TAEF: A Cross-Distance/Environment Face Recognition Method

Chun-Ting Huang† Zhengning Wang⋆ C.-C. Jay Kuo†
University of Southern California†
University of Electronic Science and Technology of China⋆

Abstract

A solution to long distance outdoor face recognition is presented in this work. The proposed method, called the Two-Stage Alignment/Enhancement Filtering (TAEF) system, consists of three main components: a cross-distance face alignment technique, a cross-environment face enhancement technique, and a two-stage filtering system. Given a probe image, the procedure of face alignment, enhancement and matching is executed against all gallery images to eliminate unlikely candidates at once at the first stage for efficiency. Then, the procedure is conducted for every individual probe/gallery image pair for higher accuracy at the second stage. The first rank recognition rates of the TAEF method are 100%, 100% and 97% for 60-, 100- and 150-meter visible-light images in the LDHF database, respectively.

1. Introduction

Although there have been progressive developments in automatic face recognition, most efforts focus on the situation where probe faces are located at a close distance with varying poses. Much less work has been conducted for Face Recognition at A Distance (FRAD), which is common in the video surveillance application. For convenience, long distance face images and short distance face images for matching are called *probe* and *gallery* images, respectively. The FRAD problem is challenging since the quality of probe images, which are often captured in an outdoor environment, is degraded by both distance and environmental effects while gallery images are typically captured in a controlled indoor environment. To address this cross-distance and cross-environment face matching challenge, we propose a solution called the Two-Stage Alignment/Enhancement Filtering (TAEF) method. The TAEF method consists of three main components: cross-distance face alignment, a cross-environment face enhancement and two-stage filtering.

Two-Stage Filtering. A coarse-to-fine two-stage filtering mechanism consisting of initial screening (or coarse-scale processing) and iterative refinement (or fine-scale processing) is proposed in this work. Given a probe image, the procedure of face alignment, enhancement and matching is executed against all gallery images to eliminate unlikely candidates at once at the first stage for efficiency. Then, the procedure is conducted for every individual probe/gallery image pair for higher accuracy and the least probable one in the candidate pool is removed one by one until the last one at the second stage. With the two-stage filtering, TAEF strikes a balance between computational efficiency and matching accuracy.

The superior performance of the TAEF method is demonstrated by experiments conducted on the LDHF database. Its visible-light outdoor images captured in daytime at a distance of 60, 100 and 150 meters are used as probe images while its indoor images shot at 1 meter are used as gallery images. TAEF gives 100%, 100% and 97% first rank recognition rates for 60-, 100- and 150-meter visible-light images. The rest of this paper is organized as follows. The background and related work are discussed in Section 2. The TAEF method is presented in Section 3. Experimental results are shown and evaluated in Section 4. Finally, concluding remarks are given in Section 5.

2. Related Work

Face Alignment. Face alignment involves two steps: finding an initial rough face shape reference, and approaching the ground truth via iterative optimization. A well known technique is facial landmark localization that finds the coordinates of essential components iteratively. It can be further categorized into the model-based and regression-based approaches.

The model-based approach includes the Active Shape Model (ASM) [4] and the Active Appearance Model (AAM) [2]. They were derived based on the principal component analysis (PCA) of shape and appearance of landmarks. As an extension, a descriptor is used to capture the appearance for each landmark while these descriptors are constrained by a shape model in the AAM with a constrained local model (AAM/CLM) [5]. Saragih *et al.* [23] incorporated a mean-shift filter into AAM/CLM to achieve

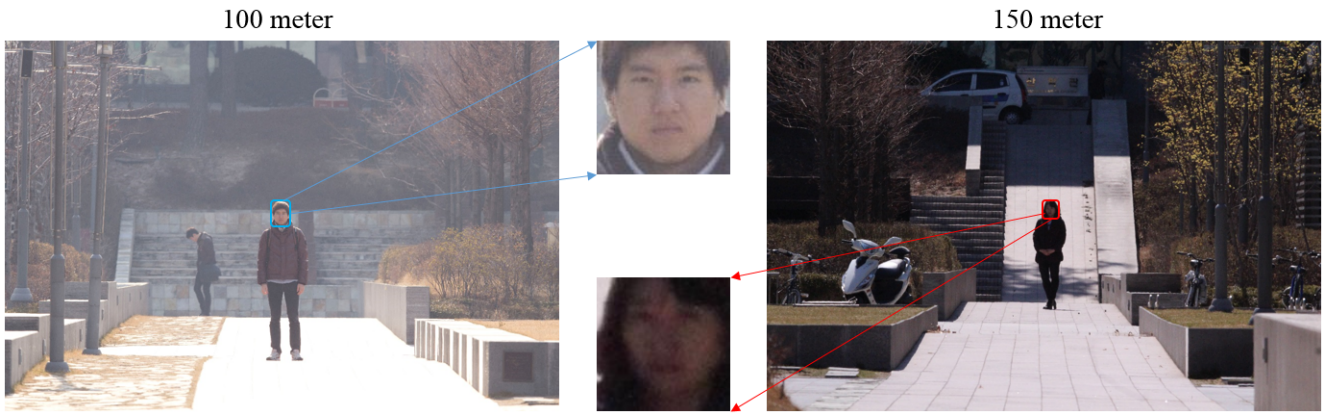

Figure 1. Two exemplary images in the LDHF database taken at a distance of 100 and 150 meters.

better matching capability. Cootes *et al.* [3] adopted a random forest method to compute accumulated votes to improve the alignment performance.

