Local-Global Landmark Confidences for Face Recognition

KangGeon Kim[1], Feng-Ju Chang[1], Jongmoo Choi[1], Louis-Philippe Morency[2]
Ramakant Nevatia[1] and Gérard Medioni[1]
[1] Institute for Robotics and Intelligent Systems, University of Southern California, CA, USA
[2] Language Technologies Institute, Carnegie Mellon University, PA, USA

***Abstract*— A key to successful face recognition is accurate and reliable face alignment using automatically-detected facial landmarks. Given this strong dependency between face recognition and facial landmark detection, robust face recognition requires knowledge of when the facial landmark detection algorithm succeeds and when it fails. Facial landmark confidence represents this measure of success. In this paper, we propose two methods to measure landmark detection confidence: local confidence based on local predictors of each facial landmark, and global confidence based on a 3D rendered face model. A score fusion approach is also introduced to integrate these two confidences effectively. We evaluate both confidence metrics on two datasets for face recognition: JANUS CS2 and IJB-A datasets. Our experiments show up to 9% improvements when face recognition algorithm integrates the local-global confidence metrics.**

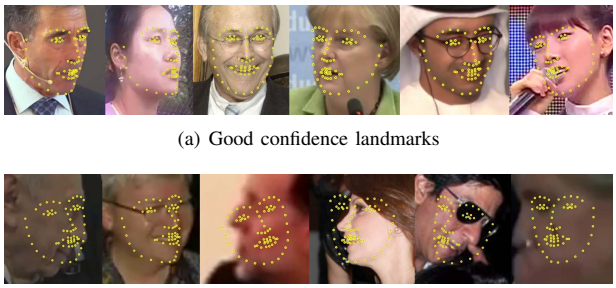
I. Introduction

Facial landmark detection is an essential preliminary step for many facial analysis tasks such as face recognition [26]. Accurate facial landmarks are needed in order to align faces between images, improving robustness of face recognition. When face alignment is performed with poorly detected facial landmarks, face recognition is likely to fail. Zhang *et al*. [26] shows the importance of alignment by proposing pose-aligned network.

The problem of automatic facial landmark detection has seen much progress over the past years [6], [2], [23], [25], [27], [12]. However, few studies focus on measuring the confidence of detected landmarks. Fig. 1 shows an example of good and bad landmarks. If we can predict the confidences of the landmarks, it would benefit face recognition.

In this paper, we propose a new method for measuring landmark confidence. One is the constrained local method, and the other is rendering based global method. The local method can measure accuracy based on local predictors of each facial landmark, and the global method can predict the confidence from 3D rendered faces. While the local confidence measures the goodness of landmarks based on the 2D aligned images, the global confidence measures it from the 3D rendering point of view. The merit of using these two confidences is they can help alleviating the influences of the poorly aligned face images on face recognition so as to improve recognition accuracy.

This research is based upon work supported in part by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via IARPA 2014-14071600011. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purpose notwithstanding any copyright annotation thereon.



(a) Good confidence landmarks

(b) Poor confidence landmarks

Fig. 1. *Example landmark detection results.* How do we determine whether detected landmarks are good or not?

Fig. 2 illustrates how the proposed confidences are integrated in the face recognition framework. In this framework, we match not only the face images with 2D alignment (affine transformation) but also the cropped ones (without alignment) given the bounding boxes from the ground truth annotation or face detector. Namely, each image pair (one is from the gallery set and the other from the probe set) is associated with two matching scores (denoted as Score 1 and Score 2 in Fig. 2). While the 2D-aligned images are used to measure the local landmark confidence, the 3D-aligned ones are employed to obtain the global confidence. In order to determine if this face image pair comes from the same person or not, we propose an approach to integrate local-global confidence using the nonlinear transform fusion. In this way, we can determine the fusion weights effectively for both matching scores and enhance the face recognition accuracy.

In the following sections, we first present the work related to our approach. In Section III, we describe our local-global confidence metrics and their application on face recognition. We follow this with the experiments (Section IV) and results (Section V) to evaluate the effectiveness of our landmark confidences and score fusion method for face recognition on JANUS CS2 dataset and IJB-A dataset [13]. Finally, we conclude our work in Section VI and propose future directions.

