

An Associate-Predict Model for Face Recognition

Qi Yin^{1,3}

Xiaoou Tang^{1,2}

Jian Sun³

¹Department of Information Engineering
The Chinese University of Hong Kong

²Shenzhen Institutes of Advanced Technology
Chinese Academy of Sciences, China

³Microsoft Research Asia

Abstract

Handling intra-personal variation is a major challenge in face recognition. It is difficult how to appropriately measure the similarity between human faces under significantly different settings (e.g., pose, illumination, and expression). In this paper, we propose a new model, called “Associate-Predict” (AP) model, to address this issue. The associate-predict model is built on an extra generic identity data set, in which each identity contains multiple images with large intra-personal variation. When considering two faces under significantly different settings (e.g., non-frontal and frontal), we first “associate” one input face with alike identities from the generic identity data set. Using the associated faces, we generatively “predict” the appearance of one input face under the setting of another input face, or discriminatively “predict” the likelihood whether two input faces are from the same person or not. We call the two proposed prediction methods as “appearance-prediction” and “likelihood-prediction”. By leveraging an extra data set (“memory”) and the “associate-predict” model, the intra-personal variation can be effectively handled.

To improve the generalization ability of our model, we further add a switching mechanism - we directly compare the appearances of two faces if they have close intra-personal settings; otherwise, we use the associate-predict model for the recognition. Experiments on two public face benchmarks (Multi-PIE and LFW) demonstrated that our final model can substantially improve the performance of most existing face recognition methods

1. Introduction

In the past two decades, the appearance-based approaches [7, 9, 12, 13, 20, 23, 28, 30, 32, 34] have dominated the face recognition field due to their good performance and simplicity. However, large intra-personal variation, like pose, illumination, and expression, remains an inevitable obstacle because it results in significant appearance change, geometric misalignment, and self-occlusion. For example, A and B in Figure 1 are two photos of the same

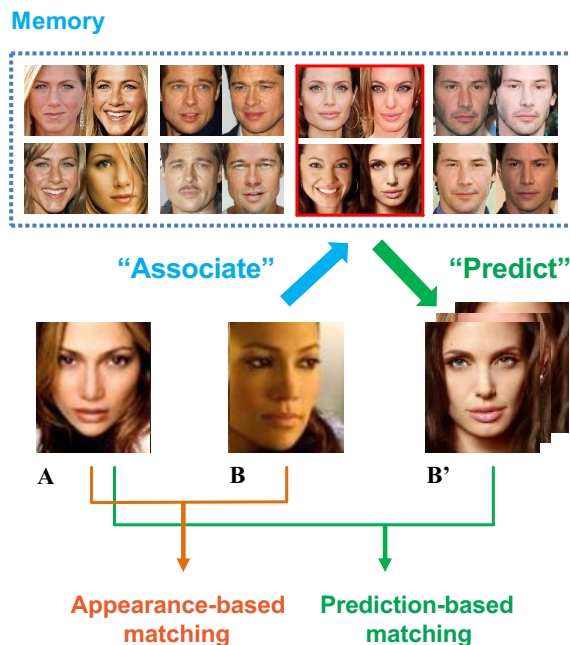


Figure 1. The “Associate-Predict” model. A and B are two faces of “Jennifer Lopez”, with significantly different settings (e.g., pose, lighting, and expression). Conventional methods (“Appearance-based matching”) compare them directly. In our model, we associate a similar generic identity of the face B at the first step. Then, we predict the new appearance (B’) of B under the similar setting of A. Using the predicted new face(s), we perform “Prediction-based matching”. Note that this figure is only for the illustration purpose. We use facial components (instead of the whole face shown here) as the basic element.

person. The appearances of the two faces are so different that any appearance-based approach may draw a conclusion that they are not the same person. But why our human brain has the ability to recognize faces with large intra-personal variation?

Built upon the studies of brain theories, Jeff Hawkins [8] gave his definition on the intelligence: “your brain receives patterns from the outside world, stores them as memories, and makes predictions by combining what it has seen before and what is happening now.”. In a nutshell, he measures the

intelligence by the capacity to remember and predict patterns in the world. Despite that this definition is arguable, we may still get useful inspirations from it.

In this paper, we conjecture that *our brain adopts prior knowledge/memories to predict the possible transition between two faces under significantly different settings*. For example, our brain may be able to associate similar facial patterns from all face images we have seen in our life, and imagine the reasonable appearance of the input face under a different setting (for example, predict from a non-frontal face to its frontal version). After the association and prediction, our brain can do direct appearance comparison between two faces with consistent settings.

