IOP Conf. Series: Materials Science and Engineering **231** (2017) 012021 doi:10.1088/1757-899X/231/1/012021

The review and results of different methods for facial recognition

Yifan Le*

*School of Physics and Electronics, Central South University, Hunan 410083, China

E-mail: yifan_le@163.com

Abstract. In recent years, facial recognition draws much attention due to its wide potential applications. As a unique technology in Biometric Identification, facial recognition represents a significant improvement since it could be operated without cooperation of people under detection. Hence, facial recognition will be taken into defense system, medical detection, human behavior understanding, etc. Several theories and methods have been established to make progress in facial recognition: (1)A novel two-stage facial landmark localization method is proposed which has more accurate facial localization effect under specific database; (2)A statistical face frontalization method is proposed which outperforms state-of-the-art methods for face landmark localization; (3)It proposes a general facial landmark detection algorithm to handle images with severe occlusion and images with large head poses; (4) There are three methods proposed on Face Alignment including shape augmented regression method, pose-indexed based multi-view method and a learning based method via regressing local binary features. The aim of this paper is to analyze previous work of different aspects in facial recognition, focusing on concrete method and performance under various databases. In addition, some improvement measures and suggestions in potential applications will be put forward.

1. Introduction

In the recent years, intelligent security technology has stepped into a higher platform with the face recognition technology (FRT) standing out among variety of biometric technologies, which is friendly, convenient and efficient. For instance, an advanced face recognition technology could be used in monitoring in complex environment like railway stations, airport, and bank, street to gain first-hand images for reference if accident happens. According to Alibaba's CEO Jack Ma, the FRT will be soon applied in financial use with the name "Smile to pay", offering safety and speed.

In this paper, we will first analyze previous work of four aspects in facial recognition technology, focusing on both improvements than before and their limitation in practical. The brief analysis will be shown as following:

(1) Wei et.al [1] proposed a two-stage facial landmark localization method based on a previous method which integrated face detection and facial landmark localization into one framework. A Supervised Descent Method (SDM) is added to refine the locations of facial features. The experimental results prove that the proposed method performs better with more accurate facial localization on CK+ datasets.

(2) A new Robust Statistical Face Frontalization method by Sagonas et.al [2] is proposed. It use only a small set of frontal images to jointly achieve face frontalization, landmark localization and pose-invariant face recognition, with the point that the rank of a frontal facial image is much smaller

than those in other poses. The RSFC is proved to outperform state-of-the-art methods for face landmark localization under several Databases even CAT.

(3) In Robust Facial Landmark Detection under Significant Head Poses and Occlusion [3], Wu et.al proposed a general facial landmark detection algorithm especially used in severe occlusion and images with large head poses. The occlusion prediction is handled in one unified framework and facial landmark detection with occlusion is treated differently on visibility probabilities of concrete points. It assumes all the landmarks visible at the beginning and the algorithm is optimized to achieve convergence by updating the visibility probabilities and the landmark locations across iterations while the prior occlusion pattern is set as the constraint. This method has been proved to achieve its purpose on both images with severe occlusion and large head poses and less challenging ones.

(4) There are three methods for face alignment shown in this paper—shape augmented regression method, pose-indexed based multi-view method and a learning based method via regressing local binary features: 1. The shape augmented face alignment method [4] adds the shape information more than the appearance features which results in that regression prediction function will change according to different face shapes. 2. The pose-indexed based multi-view (PIMV) face alignment method [5] first establish the pose-indexed shape searching space by a series of pose-shape pairs corresponding to face shapes. And the estimated pose generated by the multi-view generative model sets as the initialization in following iterative stages. Then the Local binary features are extracted from the original image, and PIMV will use repressors to predict the shape increment until convergence. 3. The learning based method [6] contains a locality principle to learn a set of local binary features, which focuses on the most discriminative texture information around the estimated landmark from the previous stage and the information of the shape context and the local texture of the landmark.