The regression-based approach utilizes local descriptors and regressors to reduce the matching error. The cascaded regression was introduced by Dollar *et al.* [8] for pose estimation in image sequences. Later, it was applied to face alignment. Cao *et al.* [1] proposed a regression method with two-stage training, where the cascaded regression was extended to the context of an affine transform. Xiong *et al.* [29] applied cascaded regression with the SIFT feature, examined the derived solution from a gradient descent view, and called it the Supervised Descent Method (SDM). Yan *et al.* [30] adopted a similar framework with the "learn-to-rank" and "learn-to-combine" modules placed in the front and the back of the main alignment module, respectively. Moreover, deep learning was introduced in [24, 25], which offers competitive performance.

Automatic face alignment techniques and systems have been extensively tested. For instance, Wagner *et al.* [28] used the spare representation in their alignment algorithm for the Multi-PIE database. Geng and Jiang [9] developed an automatic alignment system based on both holistic and local features and conducted experiments on the AR, GT and ORL databases. Deng *et al.* [7] proposed a Transform-Invariant PCA (TIPCA) method to achieve automatic alignment and tested its performance on the FERET dataset.

In this work, a face alignment method using cascaded regression is adopted. To address alignment distortion caused by the distance effect, we design a filter to mimic the long distance effect and generate facial landmarks accordingly.

Face Enhancement. Face image enhancement for FRAD received little attention before. However, it is much needed for FRAD as evidenced by the exemplary images shown in Figure 1. A system that incorporates wavelet decomposition, deblurring, denoising and linear stretching was proposed in [31] to recover quality loss due to the long distance. The reported recognition performance ranges from 50% to 70% due to low image resolution and quality. One face enhancement technique rooted in retinex theory was proposed by Land and McCann [15]. It examined relative lightness (instead of absolute lightness) in a local region to mimic the human visual experience. Later, Land [14] presented another approach to lightness computation. Based on this foundation, Jobson *et al.* [12] proposed a Single-Scale Retinex (SSR) method for the trade-off between rendition and dynamic range compression, and extended it to the Multi-Scale Retinex (MSR) method in [11]. More recently, Petro *et al.* [20] proposed a new method called MSR with Color Restoration (MSRCR).

In this work, the Multiscale Retinex with Color Restoration (MSRCR) method is adopted [20] for face enhancement to handle foggy and back-lighted conditions in the outdoor environment. To the best of our knowledge, this is the first time for MSRCR to be used in a face alignment/recognition system. We show that MSRCR can restore face quality in the Long Distance Heterogeneous Face (LDHF) database, where local contrast enhancement is used to overcome the back-lighted or foggy challenge while maintaining image color balance.

Long Distance Face Database. There are few FRAD databases accessible to the public. The UTK-LRHM database [31], which was built in 2008, contains 55 subjects in an indoor environment with distances ranging from 10 to 16 meters and 48 subjects in an outdoor environment with distances from 50 to 300 meters. Another database, built by Rara *et al.* [22] for stereo reconstruction in 2009, has 30 subjects with three distances (i.e., 3, 15, and 33 meters). Tome *et al.* [26] evaluated distance degradation using standard approaches and matchers on the "Face still dataset" of the NIST Multiple Evaluation Grand Challenge (MBGC) [21]. The NFRAD database constructed by Maeng *et al.* [17] has 50 subjects, whose images were take at 1- and 60-meter distances under both fluorescent light (day) and infrared light (night). The LDHF database [13, 18] was released in 2012. It contains 1-meter indoor, 60-, 100- and

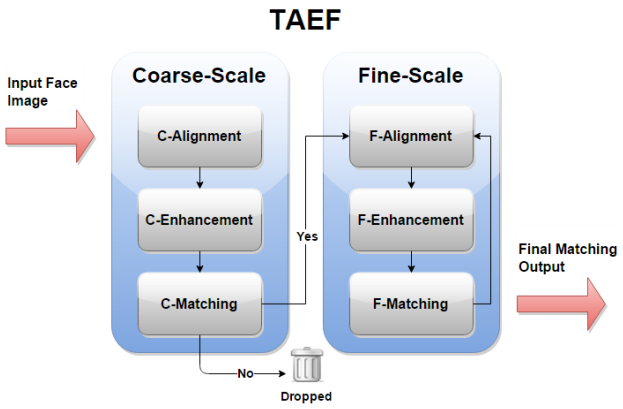

Figure 2. The system diagram of TAEF.