978-1-5090-4023-0/17 $31.00 © 2017 IEEE
DOI 10.1109/FG.2017.84

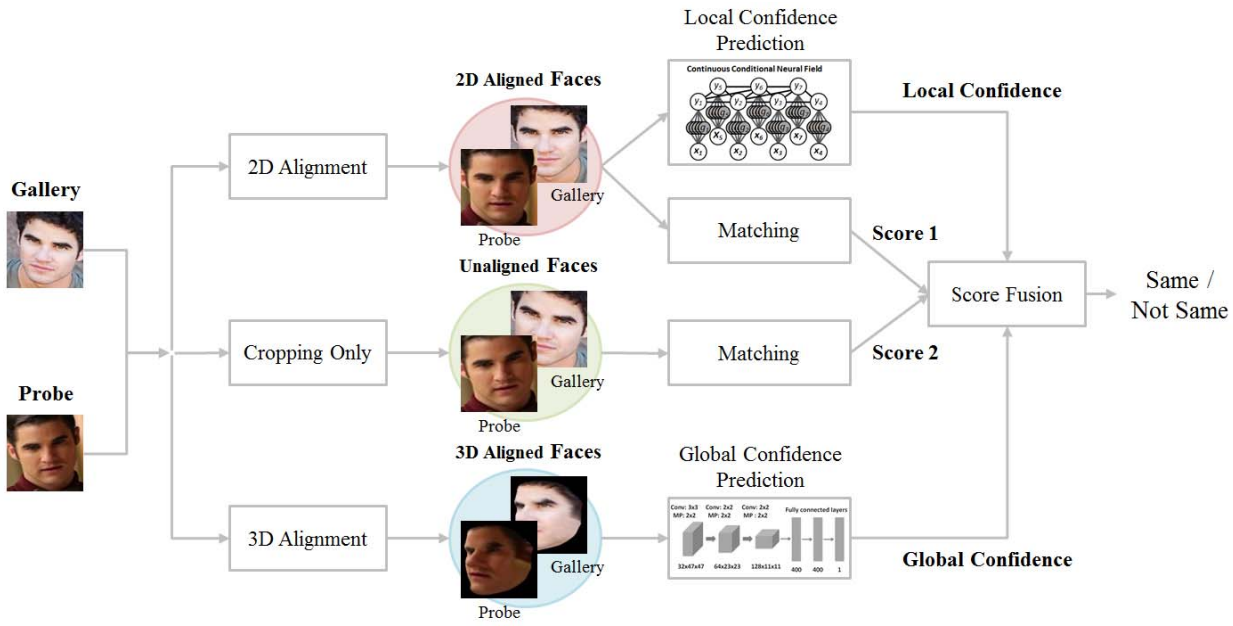


Fig. 2. *The Overview of our face recognition framework and the way of generating two types of landmark confidences.* The probe and gallery images are from JANUS CS2 dataset. The fusion of matching scores from 2D Alignment and cropping only is achieved with the help of landmark confidences.

II. Related Work

There have been various studies on facial landmark detection. Their approaches can be divided into two ways.

The first method is based on a cascaded regression approach which directly applies regressor into appearance in order to estimate landmarks [6], [23], [27]. Recently, there have been several studies done using deep learning techniques, such as Zhang *et al*. [25], and Sun *et al*. [22].

The second method is to extract appearance descriptor, and compute local response map, then fit the shape model based on the local predictions. The most representative methods are the Constrained Local Model (CLM) [20], Constrained Local Neural Fields (CLNF) [2] and Discriminative Response Map Fitting (DRMF) [1]. Rajamanoharan and Cootes [18] and Kim *et al*. [12] proposed the more robust method on large pose variations.

Despite of many studies on landmark detection, there are few studies on confidence which evaluate landmark confidence. Steger *et al*. [21] proposed failure detection method for facial landmark detector and this method allows recovery from failures. Steger *et al*. [10] proposed landmark localizations quality assessment using a regression to the Area Under the Point Accuracy Curve (AUPAC) which defined for the trade-off curve between the distance threshold and the obtained recall. Confidence is important in various aspects: it allows us to improve the landmark detection, furthermore, it can be used in later pipeline by weighting on the scores.

Score Fusion for Face Recognition Face recognition is usually achieved by matching one face image to another, and assigning a matching score to show the possibility these two face images are from the same person or not. It is often the case that we have multiple matching scores due to multiple medias for a face (e.g., 2D and 3D) or (and) multiple matching methods (e.g. Euclidean distance or inner product) or (and) multiple features (e.g. CNN features learned from aligned images or cropped images as in Fig. 2) for example. Therefore, we need to obtain a single score for the considered matching pair by score fusion. The most widely used fusion methods are the average, maximum or minimum fusion [19], [5]. In [15], the ensemble SoftMax method is shown to be effective on fusing the scores from matching face images of different poses. Recently, the quality based fusion approaches have been proposed [5], [17] to integrate the quality of the score into fusion. In [5], the qualities of individual media are directly utilized as the weights to combine five types of media: 2D face image, video, 3D face model, sketch, and demographic information. However, the media quality itself might be noisy and a better transformation from it to the fusion weight is needed. [17] proposes a unified Bayesian framework that incorporates the quality information elegantly for multimodal biometric fusion. One of the drawbacks of this approach is that we need the accurate probability estimation of class labels given the features and qualities. In this work, we fuse the matching scores from 2D alignment and cropping only by incorporating the landmark confidence, which is transformed to the fusion weight by a sigmoid function.