Based on the above conjecture, we propose an “Associate-Predict” (AP) model for face recognition. The model is built on a prior identity data set (“memory”), where each identity has multiple face images with large intra-personal variation. This data set is treated as a critical “bridge” to reveal how the same face may vary under different intra-personal settings. If we are going to compare a face pair, we first associate the input face with a few of most similar identities from the “memory”, and then predict the new appearance of the input face under different settings (“appearance-prediction”) or directly predict the recognition output (“likelihood-prediction”), as shown in Figure 3.

In the “appearance-prediction”, given two input faces, we select a specific face image from the associated identity to replace one input face. The selected face is required to have the consistent intra-personal setting with the other input face. The selected face is our predicted/imagined image from one setting to the other setting. As a result, the two input faces are “transited” into the same setting, and we can appropriately use any existing appearance-based approach to match them. For example, suppose face A is frontal while face B is left-oriented. To compare them, we select a left-oriented face A’ from the associated identity of A to replace the original face A and compute the appearance-based similarity between A’ and B.

In the “likelihood-prediction”, we first associate a few most similar identities of one input face. Then, we construct a discriminative classifier using the associated faces of these identities as positive samples and a fixed set of “background” face images as negative samples. The trained person-specific classifier is used to tell if two input faces are from the same person. Since the associated positive samples cover wider intra-personal variation, the resulting classifier can preserve person-specific discriminative information better and be more robust to apparent intra-personal variation.

In the above text, we refer the term “face” as our basic matching element for clarity. In our system, we use twelve facial components as the basic elements for the association/prediction since it is easier to associate a very alike

generic identity at the component level than at the holistic face level.

Our associate-predict model works best when the setting of two input faces are quite different. But due to the limited size of our identity data set, our model may be less discriminative than the direct appearance comparison when two input faces have similar settings. This is true and verified in our experiments. Here, we make the second conjecture that *our brain does direct appearance matching between two faces under similar settings*.

To realize the second conjecture, we enhance our model with a switching mechanism - we do direct appearance matching if two faces have close intra-personal settings; otherwise, we apply the associate-predict model to handle the large intra-personal variation. This hybrid model helps us to get the best of both worlds.

With the associate-predict model and the switching mechanism, our approach significantly improves over the current appearance-based systems. The evaluations on two complementary benchmarks, Multi-PIE and LFW, demonstrated that our system can consistently achieve the leading performance¹, while maintaining very good generalization ability.

2. Related Works

The descriptor-based methods [3, 5, 9, 16, 18, 24, 31, 32] and subspace-based methods [1, 14, 15, 17, 19, 26, 28, 29, 30, 35] are two representative appearance-based approaches. The descriptor-based methods extract discriminative information from the facial micro-structures, and the subspace-based algorithms learn an optimal subspace for recognition. Generally, all the appearance-based methods confront the tradeoff between the discriminative ability and the invariance to intra-personal variation.

To cope with the intra-personal variation, many current researches [10, 21, 27, 32, 33] apply prior knowledge in face recognition. Blanz *et al.* [2] used the prior morphable 3D models to simulate the 3D appearance transformation of the input face. Wolf *et al.* [32, 33] proposed to train binary classifier using a single positive sample and a set of prior background negative samples. To solve the problem of limited positive samples, Kim *et al.* [10] and Wang *et al.* [21] adopted manually designed transformation to generate virtual positive samples. But the virtual positive samples are usually limited in quality and variation degree.

Our work is also related to an interesting face-sketch transformation idea proposed in [25]. In order to compare one person’s photo with another person’s sketch, an identity database of photo-sketch pairs was constructed. An input photo is reconstructed by a linear combination of the

¹Since we use an extra identity data set as prior knowledge, it may be unfair to be compared with some approaches without requiring extra data. But our cross-data experiments show that we do not overfit the data.



Figure 2. Demo of identity data set. Each identity has seven pose categories P_1 - P_7 (from left to right) and four illumination conditions L_1 - L_4 (from top to bottom).

prior identity photos, and then a new sketch of the input photo is generated based on the prior photo-sketch pairs. In a way, our method also tries to use a few similar identity to help transfer the face settings. But unlike dealing with a well defined problem with strict matching pairs, we have to deal with many uncontrolled settings. Recently, Kumar *et al.* [12] also organized an extra identity data set as prior knowledge. They built a kind of high-level face representation, Simile Classifiers, based on the similarity relations between identities. Also relying on the extra identity data, Su *et al.* [22] proposed an adaptive generic learning to fit the possible variation in the test set. However, these methods are globally optimized without considering the specific setting of the input faces, thus still have difficulty to deal with varying settings of the input faces.