And as a result, all three methods mentioned above to optimize face alignment have made progress comparatively on robustness, accuracy, speed and other performance with state-of-the-art method. In a nutshell, those theories and methods mentioned above represent the significant leaps on the promising area, facial recognition technology, with the achievements including not only on speed, accuracy, universality, stability. However, doubts and limitations still exist like insufficient testing databases, universality on all images, latency due to complicate algorithm and stability under hard condition. Furthermore, we will discuss more about the application foundation and future development of facial recognition technology based on those work, which contains: (1) the possible applications of facial recognition technology on several direction in the real world; (2) the potential usage of the specific algorithm or theories produced during the development of RSF; (3) the direction and focus of those methods mentioned above in the future development.

The overall structure of this paper comprise two parts, Section II and Section III. Section II gives detailed description, experiments results and our evaluation of two improvements on Facial landmarks detection and Facial alignment, as well briefly introducing another two achievements on facial landmark localization and facial frontalization. In Section III, a conclusion of the literature review will be presented.

2. Methods and Results of Recent Advance in Face Recognition Technologys

In this paper, we have divided recent advance in face recognition technology into four aspects: (1) face alignment, (2) facial landmark localization, (3) face frontalization, (4) facial landmark detection. And then this paper will focus on four specific novel method corresponding to four aspects mentioned above, including the core of method, experiments results and evaluation from various angles. Two of them will be introduced with details like algorithm while it will directly come into conclusion for another two.

2.1. Facial Landmark Detection Under Severe Head Poses and Occlusion

Facial landmark detection algorithm contains three major categories: the holistic methods, the constrained local methods and the regression based methods. And the algorithm we will discuss about belongs to regression based methods, the most promising one recently.

2.1.1. Method

The algorithm [3] is supposed to find out the mapping from image I to landmark locationsx $= R^{2D_1}$, where D_1 is the number of facial landmarks. Different from others, the landmark visibility probability variable $P \in [0,1]^{D_1}$ are used to handle images with severe occlusion and large head poses. When updating the visibility probabilities, we denote f_t as constrained supervised regression model to predict the landmark visibility probability update Δp^t based on the image, x^{t-1} as the current landmark locations, Loss (.) as the loss function of occlusion pattern. When updating the landmark locations, we denote g_t as regression function to predict the landmark location update Δx^t, p^t as the visibility probabilities, and x^{t-1} as the current landmark locations. The general framework is shown as follows:

Initialize the landmark locations x^0 using the mean face.
Assume all the landmarks are visible $p^0 = 1$
For t=1, 2, T or convergence do
Update the landmark visibility probabilities given the images, the current landmark location, and the occlusion pattern$\mathrm{LOSS}(.)$.
f_t: I, x^{t-1}, $\mathrm{LOSS}(.)$→Δp^t
$p^t = p^{t-1} + \Delta p^t$
Update the landmark locations given the images, the current landmark locations, and the landmark visibility probabilities.
g_t: I, x^{t-1}, p^t→Δx^t
$x^t = x^{t-1} + \Delta x^t$
End

Then it will output the estimated landmark locations x^T and the binary occlusion vector based on the predicted visibility probabilitiesp^T.

The procedure [3] to update the landmark visibility probability is based on the appearance and shape information from all points with the learned explicit occlusion pattern as a constraint. First it denotes SIFT features of the local patches centered at the current landmark locations as $\emptyset\ (I, x^{t-1}) \in R^{D_1 D_f}$, where $D_f = 128$ is the dimension of features. Then it gets the shape features denoted as$\varphi\ (x^{t-1}) \in R^{D_1(D_1-1)}$ and a concatenated feature vector denoted as$\psi\ (I, x^{t-1}) = [\emptyset\ (I, x^{t-1});\ \varphi\ (x^{t-1})]$. The landmark visibility probabilities p^t will be update for the next iteration ($E_{p^t}[\mathrm{LOSS(c)}]$ represents the expectation):