III. Landmark Confidences and Their Application on Face Recognition

In this section, we introduce our confidence model for facial landmark detection. First, we start by describing a con-

strained local confidence model (Section III-A). We follow this by a description of global confidence model in Section III-B.

A. Constrained Local Confidence

Our constrained local confidence builds from the Constrained Local Neural Field (CLNF) [2] deformable model. We compute the facial landmark by using CLNF. For a given set of k facial landmark positions $\mathbf{x} = \{x_1, x_2, ..., x_k\}$, our model defines the likelihood of the facial landmark positions conditioned on image $\mathcal{I}$ as follows:

$$p(\mathbf{x}|\mathcal{I}) \propto p(\mathbf{x}) \prod_{i=1}^{k} P(x_i|\mathcal{I}). \tag{1}$$

In Equation (1), $p(\mathbf{x})$ is a prior distribution over set of landmarks $\mathbf{x}$ following a 3D point distribution model (PDM) with orthographic camera projection. Similarly to Saragih *et al*. [20], we impose a Gaussian prior on the non-rigid shape parameters on the model. P is a probabilistic patch expert that describes the probability of a landmark being aligned (response map). Patch experts (also called local detectors) are a very important part and from these patch experts, we can infer local confidence. P can be any model producing a probabilistic predictions of landmark alignment. In our work, we define P as a multivariate Gaussian likelihood function of a Continuous Conditional Neural Field (CCNF) [3]:

$$P(\mathbf{y}|\mathcal{I}) = \frac{exp(\Psi)}{\int_{-\infty}^{\infty} exp(\Psi)\ dy}, \tag{2}$$

$$\Psi = \sum_{y_i} \sum_{k=1}^{K1} \alpha_k f_k(y_i, \mathcal{W}(y_i;\mathcal{I}), \theta_k) + \sum_{y_i, y_j} \sum_{k=1}^{K2} \beta_k g_k(y_i, y_j), \tag{3}$$

Above $\mathbf{y}$ is a $m \times m$ area of interest in an image around the current estimate of the landmark (the area we will be searching in for an updated location of the landmark), $\mathcal{W}(y_i;\mathcal{I})$ is a vectorized version of an $n \times n$ image patch centered around y_i and is called the support region (the area based on which we will make a decision about the landmark alignment, typically $m > n$), f_k is a logistic regressor, and g_k is a smoothness encouraging edge potential [3]. It can be shown that (2) is a Multivariate Gaussian function [3], making the exact inference possible and fast to compute. The model parameters $[\alpha, \beta, \theta]$ of the CCNF are learned by using Maximum Likelihood Estimation (using BFGS optimization algorithm).

In order to optimize (1), we use Non-Uniform Regularized Landmark Mean-Shift which iteratively computes the patch responses and updates the landmark estimates by updating the PDM parameters [2].

After optimizing (1), the model likelihoods from each response map are computed. To obtain the local confidence, we take a sum of log probabilities over all of the likelihoods and normalize them to the range between 0 and 1. We used multi-view (three-view) initialization for CLNF model, therefore, confidences need to be normalized according to their view.

TABLE I

THE STRUCTURE OF OUR CONVOLUTIONAL NETWORK USED FOR GLOBAL CONFIDENCE ESTIMATION.

Name	Type	Filter Size	Stride	Output Size
Conv1	convolution	32x3x3	1	32x94x94
Pool1	max-pooling	2x2	2	32x47x47
Conv2	convolution	64x2x2	1	64x46x46
Pool2	max-pooling	2x2	2	64x23x23
Conv3	convolution	128x2x2	1	128x22x22
Pool3	max-pooling	2x2	2	128x11x11
Dense1	fully connected			400
Dense2	fully connected			400
Dense3	fully connected			1

B. Rendering based Global Confidence

We have observed that various facts can affect the quality of rendered images. Not only the alignment error from poor landmarks, but also image quality, illumination or rendering error itself can affect the rendering quality and this has a big influence on the recognition performance. Local confidence can only estimate landmark quality, therefore, we need to predict global confidence for rendered images. Our proposal is motivated by these factors.

We use rendering technique to generate synthetic faces [15], and rendered faces were used for the matching. We render images not only to frontal (0°), but also to half-profile (40°) and profile view (75°) in order to cope with extreme yaw angles. We use three CNN (Convolutional Neural Network) for each rendered view to train our global confidence for rendered images. Table I shows the CNN architecture used in our work. A gray-scale rendered image of size 96×96 is used for input and pixel values are normalized to the range between 0 and 1. The output is a value which represents the global confidence. We use the Euclidean distance as the network loss:

$$loss_{confidence} = \|\hat{\mathbf{C}}_i - \mathbf{C}_i\|_2^2, \tag{4}$$

where $\hat{\mathbf{C}}_i$ is the annotated value, and $\mathbf{C}_i$ is the predicted confidence for training image $\mathcal{I}_i$.