Our "Associate-Predict" model is specifically computed for each person. This person-specific approach can somehow relate to an earlier appearance-based method, unified subspace method [28]. For each input face, they selected a few most similar faces to train a specific LDA classifier for the person. The major difference from our work is that, we use the similar identities as "friends" to distinguish from other faces, while in their setting, they are trying to classify the input face against those most similar faces. So the purpose is different, but the person specific training is similar.

3. Identity Data Set and Face Representation

Identity data set. Our extra generic identity data set is built from 200 identities² from the Multi-PIE data set [6]. In this data set, each identity contains 7 different pose categories and 4 illumination conditions, shown in Figure 2. The pose categories range from -60% to +60% to cover the

²These 200 generic identities are selected as the ID 1 - 200 from the Session 1 (with neural expression) of the Multi-PIE data set.

horizontal in-plane rotation and the four illumination conditions are no-flash, left-flash, right-flash, and left-right-flash. This identity data set will be used as the prior knowledge ("memory") for our associate-predict model. Notice that the generic identities are mutually exclusive to all of our test data.

Face representation. We extract face representation at the facial component level. Given an input face image, four landmarks (eye centers and mouth corners) are automatically detected as shown in Figure 4. Then twelve facial components (*e.g.*, left eye and right cheek) are separately aligned based on these four detected landmarks using a 2D affine transformation. We evaluate four representative low-level descriptors: LBP [18], SIFT [16], Gabor [31], and Learning-based (LE) descriptor [3]. For the LBP, we use 8 uniform-spaced circular neighbor sets (radius = 3) [18] and 59-code encoding. We use the default parameter described in [31] for the Gabor descriptor and 32-orientation quantization for the SIFT descriptor. For the LE descriptor, we use the best sampling pattern and 256-code encoding suggested in [3]. To obtain more discriminative representation, we apply PCA technique [3] to get a compact descriptor f_i for each component. The whole face is represented as $\mathcal{F} = (f_1, f_2, \dots, f_{12})$. The final similarity distance of two faces is the fusion of the component-wise L2 distances by a linear SVM [4].

Intra-personal setting estimation. In this paper, we estimate two kinds of intra-personal settings: pose and illumination. In principle, our framework is flexible to incorporate more settings or attributes such as expression, gender, and age [11, 12]. Here, we adopt a very simple but effective setting estimation method based on our generic data set. As shown in Figure 2, we denote seven pose categories and four illumination conditions as $\{P_i | i = 1, 2, 3, \dots, 7\}$ and $\{L_i | i = 1, 2, 3, 4\}$. To estimate the pose, we average the face descriptors of all the identities (in the identity data set) with the specific pose to generate a template model for each pose category. Then we measure the similarity between the input face and the template faces, and assign the pose category of the most alike template. The illumination estimation adopts the same mechanism to assign the illumination setting to each input face. We believe a more sophisticated setting classifier can be trained to obtain more accurate estimation.

4. Associate-Predict model

With the collected generic identity data set, extracted face representations, and estimated intra-personal setting, we are ready to apply the associate-predict model. In this paper, we have designed two kinds of models: "appearance-prediction" model and "likelihood-prediction" models. Two models have two similar steps: 1) associate alike generic identities for each input facial component; 2) predict the