150-meter outdoor visible-light and near-infrared images of 100 subjects.

In this work, we focus on visible-light images taken during daytime only. Two exemplary LDHF visible-light images are shown in Figure 1, which were taken at 100 and 150 meters, respectively. They have the same image size (i.e., 5184 $\times$ 3456 pixels) but different face sizes (i.e., 220 $\times$ 220 pixels for the 100-meter image and 120 $\times$ 120 pixels for the 150-meter image on average). The illumination in the LDHF database can be roughly categorized into three types, such as normal, foggy, and back-lighted. The two images in Figure 1 show how image quality can be affected by the foggy and the back-lighted environments. Apparently, both face alignment and enhancement techniques are needed before face matching. Furthermore, for cross-distance/environment facial matching, features contained in the interior face region could be too weak and additional information such as face contour and partial hair is helpful.

3. Proposed TAEF Method

The diagram of the TAEF system is shown in Figure 2. At the first stage, approximate facial landmarks are obtained using cascaded regression in the alignment step and MSRCR in the enhancement step. The objectives of this stage are two folds: filtering out unlikely candidates to reduce the size of the candidate pool and providing a better initialization for further processing. At the second stage, fine-scale alignment, enhancement and matching operations are performed iteratively to reduce the least probable candidate one by one until the last one is reached.

3.1. Coarse-scale Processing

Coarse-scale Alignment (C-Alignment). Typically, the initial location of the face region is provided by face detection algorithms. It can be affected by the appearance variation such as poses, expressions and occlusion. In this work, we focus on frontal faces with limited facial expression as provided by the LDHF database as the start point. With this simplified condition, we adopt the Viola-Jones [27] algorithm as the face detector and obtain an acceptable prediction outcome for the detected face region.

Most automatic alignment algorithms developed today do not work well in a long distance environment due to blurring and illumination distortions. Although the test image contains a long distance face, the ground truth is a short-distance face of higher resolution and better quality. Apparently, there is a mismatch between the training and testing data. To address the mismatch problem, we design a distortion filter that mimics the long distance effect so as to generate the facial landmarks of synthetic long distance images. The cascaded regression scheme has to be retrained using the distance-adjusted data. Besides, we need an initial location for each landmark in the cascaded regression. To achieve this goal, we conduct Procrustes analysis on all landmarks of images in the distance-adjusted training set to yield a face model formed by the averaged locations of landmarks. After the face region bounding box on the test image is generated and a face model is constructed based on the distance-adjusted training set, we map these landmarks to their corresponding locations in the test face region to generate initial landmark locations. Then, given an input face image, we apply the cascaded regression to reduce the distance between the estimated landmark position and that of the training data. This process can be written mathematically as follows.

To conduct the cascaded regression, we need initial facial shape and training set of multiple subjects. The initial facial shape is represented by the coordinates of N initial facial landmarks in form of $S^0 = [x_1, y_1, \cdots, x_N, y_N]$. The training set is denoted by $\{(I_k, \overline{S}_k)\}_{k=1}^K$, where I_k is the kth subject, $\overline{S}_k$ is the corresponding landmark-based facial representation, and K is the total number of training subjects. With these two inputs, cascaded regression generates a sequence of approximations $S^1, \cdots, S^t, \cdots, S^T$, where S^T is the converged output. The t-th facial shape is updated based on

$$S^t = S^{t-1} + R^t(I, S^{t-1}), \tag{1}$$

where

$$R^t = \arg\min_R \sum_{k=1}^{K} ||\overline{S}_k - [S_k^{t-1} + R(I_k, S_k^{t-1})]|| \tag{2}$$

is learned. R represents a regressor and S_k^{t-1} is the estimated shape produced in the previous stage. In this work, R is chosen to be a linear regressor since it can handle the desired task efficiently. To explain the above concept in words, we design a sequence of regressors, where each regressor is trained based on the difference between the estimated result from the previous stage, S_k^{t-1}, $k = 1, \cdots, K$, and the ground truth, $\overline{S}_k$. This iteration process stops when

