiwan, and the Ph.D. degree in electrical engineering from the University of Edinburgh, Edinburgh, U.K.. He is currently with the Department of College of Computer

Science and Engineering, Shandong University of Science and Technology, Qingdao, China; with the Fujian Provincial Key Laboratory of Big Data Mining and Applications, Fujian University of Technology, Fuzhou, China; and with Chaoyang University of Technology, Taichung, Taiwan. His current research interests include cryptographic computing, robot vision, and cloud computing.