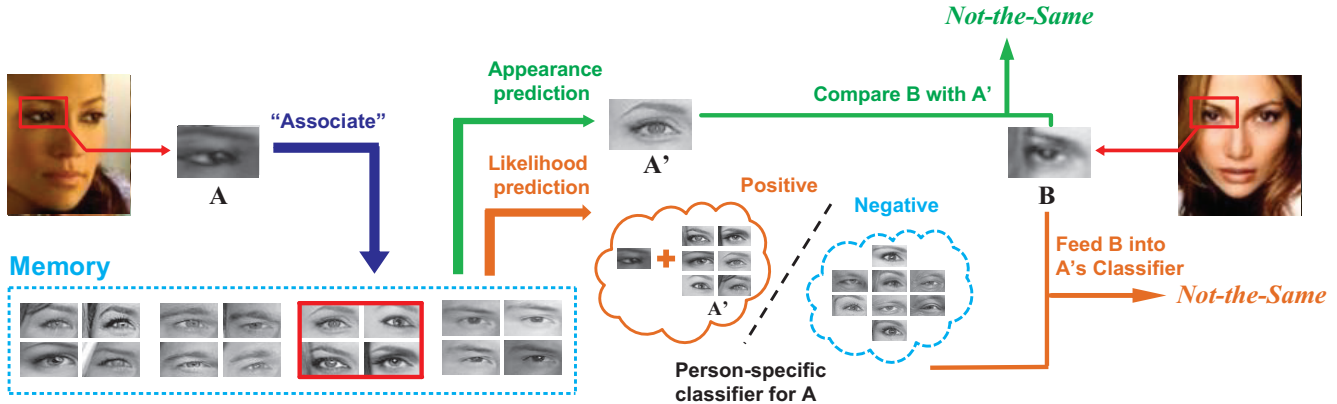


Figure 3. The framework of the “Associate-Predict” model. Given a face pair A and B, we will first “associate” A with alike generic identities based on the similarities of facial components. With the associated generic identities, we can predict the new appearance of the specific component (left eye in the figure) under a different setting (“Appearance prediction”) or directly predict the recognition output if the two components belong to the same person (“Likelihood prediction”). After implementing the “prediction” on each component, the component-level similarities are fused to give the final decision.

new appearance under a different intra-personal setting, or directly predict the recognition output by training a person-specific classifier. To associate the most alike identities, we compute the descriptor distances between the input component to all the (corresponding) components of each generic identity, and treat the averaged distance as the similarity to the specific generic identity. Since our models are working at the component level, we may associate different alike generic identities for different components. Next, we introduce the details of the two prediction models.

4.1. "Appearance-prediction" model

For two input faces, suppose s_A and s_B are their estimated intra-personal settings, and A and B are two corresponding facial components (e.g. two left eyes shown in Figure 3). For component A, we find its “nearest” generic identity which has the most alike (corresponding) component as A. Here we use L2 distance of the extracted face descriptors to measure the similarity. As introduced before, each generic identity contains 28 (7 poses \times 4 illuminations) face images. From all its 28 face images, we select the specific face image whose intra-personal setting is equal to s_B . Then, we pick the corresponding component A’ from this image. The new component A’ is indeed the “predicted” appearance of the component A from setting s_A to s_B , using A’s nearest generic identity as a bridge.

With the predicted component A’, any existing appearance-based matching method can be applied to calculate the distance $d_A = |f_{A'} - f_B|$ between the components A’ and B. Symmetrically, the distance $d_B (= |f_{B'} - f_A|)$ between B’s predicted component B’ and A can be obtained. Finally, we can simply use the average of the two distances as the final distance between A and B: $\frac{1}{2}(d_A + d_B)$.

However, we found that the discriminative power of d_A (or d_B) may be affected by the ability whether we can find a sufficiently similar identity from our limited data set. For example, if the distance between A’ and A is much smaller than the distance between B’ and B, it makes sense to trust more on d_A , rather than simply to take the average. Therefore, we design an adaptive distance d_p :

$$d_p = \frac{1}{\alpha_A + \alpha_B} (\alpha_A d_A + \alpha_B d_B), \quad (1)$$

where $\alpha_A = e^{-\gamma|f_A - f_{A'}|}$ and $\alpha_B = e^{-\gamma|f_B - f_{B'}|}$ are weights, and γ is a control parameter (default value is set to 5). The distances (d_A and d_B) should be normalized to the likelihood value between 0 and 1 before the adaptive fusion. With this adaptive fusion mechanism, the prediction-based matching will work well as long as any one of the input faces can find a good “nearest” generic identity.

After the “appearance-prediction” on all 12 facial components, we can obtain a new composite face, as shown in Figure 4. The composite face can be treated as a good approximation of the input face’s appearance under a different intra-personal setting. As a result, the prediction-based matching can always compute the similarity distance between two faces with consistent settings, largely reducing the negative effect of the intra-personal variation.

4.2. "Likelihood-prediction" model

In the likelihood-prediction model, we fix 20 generic identities to be the negative samples, and treat the rest generic identities as the candidates for positive sample augmentation. Given two input components A and B, we first build a person-specific classifier for A and measure the likelihood of B belonging to A. We use the binary LDA classi-

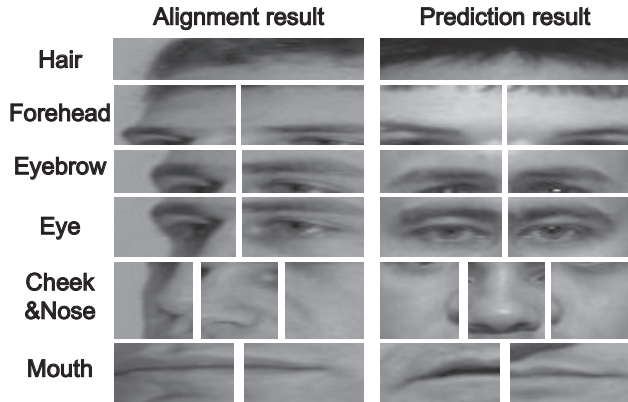


Figure 4. The effect of “appearance-prediction”. The “Alignment result” shows the original components after alignment, while the “Prediction result” demonstrates the composite face representation after “appearance-prediction” on each component. As a result, the pose category of the input face is transited from P_2 (left-oriented) to P_4 (frontal).

fier [19] as our basic classifier. The input facial component A together with all component images (under all intra-personal settings) of the K most alike generic identities are augmented to form the positive training samples. We will discuss the selection of K in Section 6.