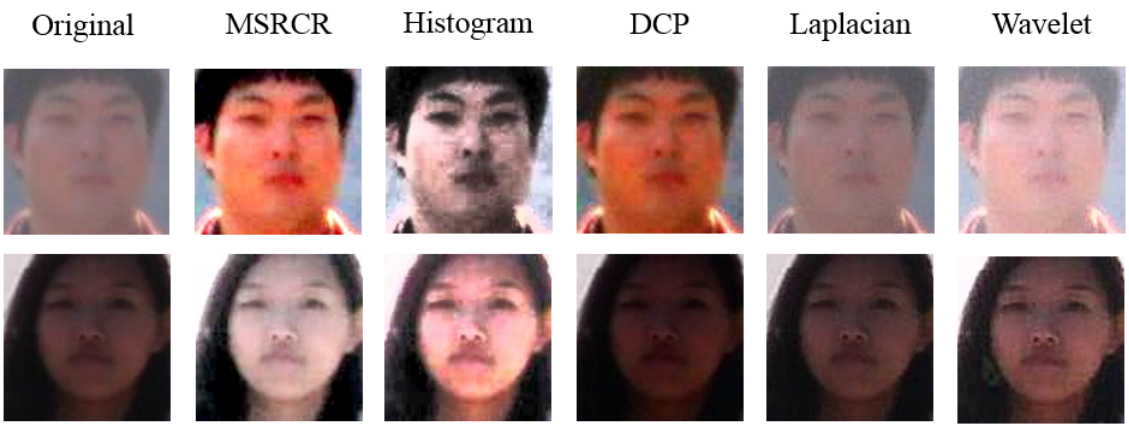

Figure 3. Two exemplary enhanced facial images (from left to right): original images and images enhanced by MSRCR, histogram equalization, dark channel prior, Laplacian sharpening and wavelet decomposition.

the training error converges. In our work, we adopt multi-scale HOG features [6] as the input descriptor for regressor's training. To be more specific, the large-scale HOG feature is extracted in the beginning stage while only a small-scale area around each estimated landmark is considered in the later stage.

Furthermore, we regulate the solution at the end of each iteration with two constraints. First, we adopt the sketch token feature from [16] as a reference for the face contour since it offers a reliable edge map using the mid-level feature. It is observed that the trained sketch token model offers an exceptional result on long distance faces in the designated region. Second, the estimated landmarks are constrained based on the face shape with the closest distance, so that no landmarks will deviate from the ground truth too much because of image quality degradation.

To sum up, the predicted facial landmarks for the kth subject at the end of the t-th iteration are obtained as the fusion of results from: 1) the predicted result from the t-th regressor, 2) the output obtained by imposing the sketch-token-based face contour constraint, and 3) the closest landmark model selected from the training set.

Coarse-scale Enhancement (C-Enhancement). After getting landmarks from the C-alignment procedure, we attempt to restore the distorted facial color so as to allow robust cross-environment facial matching. We test several enhancement algorithms, including histogram equalization, dark channel prior [10], Laplacian sharpening, wavelet decomposition and MSRCR, and conclude that MSRCR provides a superior performance on foggy and back-lighted images. Two examples are given in Figure 3 for subjective visual comparison. The top and bottom original images in Figure 3 are distorted by the foggy and back-lighted conditions, respectively. The goal of enhancement is to remove these environmental factors to allow cross-environment matching. We see that MSRCR does provide better results against the original ones.

For objective performance evaluation, we compare the Receiver Operating Characteristic (ROC) curve and the Cumulative Match Characteristic (CMC) curve of the matching result under different enhancement algorithms in Figure 4, where the matching result in this figure is generated using only the coarse-scale alignment/enhancement and will be detailed in the next subsection. We see from this figure that only MSRCR can improve the matching performance. It is worthwhile to point out that most image enhancement algorithms have been developed for white noise removal. The white noise model is however not suitable in characterizing outdoor distortions. The matching performance is actually worsen by these enhancement algorithms designed for other purposes. In contrast, MSRCR compensates the environment effect with a more suitable design and, as a result, it can offer better performance over the original one.

Facial Matching (C-Matching and F-Matching). The facial matching component appears in both the coarse-scale and the fine-scale processing modules, where the same matching method is adopted and described below.

After proper alignment and enhancement, the first step is to extract feature descriptors (e.g. HOG and SIFT) and geometric features (i.e. facial landmarks) in polar coordi-

Figure 5. Illustration of interior and bounded face regions.

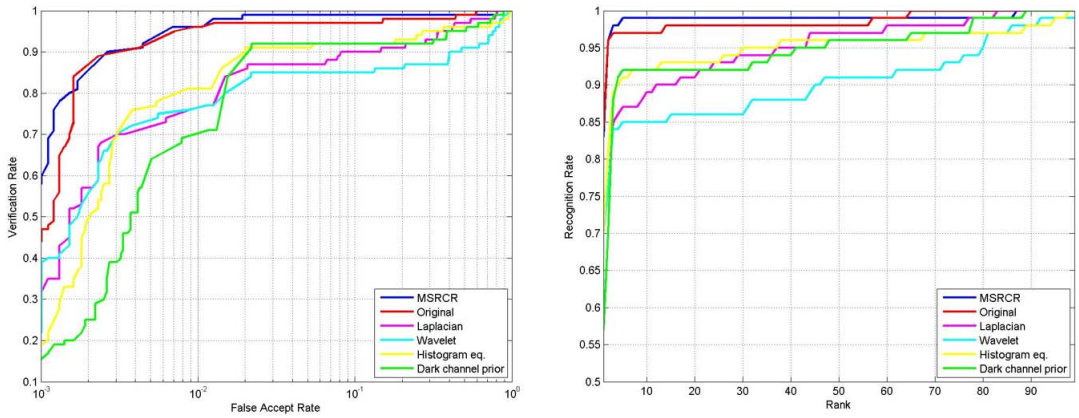

Figure 4. Comparison of the 150-to-1 meter face matching result with different enhancement methods: ROC (left) and CMC (right).