C. Applying Landmark Confidences on Face Recognition

In general, face recognition relies on matching one face image to another, and the matching score is computed to evaluate the probability whether these two images are from the same person or not, as shown in Fig. 2. In this work, the matching score is computed by the correlation distance following [15]. We have two scores (Score 1 and 2 in Fig. 2) by matching two types of face images: with 2D alignment or not (cropping only). The motivation of matching cropped images is to compensate for matching the poorly 2D-aligned images (due to the bad landmarks detected). Now, the problem becomes how to combine these two scores. We employ either or both of the proposed landmark confidences to determine the fusion weights.

While the local confidence measures the goodness of landmarks based on the 2D aligned images, the global

confidence measures it from the 3D rendering point of view. Since both of them reflect the quality of matching scores from 2D aligned images, the lower the confidence is, the lower the fusion weight of this score should be. To model the relation between the landmark confidence and the fusion weight, we exploit a sigmoid function:

$$w_A = \frac{1}{1 + \exp(-m \times (l - r))} \tag{5}$$

where m and r are the slope and offset of the sigmoid function. l is the local confidence or global confidence. The average of the local and global confidences is used when we consider both before applying the sigmoid function. w_A is the fusion weight of the matching score from the 2D alignment and the weight of the score from the cropping only is simply $1 - w_A$.

Exploiting the sigmoid function ensures the fusion weight is always in the range of $[0, 1]$ and has more flexibility to model the relationship between the confidences and fusion weights such as nonlinearlity. Note that there are two landmark confidences (local or global) for a given face image pair, we take the minimum of them[1] as a single confidence associated with this matching score.

IV. Experiments

The datasets we used and the experimental procedures we followed are presented in this section.

A. Datasets

For the CLNF patch expert training, we used Multi-PIE [9] and the training partitions of HELEN [14] and LFPW [4]. Furthermore, we used a multi-view and multi-scale approach as described in Baltrušaitis *et al*. [3].

The CASIA WebFace dataset [24] is used for CNN training for global confidence which is currently the largest publicly available dataset and contains roughly 500K face images.

We evaluate the local and global landmark confidences for the matching score fusion on JANUS CS2 dataset and IJB-A dataset [13]. JANUS CS2 dataset [2] is the extended version of IJB-A dataset [3]. Both JANUS CS2 and IJB-A datasets include unconstrained faces of 500 subjects with extreme poses, expressions and illuminations. JANUS CS2 dataset contains not only the images and sampled key frames but also the original videos of a subject. Also, JANUS CS2 dataset includes much test data for identification and verification than IJB-A dataset.

B. Methodology

In Equation (5), the slope and offset of the sigmoid function are selected in the range of $[3, 5, 7, 9]$, and $[0.1, 0.2, 0.3, 0.4, 0.5]$ respectively based on the performance of the validation set (created in the gallery set). In the following, the LandMark Confidences are shortened as *LMCs*.

Regarding the score fusion, the baselines we compare can be divided into three categories: (1) *No score fusion*: The matching scores from 2D alignment or cropping only, (2) *The fusion methods without LMCs*: The average, max, min, weighted average fusion methods. In the weighted average method, we use the grid search method and pick the fusion weight of the matching score from 2D alignment in the range of 0.1 to 0.9 with interval 0.1 based on the validation set performance, and (3) *The fusion methods with LMCs*: Score replacement, which replaces the scores from the 2D alignment with the ones from cropping only, when the confidence (local or global or the average of them) is less than a threshold (the threshold is picked in $[0.1, 0.2, 0.3, 0.4, 0.5]$ based on the validation set performance). We also compare our approach to [5] by using the landmark confidence directly as the fusion weight without any transformations (denoted as Local LMCs Only in Table II).

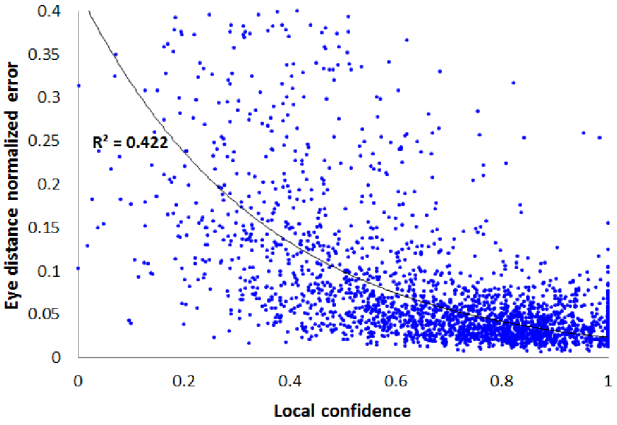


Fig. 3. *Correlation between local confidence and eye distance normalized RMS error.* When the local confidence is higher, the RMS error is smaller

V. Results and Discussion

In this section, we analyze and evaluate our two confidence models. Also, we demonstrate the efficacy of landmark confidences on face recognition by applying them on the face matching score fusion.