After the training, the component B is fed into the classifier and calculate the likelihood distance $d_A \in [0, 1]$. The same process can be applied to train another classifier to get the distance d_B . The final distance is also computed using Equation (1). Note that we use the averaged distances between A (or B) and its top K nearest identities to compute the adaptive weight α_A (or α_B).

In both prediction models, we use a linear SVM [4] to fuse the component-level distances to obtain the final distance between the whole faces.

5. Switching Mechanism

The associate-predict model has its own risk - it may reduce the inter-personal difference; it may be even worse than the direct appearance-based methods, especially when the input faces have very similar intra-personal settings. Relying on a transition through a not-exact same identity, the discriminative power may be decreased.

In this paper, we make another conjecture: *our brain adaptively switches between the direct matching and “Associate-Predict” based matching, based on the intra-personal settings*. In other words, when two faces with close intra-personal settings, the direct appearance comparison is enough to convince us if the two faces are from the same identity or not; otherwise, the matching strategy should be switched to a kind of prediction-based matching built on prior knowledge (“memories”).

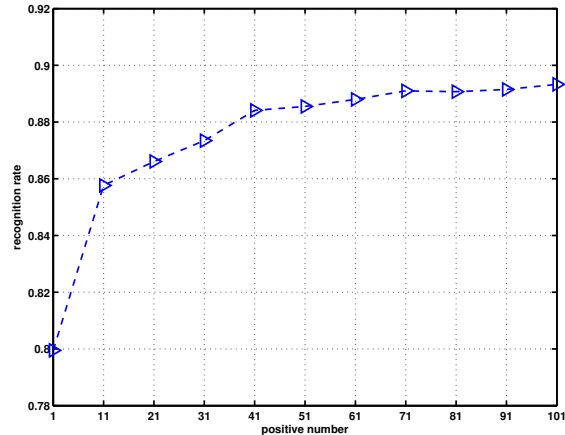


Figure 5. The effect of positive sample number. We extensively studied the recognition performance of our “likelihood-prediction” model with different numbers of associated positive samples. The experiment is conducted on the Multi-PIE benchmark, adopting LBP descriptor as the face representation.

To implement this switching mechanism, we apply a very simple switch model as follows. Given two facial components (A and B) and their settings ($s_A = \{P_A, L_A\}$ and $s_B = \{P_B, L_B\}$), we categorize the input pair into two classes: “comparable” or “not comparable”, based on the difference of s_A and s_B . We define the pair with $\{|P_A - P_B| < 3\}$ and $\{|L_A - L_B| < 3\}$ as “comparable”. The rest situations are treated as “not comparable”. The final matching distance d_{sw} can be computed by a “hard” switch model:

$$d_{sw} = \begin{cases} d_a & \text{if “comparable”} \\ d_p & \text{otherwise} \end{cases}, \quad (2)$$

where d_a and d_p are distances from the direct appearance matching and the associate-predict model.

We have also tried a “soft” switch model, which adaptively fuses d_a and d_p :

$$d_{sw} = \begin{cases} \alpha_1 * d_a + (1 - \alpha_1) * d_p & \text{if “comparable”} \\ \alpha_2 * d_a + (1 - \alpha_2) * d_p & \text{otherwise} \end{cases}, \quad (3)$$

where α_1 and α_2 are two weighting parameters learned from the training data. However, we found the “soft” switch model is only marginally better in our experiments. For simplicity, we adopt the “hard” version as our default model.

The above switching mechanism can greatly reduce the risk of inaccurate association/prediction caused by the limited identity data set. Compared with the large data volume of human memory, our current prior knowledge (200 identities) is still like an infant. We believe there is a long way to go to build a better “machine memory”.

Descriptor	LBP	SIFT	Gabor	LE
direct	80.55%	78.30%	81.00%	84.20%
appearance (H)	83.85%	82.40%	83.95%	87.55%
appearance (C)	86.75%	86.65%	86.80%	89.75%
likelihood (H)	85.65%	83.45%	87.05%	87.40%
likelihood (C)	89.05%	87.60%	89.85%	92.25%

Table 1. Holistic vs. component in the association. The “direct” is the direct appearance-based matching. The “appearance” and “likelihood” are our two prediction models. “H” and “C” in brackets separately refer to search the alike generic identities at the holistic-face level or the component level.