nates. Polar coordinates are adopted because it can represent relative locations of aligned landmarks conveniently. Feature descriptors are extracted from two cropped face regions called the *interior face region* and the *bounded face region*, respectively. Two examples are illustrated in Figure 5. The interior face region includes major facial components up to eyebrows and down to part of chin without ears and hair. The bounded face region has the whole face including the face contour and partial hair such as bangs. Since the bounded face region is sensitive to background and hairstyle change, it is not used in traditional face recognition systems. However, for the cross-distance and cross-environment face recognition problem, the information contained in the interior face region could be too little. The additional information contained in the bounded face region can be helpful, and it is not proper to discard any relevant information due to aggressive cropping. Experiments show that the performance of using information from both interior and bounded face regions is better than that from only one of them. Thus, in our implementation, both HOG and SIFT features from the two regions are used separately as individual classifiers for further processing.

Moreover, both regions share the same Interpupillary Distance (IPD) as 40 pixels. The number 40 is chosen because it is the average IPD for 150 meter images, and the cutoff range is decided based on the total average face boundary. The reason to fix the IPD in both regions is to maintain the resolution alignment since we may generate distortions during resizing by allowing images of different scales.

After collecting all features from both interior and bounded regions of aligned and enhanced face images, we can measure the Euclidean distance of feature vectors and generate rank-order lists from all classifiers. Then, a weighted voting will be used to pile all classification results into one single rank-ordered list. By gradually eliminating less probable candidates in various stages, the TAEF system will provide the final ranked result.

3.2. Fine-scale Processing

An iterative alignment/enhancement filtering process is adopted in this stage. It means that, after eliminating the least possible candidate from the selection pool through alignment, enhancement and matching, features are extracted from re-normalized images based on remaining images in the pool at the next iteration. This process is described in detail below.

Sets of probe and gallery images are denoted by $P = \{P_1, \cdots, P_k, \cdots, P_{N_p}\}$ and $G = \{G_1, \cdots, G_k, \cdots, G_{N_g}\}$, where N_p and N_g are their sizes. Furthermore, $O_1, \cdots, O_k, \cdots, O_{N_p}$ are candidate pools for probe images $P_1, \cdots, P_k, \cdots, P_{N_p}$, respectively. The iterative filtering process consists of two steps at each iteration. First, probe image P_i is geometrically and photometrically normalized with subjects left in its candidate pool O_i so that the normalized probe image can be written as

$$P_i' = \Gamma(G_j, \Lambda(G_j, P_i)), \quad G_j \in O_i, \tag{3}$$

where Λ and Γ are the fine-scale alignment (F-alignment) and enhancement (F-enhancement) operations, respectively. Afterwards, the new candidate pool is expressed as

$$O_i' = \{O_i \mid V_j \geq N_v\}, \tag{4}$$

where

$$V_j = \sum_{l=1}^{N_c} w_l \cdot C_l(\Psi_l(P_i'), j) \tag{5}$$

is the vote collected from classifiers C_l, $l = 1, \cdots N_c$ using feature transform Ψ_l and weighting factor w_l, and N_v is a threshold of vote count for the pool.

The same cascaded regression in C-alignment is applied to the probe image in the F-alignment but with one major difference. That is, it is aligned with each individual gallery image G_j in the candidate pool O_i one by one, where the

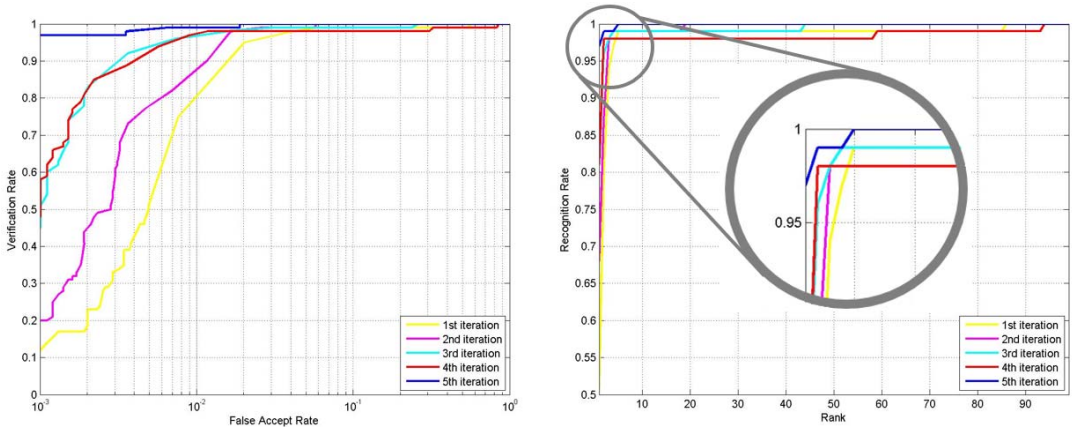

Figure 6. Performance of the 150-to-1 meter face matching result with different enhancement methods: ROC (left) and CMC (right).