A. Evaluation of Local Confidence

Fig. 3 shows the relation between local landmark confidence and eye distance normalized Root-Mean-Square (RMS) error. Since JANUS CS2 is annotated only with 3 key-points on the faces (two eyes and nose base), the error is normalized by the distance between two eyes. We use 25K images from JANUS CS2 dataset but only 2.5K points are plotted to avoid visual distractions. The equation of the regression line is $\mathbf{y} = 0.4237e^{-2897x}$. Coefficient of determination, R^2 is 0.422 and it means 42% of the points fall within the regression line. When the local confidence is above 0.6, most of the eye distance normalized RMS errors

[1]The average and max values are also experimented, but the min value perform the best empirically.

[2]The JANUS CS2 dataset is not publicly available yet.

[3]IJB-A is available under request at http://www.nist.gov/itl/iad/ig/ijba_request.cfm

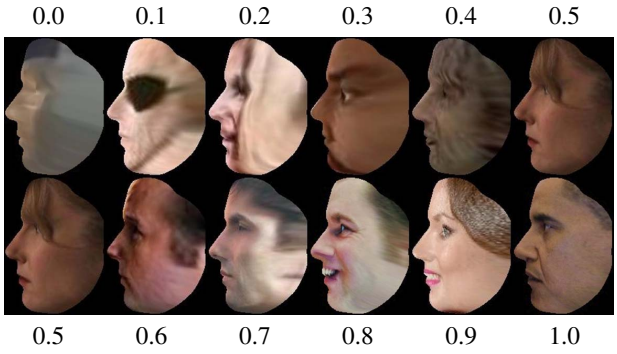


Fig. 4. *Example predicted scores from CASIA test set.* Rendered images to profile view (75°). Higher numbers are better

are below 0.1. This shows that the local confidence above 0.6 is predicting well the align error.

B. Evaluation of Global Confidence

We had two coders who produced annotations for quality of rendered images. Good (+1) and bad (0) annotations are used as labels. To measure inter-coder agreement, we use Krippendorffs alpha [11]. The Krippendorffs alpha value from the two coders is 0.875 which is relatively high since the variable is easy to code. The agreement between the two coders was considered as the gold standard in our experiments. We began training at a learning rate of 0.001 and the learning rate was dropped to $1e^{-8}$ with 0.1 step and set the momentum to 0.9.

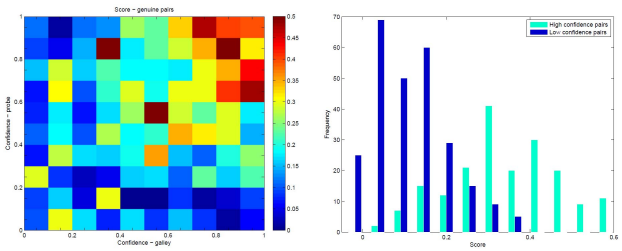
We test the CNN classifier based on the 2.5K of CASIA synthesized images. Among 10K images, three forth images were used for training and one forth images were used for test. Images in the training set were not used in testing. Accuracy of good and bad classification was 95.19% for test set and 99.91% for training set. Fig. 4 shows the example of predicted scores from CASIA test set.

C. Confidences Analysis for Face Recognition

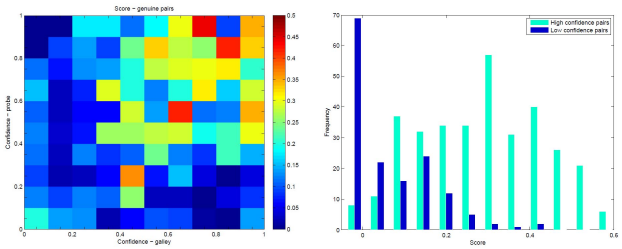
We analyze the relationship between confidences and matching scores. Only one image per template was left to get image-to-image matching scores and we obtained all scores of genuine pairs. Fig. 5 shows the correlation between confidences and matching scores. As shown in left figures, when the confidence of each pair is higher, the score is also higher (the red means higher). Right figures are the histogram of scores of both high (confidence above 0.7) and low (confidence below 0.3) confidence pairs, and they show significant difference in distribution (global confidence $p = 3.44e^{-53}$, local confidence $p = 1.59e^{-40}$ at significance level $p < 0.001$). Considering this analysis, it makes sense to put some weight from both confidences during fusion process.