6. Experimental Results

We conduct experiments on the Multi-PIE and LFW data sets. For the experiment on the Multi-PIE, we select 49 test identities which are exclusive to our identity data set. We randomly select 10 mutual exclusive folders, with 300 intra-personal and 300 extra-personal pairs in each folder. The test task is to tell whether the face pair comes from the same person. The cross-validation evaluation is adopted among these 10 folders, and we report the averaged results. As for the LFW data set, we follow the restricted protocol of the LFW benchmark for evaluation. Note that the LFW benchmark is quite different from the Multi-PIE benchmark, which verifies the generalization ability of our model.

6.1. Basic comparisons

Holistic vs. Component. First, we measure the importance of the component-level association. Using the component enables the system to have larger chances to find a more alike generic identity and thus reduce the risk of inaccurate association. We separately use the holistic level association and the component level association to find the most similar generic identities. After getting associated identities, the remaining processing is the same for the two methods. The results shown in Table 1 demonstrated that the component level association is consistently better.

Positive sample size. Next, we investigate the effect of the positive sample number. The number of positive samples is $1 + 28 * k$, where the “1” is the input sample and K is the selected number of top-alike associated identities. As shown in Figure 5, we varied the number of positive samples from 1 to 100, and plotted the recognition rate of our “likelihood prediction”. Notice that if the associated number is not a multiple of 28, a subset of face images will be randomly chosen from the K th selected generic identity. We observed that the performance increases with the positive sample number, and roughly remains the same after $K = 3$. Thus, we adopt $K = 3$ as the default parameter in the following experiments. Compared with single positive

Descriptor	LE on MPIE	LE on LFW
direct	84.20%	82.33%
appearance (no switching)	86.95%	82.95%
appearance (switching)	89.30%	88.16%
likelihood (no switching)	89.75%	84.30%
likelihood (switching)	92.25%	89.25%

Table 2. Effects of the switching mechanism in different models. The “direct” is the direct appearance-based matching, and “appearance” and “likelihood” are our two prediction models.

sample method, the large gain (about 10% improvement) revealed the significance of having an prior identity data set with large intra-personal variation.

Switching mechanism. Last, we study the function of the switching mechanism. We report the recognition rate on both Multi-PIE and LFW benchmarks using the LE descriptor. Table 2 shows that the switch model can effectively improve the results on both benchmarks. The effect of the switch model is more significant on the LFW benchmark because the face images on the LFW benchmark are more different from our generic identity data set. This result validates that our switch model is powerful to handle the limited size issue of our identity data set.

6.2. Results on benchmarks

Multi-PIE benchmark. Figure 6 compares the direct appearance-based matching and our two models, using LBP descriptor and LE descriptor. Both descriptors are compressed to 300 dimension at each component using the PCA technique described in [3]. The resulting curves clearly show that our framework can dramatically improve the recognition performance of both descriptors. Taking the LE descriptor as example, the recognition rate (equal error point) of direct appearance-based matching is 82.08%, while our model can achieve 94.03% - reducing nearly 60% errors! To get the best result, we further fuse the distances of two models by a linear SVM (using the LE descriptor as the representation). It reveals a certain complementariness between the two models.

LFW benchmark. The LFW benchmark defines two evaluation protocols: the restricted setting (with a fixed number of intra-personal and extra-personal pairs provided for training) and the unrestricted one (where arbitrary number of training pairs can be generated based on the given faces’ labels). With sufficient training samples, the unrestricted protocol is easier to produce high performance. Strictly speaking, our approach does not follow either protocols because we leverage an additional identity data set. However, unlike the unrestricted protocol, we do not use the extra training data provided in LFW. The goals of our experiments on this benchmark are twofold: 1) verify the generalization ability; 2) to see how good the face recognition could be if we want

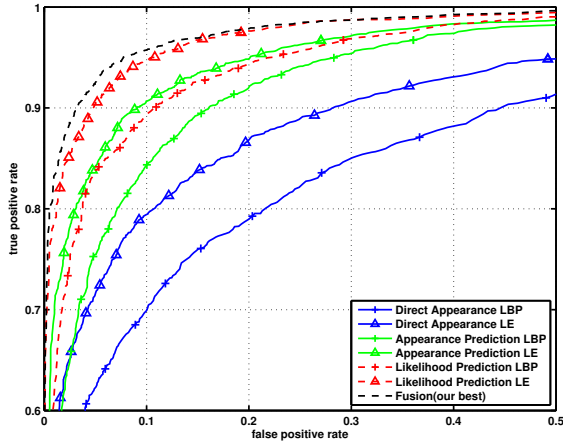


Figure 6. Effects of the proposed techniques on the Multi-PIE benchmark.