Procrustes analysis is conducted to derive the transform array. Translation, orthogonal rotation, reflection, and scale component are all calculated in this process. Furthermore, only reliable landmarks are selected to reduce the influence from inaccurate landmarks. For example, tips of eyes and the mouth often have higher steadiness and the nose rim location is difficult to determine in long distance. With these improvements, the F-alignment, denoted by Λ in Eq. (3) can reduce the estimation error based on the improvement in the last iteration.

The F-enhancement process, denoted by Γ in Eq. (3), is needed for photometric normalization, and it is achieved by region-based histogram matching. In contrast with the traditional histogram matching method that calculates the histogram of the whole face, we match histograms of the probe image and each individual gallery image in face sub-regions segmented by localized landmarks in the F-alignment step so as to differentiate images in a small candidate pool.

4. Experimental Results

Implementation Details. Since the LDHF database has a limited number of daytime gallery images to be used as training samples, we also include the MUCT database [19], which consists of 3755 images with 76 landmarks and is collected from 276 subjects. However, we only use 25 landmarks out of 76 in the C-alignment training process by focusing on visible landmarks such as center and edge tips of eyes and mouth in the long distance. The training set in C-alignment also contains gallery images from LDHF. Each face is manually labeled with 25 landmarks since the ground truth is not provided with the database. We simulate three long distance scenarios (contrast change due to long distance, fog and back-lighted) for each gallery image so that the size of the training images is tripled. This allows the regressor to learn in a cross-environment setting.

In the C-enhancement step, we apply MSRCR to the whole probe image, where its parameters are decided by the histogram distribution of each probe image. If the distribution contains a concentrated peak in the dark area, it should be under the back-lighted condition. If the distribution spans over a broad area, it should be under the foggy condition. Otherwise, it is under the normal condition. In our experiment, we set parameters $\alpha = 0.7$ and $c = 0.5$ in the Laplacian sharpening method and parameters $threshold = 50$ and $C = 2$ in the wavelet decomposition method. The enhancement performance is shown in Figure 4, where the verification rate under $FAR = 0.1\%$ is: 12% for dark channel prior, 18% for histogram equalization, 22% for wavelet decomposition, 32% for Laplacian sharpening, 44% for original and 58% for MSRCR.

In the fine-scale stage, TAEF collects votes from all classifiers to build up a candidate pool. We observe that HOG and SIFT features have the ability to select candidates of high similarity but with low first-rank accuracy. They can be used as the main features for both interior and bounded face regions, yet they need to be assisted with geometric features offered by landmarks. As a result, we have six classifiers based on the following feature sets: HOG and SIFT from interior and bounded face regions, landmark's angle and radius distributions (represented in polar coordinates). The voting mechanism collects votes from all six classifiers, and the top N candidates that receives most votes are placed in the initial candidate pool ($N = 5$ in the experiment). Then, one candidate is removed at each iteration until the final one is reached.

Performance Evaluation. We compare curves of ROC and CMC in Figure 6 to demonstrate the performance of the TAEF system. Note that we need a distance table among all candidates to draw ROC and CMC, where the distance table is built based on the received number of votes. A higher vote number means a closer distance and vice versa, and the distance is weighed by the iteration number.

We can see the performance improvement in ROC curve as the iteration number increases. For example, when

Table 1. Comparison of ROC verification rates for 150-to-1 cross-distance face recognition.

Methods	0.1% FAR	1% FAR	10% FAR
Maeng [17]	93%	99%	100%
Kang [13]	75%	87%	99%
TAEF	**97%**	99%	100%

$FAR = 0.001$, TAEF gives a verification rate of 12%, 20%, 45%, 48% and 97% at the 1st, 2nd, 3rd, 4th and 5th. The superior performance of the TAEF method is also demonstrated by the CMC plot. At the first iteration, the first rank recognition rate is 51% and it rapidly climbs up to 99% in rank 5. It follows the same aggregation pattern for later iterations. Its first rank recognition rates are 68%, 81%, 81% and 97% for iteration numbers 2, 3, 4 and 5, respectively.