Minimize Δp^t

$$\|\Delta p^t - T^t \psi\ (I, x^{t-1})\|_2^2 + \lambda E_{p^t}[\mathrm{LOSS(c)}]$$

$$p^t = p^{t-1} + \Delta p^t$$

$$0 \le p^t \le 1$$

Subject to $E_{p^t}[\mathrm{LOSS(c)}] = \sum_{k=1}^{2D_l} Loss(c_k) P(c_k, : p^t)$

$$p(c; p) = \prod_{d-1}^{D_l} p_d^{c_d} (1 - p_d)^{1-c_d}$$

After landmark visibility prediction, the method [3] will use the Autoencoder model to learn the loss function $\mathrm{LOSS}(.)$. The parameters$\theta = \{W_1, b_1, W_2, b_2\}$ of the autoencoder model are based on the feasible landmark occlusion label from the real training date and the synthetic data. Then it could get:

The minimized reconstruction errors

$$\theta^* = \mathrm{argmin} \sum_i \|c_i - \sigma(W_2 \sigma(W_1 c_i + b_1) + b_2)\|_2^2$$

The loss function

$$\mathrm{LOSS}(\mathrm{c}; \theta) = \|c - \sigma(W_2 \sigma(W_1 c + b_1) + b_2)\|_2^2$$

And it uses standard least squares formulation to estimate the linear regression with parameter T^t in each iteration:

$$T^{t^*} = \operatorname{argmin} \sum_i \left\| \Delta p^t - T^t \psi\left(I_i, x_i^{t-1}\right) \right\|_2^2$$

To estimate Δp^t given the appearance and shape features $\psi\,(I, x^{t-1})$, the currently estimated visibility probabilities p^{t-1}, model parameter T^t, and the loss function LOSS(.), it uses Monte Carlo approximation and gradient descent algorithm:

$$E_{p^t}[\mathrm{LOSS(c)}] \approx \frac{const}{K} \sum_{k=1}^{K} \mathrm{LOSS}(\widetilde{c_k})\, p^t(\widetilde{c_k})$$

$$const = 2^{D_l}$$

$$\delta = 2\left(\Delta p^t - T^t \psi\,(I, x^{t-1})\right) + \lambda \frac{const}{K} \sum_{k=1}^{K} \mathrm{LOSS}(\widetilde{c_k})\, p^t(\widetilde{c_k})$$

$$\Delta p^t \in [0,1]^{D_1}$$

2.1.2. Results of experiments

(1) Images with severe occlusion and large head poses

With images from COFW database with severe occlusion, the proposed method [3] shows close performance with human on landmark detection error, which is better than all the other state-of-the-art works like CRC, OC, SDM, etc. While on occlusion prediction, it also has comparable precision and better result on recall values. With images from FERET database with large head poses, the proposed method is better than the FPLL algorithm and Pose-free algorithm with hardly wrong estimated pose which could result in large facial landmark detection error.

(2) General "in-the-wild" images

In this part, the experiment [3] uses images from Helen database and LFPW database. And the result shows that the proposed method achieves the least ratio of detection error among all the state-of-the-art works taking test. In addition, the speed of the proposed algorithm is comparable to others especially for the model without the explicit occlusion pattern.

2.1.3. Conclusion and Evaluation

The core of the whole method is updating the landmark visibility probabilities and landmark locations through iterations. To be more specific, it uses one unified model to deal with occlusion with prior occlusion pattern as the constraint and it rely more on the information from points with high visibility probabilities for better landmark detection. And the experimental results have proved its performance much better than state-of-the-art works while handling images with severe occlusion and images with large head poses. When it comes to general images, it is able to keep its robustness and efficiency comparable with other works. However, it remains a problem to solve: the strategy that treating points differently will result in latency of the whole facial detection system due to the larger amount of calculation, which might influence its performance in real-time tracking. Hence, in the future, the proposed method should improve its algorithm for real-time tracking and apply it into more complicate images like other primate or even cats with real world conditions like significant illumination change or low resolution. Here, we put forward to combine probabilistic random forest (PRF) and the algorithm mentioned above to achieve better performance. The random forest are useful for face aligning via shape regression, which has been proved well fit for locating facial landmarks in[7]. From experiments results [3], application of PRF is able to improve system accuracy when facing the pictures under variation of conditions. And it would help the proposed algorithm to achieve real-time tracking.