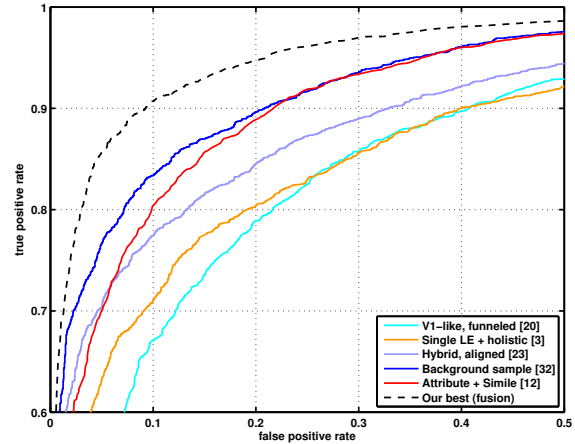


Figure 8. Recognition performance comparison on the LFW benchmark with results in restricted protocol.

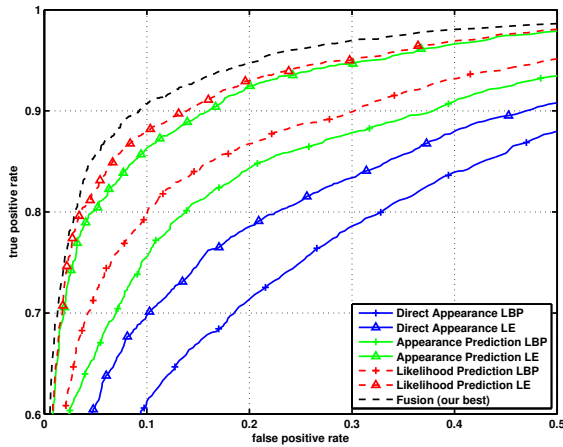


Figure 7. Effects of the proposed techniques on the LFW benchmark.

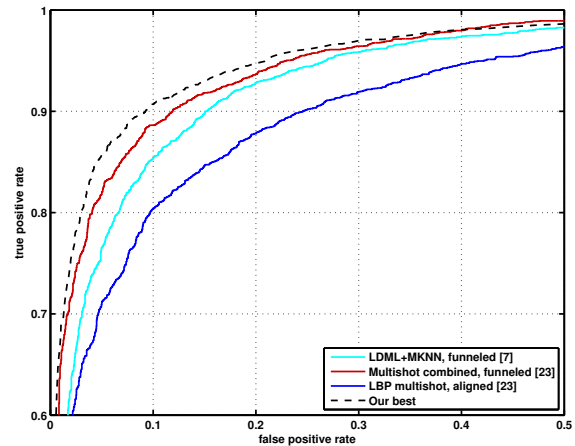


Figure 9. Recognition performance comparison on the LFW benchmark with results in unrestricted protocol.

to use the prior knowledge.

In our experiment, beside the additional identity data set, we strictly follow the restricted protocol, including PCA learning and SVM training. In Figure 7, we repeat the same comparison on the LFW data set. Again, we can see our associate-predict model significantly boost the recognition performance of the two descriptors. Furthermore, Figure 8 and Figure 9 show our results along with current state-of-the-art results. Under the restricted protocol, the current best method [33] can achieve 86.83% by combining multiple descriptors and leveraging the metric learning. Our best system can achieve 90.57% (reducing nearly 30% errors) by a single descriptor and without any metric learning. Under the unrestricted protocol, our best result still outperforms the best existing methods.

7. Concluding Remarks

Motivated by Jeff Hawkins’s studies on the intelligence, we have presented an “Associate-Predict” model for handling intra-personal variation in face recognition. Our model has two main advantages over the traditional appearance-based methods: 1) explicitly handle the intra-personal variation issue in a more principled way: using generic identities as a bridge; 2) use the prior knowledge adaptively by a switch model.

We have seen the power of the proposed model on one computer vision problem - face recognition. We believe that some basic elements may also be useful to other vision tasks. As more effective prior knowledge can be built and more principled image matching can be made, we hope that we will eventually be able to build intelligent vision system that can be comparable with our human vision.