D. Local and Global Confidences for Face Recognition

In this section, we evaluate the local and global landmark confidences for fusing the matching scores from the 2D



(a) Relation between global confidence and matching scores



(b) Relation between local confidence and matching scores

Fig. 5. *Correlation between confidences and matching scores of genuine pairs* Image-to-Image matching score (left) and histogram of scores of both high and low confidence pairs (left).

alignment and cropping on JANUS CS2 dataset and IJB-A dataset [13]. Since there are multiple scores for each probe-to-gallery image pair, we use the ensemble SoftMax method [15] to obtain a single matching score to measure the similarity between a probe image set and a gallery image set. Following [13], the average verification performances (in terms of TARs at 1%, 0.1%, and 0.01% FARs) and identification performances (in terms of Rank-1, Rank-5, and Rank-10 identification rates) are employed for evaluation and reported in Table II and Table IV. The LandMark Confidences are shorten as LMCs in both tables. The benchmark evaluation measures including the Receiver Operating characteristic curve (ROC) and the Cumulative Match curve (CMC) for Janus CS2 dataset are also displayed in Fig. 6 and 7. Table III demonstrates the comparisons of our method to the commonly used face recognition systems including COTs [13], GOTs [13], Fisher Vector [8], and J.-C. Chen *et al*. [7] on Janus CS2 dataset.

As can be seen in Table II, our method (local LMCs + Sigmoid) outperforms no score fusion, the methods without considering LMCs, and use the LMCs directly as weights by $\sim 21.2\%$, $\sim 8.8\%$, $\sim 9.1\%$ in average for TAR at 0.01% FAR respectively. It can also be seen that the rendering based global confidence achieves the similar improvement, showing the goodness of rendered images also help to evaluate the matching score quality and determine the fusion weights properly. Considering both the local and global confidences further improves the face verification rate in low FARs and the identification rate in rank-1, which can also be seen in Fig. 6 and 7. In Table IV, similar improvements can be oberserved. Moreover, the influence of global confidences is most significant when requiring large TAR at small FAR (e.g. FAR@1e-4), as shown in Table II and Table IV.

TABLE II

AVERAGE PERFORMANCES OF THE SCORE FUSION ON JANUS CS2 DATASET WITH LOCAL AND GLOBAL CONFIDENCES.

Matching method ↓	**TAR(%)**			**Identification Rate (%)**		
Eval. Measures →	FAR@1e-2	FAR@1e-3	FAR@1e-4	Rank-1	Rank-5	Rank-10
2D Alignment [15]	86.91	59.02	24.90	83.66	92.10	94.31
Cropping Only (Unaligned) [16]	72.33	39.34	16.05	69.60	86.11	90.74
Average [19]	86.64	64.01	35.64	82.93	92.24	94.68
Max [19]	82.96	53.05	23.26	79.00	90.82	93.78
Min [19]	85.15	63.39	36.81	81.92	91.96	94.23
Weighted Average	87.56	65.51	35.78	83.76	92.58	94.85
Score Replacement	87.56	63.33	32.06	83.27	92.29	94.60
Local LMCs Only [5]	85.95	61.77	33.11	81.74	91.74	94.21
Local LMCs + Sigmoid	**88.83**	68.90	41.64	**85.03**	**92.96**	**95.13**
Global LMCs + Sigmoid	87.36	67.00	42.51	83.32	92.29	94.45
Local + Global LMCs + Sigmoid	88.76	**69.24**	**43.19**	**85.03**	92.86	94.97

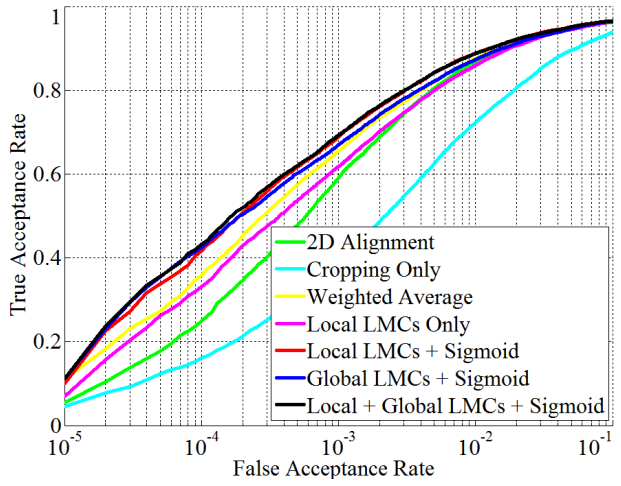


Fig. 6. Comparative performance (ROC) on JANUS CS2 dataset for our confidence based fusion method for face verification. Only TARs of 0% to 1% FARs are shown.