2.2. Face Alignment By Regressing Local Binary Features

2.2.1. Method

For general shape regression method [6], a shape increment ΔS^t at stage t could be regressed as: $\Delta S^t = W^t \emptyset^t \ (I, S^{t-1})$, where I is the input image, S^{t-1} is the shape from the previous stage, $\emptyset^t$ is a feature mapping function, and W^t is a linear regression matrix. And this method [6] would focus on regularization for learning local binary features $\emptyset^t$ and regularization for learning global linear regression W^t (Locality principle is the core of the theory).

Firstly, the feature mapping function is composed of a set of ones: $\emptyset^t = [\emptyset_1^t, \emptyset_2^t, \dots, \emptyset_L^t]$; the ground truth shape increment:

$$\Delta \hat{S}^t = \min_{w^t, \emptyset_l^t} \sum\nolimits_{i=1} \left\| \pi_l o \Delta \hat{S}_i^t - w_l^t \emptyset_l^t (I_i, S_i^{t-1}) \right\|_2^2$$

Where $\pi_l o \Delta \hat{S}_i^t$ is the ground truth 2D-offset of *l*th landmark in *i*th training sample. A standard regression random forest where the split nodes in the tree are trained using pixel-difference feature, is used to learn each local mapping function$\emptyset_l^t$. Then each leaf node could gain a 2D offset vector represents average of all the training examples. While testing, the output could be written as "$w_l^t \emptyset_l^t (I_i, S_i^{t-1})$ ", which is the summation of the output stored in those leaf nodes the sample traverses through. And the value is 1 if the sample reaches the corresponding leaf node and 0 otherwise for each dimension in $\emptyset_l^t$ so that it could be called as local binary features.

Secondly, the method [6] would get another W_1^t rather than using the local regression output w_l^t by the following:

$$\min_{W^t} \sum\nolimits_{i=1}^{N} \left\| \Delta \hat{S}_i^t - W_l^t \emptyset^t \ (I, S^{t-1}) \right\|_2^2 + \lambda \| W^t \|_2^2$$

where the first term is the regression target and the second is the regularization on W^t.

In addition, a dual coordinate descent method [6] is used to deal with the highly sparse linear system.

2.2.2. Results of experiments

Datesets [6]: LFPW (29 landmarks), Helen (194 landmarks), 300-W (68 landmarks).

When comparing accuracy with state-of-the-art methods, the proposed method (LBF) [6] has the best performance on all datasets tested and similarly the faster version (LBF fast) is comparable with the best previous one. What's more, the LBF complicate its achievement on the challenging IBUG subset with significant error reduction. When comparing speed with state-of-the-art methods, the LBF shows obvious superiority comparing with ESR and SDM especially the LBF fast could reach tens of times fast of others, even thousands of FPS.

2.2.3. Conclusion and Evaluation

By experiments results, the writer has proved that the proposed method by learning local binary features could achieve fast face alignment with comparable accuracy with the state-of-the-art methods. From other angles, it has several advantages as follows:

(1) It could achieve real-time face alignment faster than those on mobile phone;

(2) The sparse binary features result in high speed of thousands of FPS;

(3) The LBF reduces error successfully for the good generalization ability including on challenging database;

(4) In this method, the pixels are indexed only in a local region rather than over the global shape so that it is able to find much better features;

(5) High dimensional binary features take place of the local random forest' regression output of each landmark as features, which could remain the completed information to reach better performance.

(6) The standard regression random forest could result in better geometrical invariability with relatively lower learning cost.

However, there are problem existing:

(1)The learning local binary features and function of the learning global linear regression are based on specific quantity of training database so that it could be too average and simple while facing much more difficult head poses and complicate background.

(2) The proposed method focus on dealing with shape increment based on an initial shape so that the selection of initial shape is significant in the whole regression. And it is supposed to prove its good selection of initial shape from images with severe occlusion and large poses.

(3) The whole regression uses feature mapping function and linear regression matrix so that its accuracy remains doubts.

(4) It only consider candidate features in a local region rather than the global face image, which improves the efficiency but increases mistakes at the same time.