References

- [1] M. Belkin and P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural computation*, 15(6), 2003.
- [2] V. Blanz and T. Vetter. Face recognition based on fitting a 3 D morphable model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(9), 2003.
- [3] Z. Cao, Q. Yin, J. Sun, and X. Tang. Face recognition with Learning-based Descriptor. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2010.
- [4] C. Chang and C. Lin. LIBSVM: a library for support vector machines, 2001. *Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>*, 2001.
- [5] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2005.
- [6] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker. Multi-PIE. In *International Conference on Automatic Face and Gesture Recognition*, 2008.
- [7] M. Guillaumin, J. Verbeek, C. Schmid, I. LEAR, and L. Kuntzmann. Is that you? Metric learning approaches for face identification. In *Proc. IEEE International Conference on Computer Vision*, 2009.
- [8] J. Hawkins and S. Blakeslee. *On Intelligence*. Times Books, Henry Holt and Company, New York, 2004.
- [9] G. Hua and A. Akbarzadeh. A robust elastic and partial matching metric for face recognition. In *Proc. IEEE International Conference on Computer Vision*, 2009.
- [10] T. Kim and J. Kittler. Locally linear discriminant analysis for multimodally distributed classes for face recognition with a single model image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005.
- [11] N. Kumar, P. Belhumeur, and S. Nayar. FaceTracer: A search engine for large collections of images with faces. *Computer Vision—ECCV 2008*, 2008.
- [12] N. Kumar, A. Berg, P. Belhumeur, and S. Nayar. Attribute and Simile classifiers for face verification. In *Proc. IEEE International Conference on Computer Vision*, 2009.
- [13] P. Lee, G. Hsu, and Y. Hung. Face Verification and Identification using Facial Trait Code. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2009.
- [14] Q. Liu, R. Huang, H. Lu, and S. Ma. Face recognition using kernel based fisher discriminant analysis. In *Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition*. IEEE Computer Society, 2002.
- [15] Q. Liu, X. Tang, H. Lu, and S. Ma. Face recognition using kernel scatter-difference-based discriminant analysis. *Neural Networks, IEEE Transactions on*, 17(4), 2006.
- [16] D. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2), 2004.
- [17] B. Moghaddam, T. Jebara, and A. Pentland. Bayesian face recognition. *Pattern Recognition*, 33(11), 2000.
- [18] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7), 2002.
- [19] N. Peter, P. João, and J. David. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7), 1997.
- [20] N. Pinto, J. DiCarlo, and D. Cox. How far can you get with a modern face recognition test set using only simple features. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2009.
- [21] S. Shan, B. Cao, W. Gao, and D. Zhao. Extended Fisherface for face recognition from a single example image per person. In *IEEE International Symposium on Circuits and Systems*, 2002.
- [22] Y. Su, S. Shan, X. Chen, and W. Gao. Adaptive generic learning for face recognition from a single sample per person. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2010.
- [23] Y. Taigman, L. Wolf, T. Hassner, and I. Tel-aviv. Multiple One-Shots for utilizing class label information. In *British Machine Vision Conference*, 2009.
- [24] X. Tan and B. Triggs. Enhanced local texture feature sets for face recognition under difficult lighting conditions. *Lecture Notes in Computer Science*, 4778, 2007.
- [25] X. Tang and X. Wang. Face sketch recognition. *Circuits and Systems for Video Technology, IEEE Transactions on*, 14, 2004.
- [26] M. Turk and A. Pentland. Face Recognition using Eigenfaces. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 1991.
- [27] J. Wang, K. Plataniotis, J. Lu, and A. Venetsanopoulos. On solving the face recognition problem with one training sample per subject. *Pattern recognition*, 39(9), 2006.
- [28] X. Wang and X. Tang. A unified framework for subspace face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(9), 2004.
- [29] X. Wang and X. Tang. Dual-space linear discriminant analysis for face recognition. 2004.
- [30] X. Wang and X. Tang. Random sampling for subspace face recognition. *International Journal of Computer Vision*, 70(1), 2006.
- [31] L. Wiskott, J. Fellous, N. Krüger, and C. Von der Malsburg. Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7), 1997.
- [32] L. Wolf, T. Hassner, and Y. Taigman. Descriptor based methods in the wild. In *Faces in Real-Life Images Workshop in ECCV*, 2008.
- [33] L. Wolf, T. Hassner, and Y. Taigman. Similarity scores based on background samples. In *Proc. Asian Conference on Computer Vision*, 2009.
- [34] J. Wright and G. Hua. Implicit elastic matching with random projections for pose-variant face recognition. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2009.
- [35] D. Zhao, Z. Lin, R. Xiao, and X. Tang. Linear Laplacian discrimination for feature extraction. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2007.