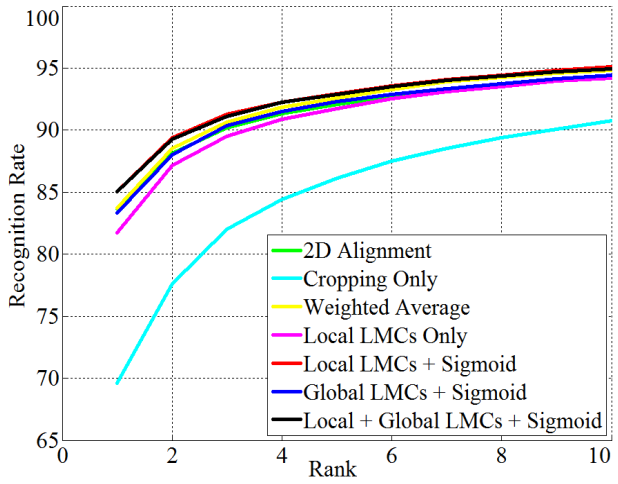


Fig. 7. Comparative performance (CMC) on JANUS CS2 dataset for our confidence based fusion method for face identification.

Table III demonstrates the comparisons of our method to the commonly used face recognition systems including COTs [13], GOTs [13], Fisher Vector [8], and J.-C. Chen *et al*. [7] on Janus CS2 dataset. As can be seen, our method performs better than the other face recognition systems.

Notice in Fig. 6 and 7 that using LMCs as fusion weights directly (denoted as Local LMCs Only) performs worse than the weighted average (without LMCs) in low FARs, but performs much better after applying the sigmoid function. This implies that the relationship of the landmark confidences and fusion weights is nonlinear, and can be better modeled by adjusting the slope and offset of the sigmoid function. To sum up, the matching score fusion in face recognition can benefit from both the proposed landmark confidences and the nonlinear transformation.

VI. CONCLUSION

In this paper, we propose two methods for measuring the landmark confidence: local confidence based on local predictors of each facial landmark, and global confidence based on a 3D rendered face model. Both confidence metrics are analyzed and evaluated. Our experiments show that both confidences are beneficial to face recognition by up to 9% improvements compared to the methods without landmark confidences.

In our future work, we will explore the landmark confidences of videos. A robust and accurate tracking in the video is still a huge challenge today. We will model a new confidence with temporal information in order to help the face tracking and the video based face recognition.

REFERENCES

[1] A. Asthana, S. Zafeiriou, S. Cheng, and M. Pantic. Robust discriminative response map fitting with constrained local models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3444–3451, 2013.

[2] T. Baltrusaitis, P. Robinson, and L.-P. Morency. Constrained local neural fields for robust facial landmark detection in the wild. In *The IEEE International Conference on Computer Vision (ICCV) Workshops*, June 2013.

[3] T. Baltrušaitis, P. Robinson, and L.-P. Morency. Continuous conditional neural fields for structured regression. In *Computer Vision–ECCV 2014*, pages 593–608. Springer, 2014.

TABLE III

PERFORMANCE COMPARISONS OF THE PROPOSED METHOD TO THE PREVIOUS FACE RECOGNITION APPROACHES ON JANUS CS2 DATASET.

Matching method ↓	**TAR(%)**			**Identification Rate (%)**		
Eval. Measures →	FAR@1e-2	FAR@1e-3	FAR@1e-4	Rank-1	Rank-5	Rank-10
COTs [13]	58.1	37	-	55.1	69.4	74.1
GOTs [13]	46.7	25	-	41.3	57.1	62.4
Fisher Vector [8]	41.1	25.0	38.1	55.9	63.7	-
J.-C. Chen *et al*. [7]	87.6	-	-	83.8	92.4	94.9
Ours(Local + Global LMCs + Sigmoid)	**88.76**	**69.24**	**43.19**	**85.03**	**92.86**	**94.97**

TABLE IV

AVERAGE PERFORMANCES OF THE SCORE FUSION ON IJB-A DATASET [13] WITH LOCAL AND GLOBAL CONFIDENCES.

Matching method ↓	**TAR(%)**			**Identification Rate (%)**		
Eval. Measures →	FAR@1e-2	FAR@1e-3	FAR@1e-4	Rank-1	Rank-5	Rank-10
2D Alignment [15]	85.32	49.62	19.14	89.97	95.53	96.73
Cropping Only (Unaligned) [16]	79.06	35.56	8.87	85.37	94.28	96.16
Average [19]	85.96	47.47	16.72	89.28	95.51	96.94
Max [19]	84.82	47.34	16.60	89.45	95.49	96.83
Min [19]	83.24	41.19	12.19	87.97	95.21	96.62
Weighted Average	86.42	51.75	19.64	90.18	95.67	96.92
Score Replacement	86.81	52.84	20.51	90.09	95.59	96.90
Local LMCs Only [5]	85.67	49.14	18.01	89.03	95.38	96.85
Local LMCs + Sigmoid	87.89	56.11	22.91	90.39	95.66	**97.05**
Global LMCs + Sigmoid	87.88	55.93	**24.66**	90.27	**95.76**	97.00
Local + Global LMCs + Sigmoid	**88.29**	**56.74**	23.83	**90.42**	95.75	97.02

[4] P. N. Belhumeur, D. W. Jacobs, D. J. Kriegman, and N. Kumar. Localizing parts of faces using a consensus of exemplars. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(12):2930–2940, 2013.