Last but not the least, we would put forward some suggestions and expectations for the proposed method:

(1) The concept of locality principle could be used in other relevant areas like anatomic structure segmentation and human pose estimation as the author thought.

(2) The proposed method still worth improving refitting strategy when regression trees applied in other scenarios.

(3) Since the proposed method could reach the required FPS to apply in mobile phones, it actually make sense to concentrating on optimization on real-time tracking.

(4) Aiming at reduction of error from the initial shape selection, we proposed that the concept of multiple initialization could be used to optimize the selection of initial shape for better accuracy. The method that has been proposed in [8] uses several different poses with different ranks by probability as initialization for best estimation.

2.3. A Two-stage Facial Landmark Localization Method

The novel method [2] draws on the experience of a facial analysis method to integrated face detection and facial landmark localization into one framework while face images would be divided into different expressions, each of which has a certain configuration. Except it, the authors added SDM to make facial features more accurate. To be specific, the proposed method uses multi-tree model with a shared pool of parts to handle multi-expression and supervised descent method (SDM) are used to refine facial landmarks in iterations.

Experiment results have shown that the proposed method make sense in improving accuracy of facial localization with several features: face detection, facial landmarks localization and face expression location can be handled at the same time (using multi-tree model);it introduces SDM to refine the location of facial features, which results in more accurate localization (close to ground truth landmarks). However, it still need to take more tests on other challenging database rather than only CK+ datasets before applied in system widely.

2.4. A Robust Statistical Face Frontalization

The author in [2] proposed a novel and robust face frontalization. It is worth mentioning that it use only a small set of frontal images to jointly achieve face frontalization, landmark localization and pose-invariant face recognition, with the point that the rank of a frontal facial image is much smaller than those in other poses in a linear space.

The RSFC is proved through sufficient experiments [2] to outperform state-of-the-art methods for face landmark localization on Databases of LFPW, HELEN, AFW, LEW, FERET, Multi-PIE, and FS, even CAT. And the method brings much more contributions than a novel RSF method on both technology and applications in computer vision, such as a developed algorithm for the RSF, the first generic landmark localization method using only a model of frontal images with considerable consequence.

3. Conclusion

This paper mainly introduces several previous work on face recognition technology about detailed algorithm, experiment results and analytical evaluations. It could be obviously concluded from the four advanced achievements of corresponding aspects that FRT has been significantly improved on accuracy, robustness, speed, processing capacity on challenging images, etc. But before widely application in mature system, it is supposed to make improvements on real-time tracking, stability under changing conditions with less reliance on limited training objects.

References

[1] Wei, W., Tian, C., & Zhang, Y. (2015). A two-stage facial landmark localization method. International Conference on Orange Technologies(pp.157-160).
[2] Sagonas, C., Panagakis, Y., Zafeiriou, S., & Pantic, M. (2015). Robust Statistical Face Frontalization. IEEE International Conference on Computer Vision (pp.3871-3879). IEEE.
[3] Wu Y, Ji Q. Robust Facial Landmark Detection Under Significant Head Poses and Occlusion [C] // IEEE International Conference on Computer Vision. IEEE, 2015:3658-3666.
[4] Wu, Y., & Ji, Q. (2015). Shape Augmented Regression Method for Face Alignment. IEEE International Conference on Computer Vision Workshop (pp.979-985). IEEE.
[5] Qi H, Zhao Q, Wang X, et al. Pose-indexed based multi-view method for face alignment [C] //Image Processing (ICIP), 2016 IEEE International Conference on. IEEE, 2016: 1294-1298.
[6] Ren, S., Cao, X., Wei, Y., & Sun, J. (2014). Face Alignment at 3000 FPS via Regressing Local Binary Features. IEEE Conference on Computer Vision and Pattern Recognition (Vol.25, pp.1685-1692). IEEE Computer Society.
[7] Luo, C., Wang, Z., Wang, S., & Zhang, J. (2015). Locating facial landmarks using probabilistic random forest. Signal Processing Letters IEEE, 22(12), 2324-2328.
[8] Chai, X., Wang, Q., Zhao, Y., & Li, Y. (2016). Robust facial landmark detection based on initializing multiple poses. , 13(5).