For performance benchmarking, we selected the work of Maeng *et al.* [17] and that of Kang *et al.* [13]. Note that the former did not provide sufficient details on their alignment process while the latter relied on a commercial software called FaceVACS, and manually provided eye locations when the software failed to detect. For these reasons, we can only take the reported data from their papers for the comparison purpose. We list the ROC verification rates of three methods (TAEF and theirs) for 150-meter visible-light images in LDHF in Table 1. TAEF has the best performance among the three. Moreover, we test the TAEF method using 60 meter and 100 meter visible-light images in LDHF, and it gives 100% first rank recognition rate. Thus, TAEF offers the state-of-the-art performance for the FRAD problem at an outdoor setting.

We also compare the first rank recognition rate using features from only the interior or bounded face region or both under the same settings. The results for the 150-meter visible-light images at first rank are shown in Table 2. It is interesting to see that the performance of the bounded face region alone is better than that of the interior face region. Since the interior face region is blurred due to the long distance effect, its extracted features have limited discriminant power. The additional information from the bounded face region such as the face contour and hairstyle can play an important role although it is less robust. The TAEF system takes both into account and achieves the best performance.

Error Analysis. Among the 100 probe images located at the 150-meter distance, there are three failure cases for TAEF as shown in Figure 7. we show two intermediate processing results of probe images in the first two columns: the output after C-alignment in the first column and the final normalized result in the second column. Furthermore, their ground truth of the 1-meter gallery image is shown in the fourth column while their predicted match by TAEF is shown in the third column. The ground truth images of these three subjects from top to bottom rank as No. 2, No. 2, and No. 4, respectively.

Table 2. First rank recognition rates for different face regions.

Face region	Interior	Bounded	Both
1^{st} iteration	35%	46%	51%
2^{nd} iteration	49%	54%	68%
3^{rd} iteration	60%	69%	81%
4^{th} iteration	60%	72%	81%
5^{th} iteration	66%	86%	97%

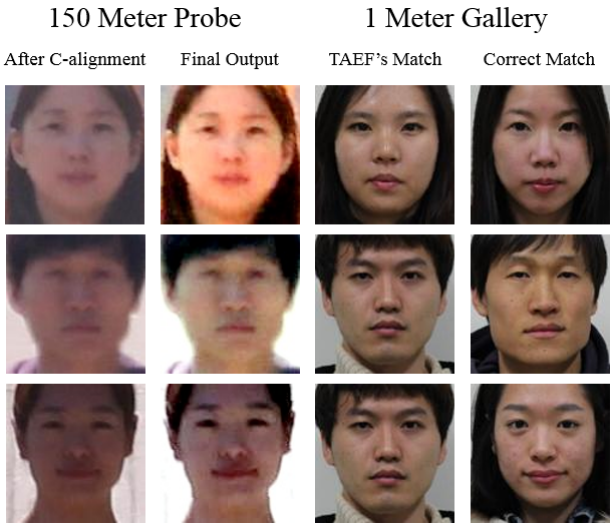

Figure 7. Error cases for 150 meter visible-light probe images.

One obvious reason for the error is attributed to the difference of the hair style between gallery and probe images. For example, for the case in the bottom row, the hairstyle of the 1-meter gallery ground truth is completely changed in her corresponding 150-meter probe image. Generally speaking, the hairstyle and the chin shape visible in the bounded face region do contribute positively to the recognition performance. This case happens to work against this policy.

Another reason is due to other environmental factors such as blurring, which is not yet considered in the TAEF system. For the first two rows, the interior face regions of the final output images from the TAEF system are still blurred. The loss of details in pupils and eye's shape can mislead HOG and SIFT descriptor classifiers. With these two blurred probe images, the TAEF system fails to choose the correct one in the last round.

5. Conclusion and Future Work

In this paper, we presented an interesting and practical face recognition problem where the probe images are located at a distance in an outdoor environment. We discussed

several challenging issues existing in this problem and proposed a solution called the TAEF method to address them. The TAEF method offers a state-of-the-art solution to this cross-distance and cross-environment face matching problem.

As compared with other face recognition databases, the size and the variety of the existing long distance face recognition database are still very limited. It is urgent to build a larger database in the research community to facilitate future research and development work along this line. This is a task that we would like to pursue in the near future.

References

[1] X. Cao, Y. Wei, F. Wen, and J. Sun. Face alignment by explicit shape regression. *IJCV*, 107(2):177–190, 2014. 2

[2] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. In *ECCV*, pages 484–498. Springer, 1998. 1

[3] T. F. Cootes, M. C. Ionita, C. Lindner, and P. Sauer. Robust and accurate shape model fitting using random forest regression voting. In *ECCV*, pages 278–291. Springer, 2012. 2

[4] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models-their training and application. *Computer vision and image understanding*, 61(1):38–59, 1995. 1

[5] D. Cristinacce and T. Cootes. Automatic feature localisation with constrained local models. *Pattern Recognition*, 41(10):3054–3067, 2008. 1

[6] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *CVPR*, volume 1, pages 886–893. IEEE, 2005. 3