[5] L. Best-Rowden, H. Han, C. Otto, B. F. Klare, and A. K. Jain. Unconstrained face recognition: Identifying a person of interest from a media collection. *IEEE Transactions on Information Forensics and Security*, 9(12):2144–2157, 2014.

[6] X. Cao, Y. Wei, F. Wen, and J. Sun. Face alignment by Explicit Shape Regression. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2887–2894. Ieee, jun 2012.

[7] J.-C. Chen, V. M. Patel, and R. Chellappa. Unconstrained face verification using deep cnn features. In *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1–9, 2016.

[8] J.-C. Chen, S. Sankaranarayanan, V. M. Patel, and R. Chellappa. Unconstrained face verification using fisher vectors computed from frontalized faces. In *IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–8, 2015.

[9] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker. Multi-PIE. In *IEEE International Conference on Automatic Face and Gesture Recognition*, pages 1–8, 2008.

[10] Y. Gurovich, I. Kissos, and Y. Hanani. Quality scores for deep regression systems. In *Image Processing (ICIP), 2016 IEEE International Conference on*, pages 3758–3762. IEEE, 2016.

[11] A. F. Hayes and K. Krippendorff. Answering the call for a standard reliability measure for coding data. *Communication methods and measures*, 1(1):77–89, 2007.

[12] K. Kim, T. Baltruaitis, A. Zadeh, L.-P. Morency, and G. Medioni. Holistically constrained local model: Going beyond frontal poses for facial landmark detection. In *Proceedings of the British Machine Vision Conference (BMVC)*, pages 095.1–095.12. BMVA Press, September 2016.

[13] B. F. Klare, B. Klein, E. Taborsky, A. Blanton, J. Cheney, K. Allen, P. Grother, A. Mah, and A. K. Jain. Pushing the frontiers of unconstrained face detection and recognition: Iarpa janus benchmark a. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1931–1939, 2015.

[14] V. Le, J. Brandt, Z. Lin, L. Bourdev, and T. S. Huang. Interactive facial feature localization. In *Computer Vision–ECCV 2012*, pages 679–692. Springer, 2012.

[15] I. Masi, S. Rawls, G. Medioni, and P. Natarajan. Pose-aware face recognition in the wild. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4838–4846, 2016.

[16] O. M. Parkhi, A. Vedaldi, and A. Zisserman. Deep face recognition. In *British Machine Vision Conference*, volume 1, page 6, 2015.

[17] N. Poh and J. Kittler. A unified framework for biometric expert fusion incorporating quality measures. *IEEE transactions on pattern analysis and machine intelligence*, 34(1):3–18, 2012.

[18] G. Rajamanoharan and T. F. Cootes. Multi-View Constrained Local Models for Large Head Angle Facial Tracking. In *ICCV*, 2015.

[19] A. Ross and A. Jain. Information fusion in biometrics. *Pattern recognition letters*, 24(13):2115–2125, 2003.

[20] J. M. Saragih, S. Lucey, and J. F. Cohn. Deformable Model Fitting by Regularized Landmark Mean-Shift. 91(2):200–215, 2011.

[21] A. Steger, R. Timofte, and L. V. Gool. Failure detection for facial landmark detectors, 2016.

[22] Y. Sun, X. Wang, and X. Tang. Deep convolutional network cascade for facial point detection. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 3476–3483, 2013.

[23] X. Xiong and F. Torre. Supervised descent method and its applications to face alignment. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 532–539, 2013.

[24] D. Yi, Z. Lei, S. Liao, and S. Z. Li. Learning face representation from scratch. *arXiv preprint arXiv:1411.7923*, 2014.

[25] J. Zhang, S. Shan, M. Kan, and X. Chen. Coarse-to-fine auto-encoder networks (cfan) for real-time face alignment. In *Computer Vision–ECCV 2014*, pages 1–16. Springer, 2014.

[26] N. Zhang, M. Paluri, M. Ranzato, T. Darrell, and L. Bourdev. Panda: Pose aligned networks for deep attribute modeling. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1637–1644, 2014.

[27] S. Zhu, C. Li, C. Change Loy, and X. Tang. Face alignment by coarse-to-fine shape searching. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4998–5006, 2015.