[7] W. Deng, J. Hu, J. Lu, and J. Guo. Transform-invariant pca: A unified approach to fully automatic facealignment, representation, and recognition. *TPAMI*, 36(6):1275–1284, June 2014. 2

[8] P. Dollar, P. Welinder, and P. Perona. Cascaded pose regression. In *CVPR*, pages 1078–1085, June 2010. 2

[9] C. Geng and X. Jiang. Fully automatic face recognition framework based on local and global features. *Machine vision and applications*, 24(3):537–549, 2013. 2

[10] K. He, J. Sun, and X. Tang. Single image haze removal using dark channel prior. *TPAMI*, 33(12):2341–2353, 2011. 4

[11] D. J. Jobson, Z.-U. Rahman, and G. A. Woodell. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *Transactions on Image Processing*, 6(7):965–976, 1997. 2

[12] D. J. Jobson, Z.-U. Rahman, and G. A. Woodell. Properties and performance of a center/surround retinex. *Transactions on Image Processing*, 6(3):451–462, 1997. 2

[13] D. Kang, H. Han, A. K. Jain, and S.-W. Lee. Nighttime face recognition at large standoff: Cross-distance and cross-spectral matching. *Pattern Recognition*, 47(12):3750–3766, 2014. 2, 7

[14] E. H. Land. Recent advances in retinex theory and some implications for cortical computations: color vision and the natural image. *Proceedings of the National Academy of Sciences of the United States of America*, 80(16):5163, 1983. 2

[15] E. H. Land and J. McCann. Lightness and retinex theory. *JOSA*, 61(1):1–11, 1971. 2

[16] J. Lim, C. L. Zitnick, and P. Dollár. Sketch tokens: A learned mid-level representation for contour and object detection. In *CVPR*, 2013. 4

[17] H. Maeng, H.-C. Choi, U. Park, S.-W. Lee, and A. K. Jain. Nfrad: Near-infrared face recognition at a distance. In *Biometrics (IJCB)*, pages 1–7. IEEE, 2011. 2, 7

[18] H. Maeng, S. Liao, D. Kang, S.-W. Lee, and A. K. Jain. Nighttime face recognition at long distance: Cross-distance and cross-spectral matching. In *ACCV*, pages 708–721. Springer, 2013. 2

[19] S. Milborrow, J. Morkel, and F. Nicolls. The muct landmarked face database. *Pattern Recognition Association of South Africa*, 2010. http://www.milbo.org/muct. 6

[20] A. B. Petro, C. Sbert, and J.-M. Morel. Multiscale retinex. *Image Processing On Line*, pages 71–88, 2014. 2

[21] P. J. Phillips, P. J. Flynn, J. R. Beveridge, W. T. Scruggs, A. J. Oąetoole, D. Bolme, K. W. Bowyer, B. A. Draper, G. H. Givens, Y. M. Lui, et al. Overview of the multiple biometrics grand challenge. In *Advances in Biometrics*, pages 705–714. Springer, 2009. 2

[22] H. Rara, S. Elhabian, A. Ali, M. Miller, T. Starr, and A. Farag. Face recognition at-a-distance based on sparse-stereo reconstruction. In *CVPR Workshops*, pages 27–32. IEEE, 2009. 2

[23] J. M. Saragih, S. Lucey, and J. F. Cohn. Deformable model fitting by regularized landmark mean-shift. *IJCV*, 91(2):200–215, 2011. 1

[24] Y. Sun, X. Wang, and X. Tang. Deep convolutional network cascade for facial point detection. In *CVPR*, pages 3476–3483, June 2013. 2

[25] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf. Deepface: Closing the gap to human-level performance in face verification. In *CVPR*, pages 1701–1708. IEEE, 2014. 2

[26] P. Tome, J. Fierrez, F. Alonso-Fernandez, and J. Ortega-Garcia. Scenario-based score fusion for face recognition at a distance. In *CVPR Workshops*, pages 67–73. IEEE, 2010. 2

[27] P. Viola and M. J. Jones. Robust real-time face detection. *IJCV*, 57(2):137–154, 2004. 3

[28] A. Wagner, J. Wright, A. Ganesh, Z. Zhou, H. Mobahi, and Y. Ma. Toward a practical face recognition system: Robust alignment and illumination by sparse representation. *TPAMI*, 34(2):372–386, 2012. 2

[29] X. Xiong and F. De la Torre. Supervised descent method and its applications to face alignment. In *CVPR*, pages 532–539, June 2013. 2

[30] J. Yan, Z. Lei, D. Yi, and S. Li. Learn to combine multiple hypotheses for accurate face alignment. In *ICCV Workshops*, pages 392–396, Dec 2013. 2

[31] Y. Yao, B. R. Abidi, N. D. Kalka, N. A. Schmid, and M. A. Abidi. Improving long range and high magnification face recognition: Database acquisition, evaluation, and enhancement. *Computer Vision and Image Understanding*, 111(2):111–125, 2008. 2