Fine-Grained Face Verification: Dataset and Baseline Results

Junlin Hu[1], Jiwen Lu[2*], Yap-Peng Tan[1]
[1]School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore
[2]Advanced Digital Sciences Center, Singapore
jhu007@e.ntu.edu.sg, jiwen.lu@adsc.com.sg, eyptan@ntu.edu.sg

Abstract

This paper investigates the problem of fine-grained face verification under unconstrained conditions. For the conventional face verification task, the verification model is trained with some positive and negative face pairs, where each positive sample pair contains two face images of the same person while each negative sample pair usually consists of two face images from different subjects. However, in many real applications, facial appearance of the twins looks very similar even if they are considered as a negative pair in face verification. Therefore, it is important to differentiate a given face pair to determine whether it is from the same person or a twins for a practical face verification system because most existing face verification systems fails to work well in such a scenario. In this work, we define the problem as fine-grained face verification and collect an unconstrained face dataset which contains 455 pairs of identical twins to generate negative face pairs to evaluate several baseline verification models for fine-grained unconstrained face verification. Benchmark results on the unsupervised setting and restricted setting show the challenge of the fine-grained face verification in the wild.

1. Introduction

Face verification, as an important task of face recognition, aims to determine whether a given pair of face images is from the same person or not. The previous face verification methods have achieved impressive performance in controlled conditions, however these methods always drop their performance if the input images are collected in uncontrolled environments [8]. In recent years, the unconstrained face verification has attracted a large attention, especially since the release of a benchmark dataset named the Labeled Faces in the Wild (LFW) dataset [9], where the face images are captured in the uncontrolled environment with significant variations such as expression, pose, partial occlusions lighting, and resolution.

Figure 1. Typical positive pairs and negative pairs sampled from the LFW dataset. Two neighbor images in each row form a face pair, and negative pair shows significant variation of appearance.

Recently, a variety of face verification approaches have been proposed to advance the performance of the unconstrained face verification [9, 6, 3, 7, 21, 13, 19, 15] on the LFW dataset. For example, Guillaumin *et al*. [6] presented a logistic discriminant method to learn a Mahalanobis distance from labelled image pairs in the probabilistic view, which expects the distance between a positive pair to be smaller than that of a negative pair. Cui *et al*. [3] presented a robust spatial face region descriptor under the Bag-of-Feature framework as well as a pairwise-constrained multiple metric learning to learn multiple distance metrics for various descriptors. Hu *et al*. [7] proposed a discriminative deep metric learning method to learn a nonlinear distance metric via a architecture of neural network for face verification in the wild. Taigman *et al*. [21] developed a deep neural network to learn a face representation from a outside training dataset with four million face images, and obtained an accuracy of 97.25%, closely approaching human-level performance on the LFW dataset. Sun *et al*. [19] designed a deep convolutional networks model to learn deep identification-verification features (DeepID2) by using both face identification and verification signals as supervision,

*Corresponding author.

and achieved 99.15% verification accuracy under the standard LFW protocol. However, these high accuracies are also high in the underlying false accept rate (FAR), which is not reliable for for most practical applications. Liao *et al*. [13] developed a new benchmark protocol based on this dataset, under which the best approach obtains 41.66% verification accuracy at FAR $= 0.1\%$. These results show that unconstrained face recognition is still a very challenging problem.

However, most existing face verification methods are learned with some positive and negative face pairs, where each positive sample pair contains two face images of the same person while each negative sample pair usually consists of two face images from different subjects. Figure 1 shows some positive and negative pairs in the LFW protocol, two images of a negative pair are different in appearance even gender. However, in many real applications, facial appearance of the twins looks very similar even if they are considered as a negative pair in face verification. Therefore, it is important to differentiate a given face pair to determine whether it is from the same person or twins for a practical face verification system because most existing face verification systems fail to work well in such a scenario. In this paper, we collect a new unconstrained face dataset which contains 455 pairs of identical twins to form negative face pairs with quite similar appearance for the challenging fine-grained unconstrained face verification, as well as develop a benchmark protocol for evaluating several baseline verification approaches for this problem. Different from the works [20, 18, 11, 17] in distinguishing between identical twins by face recognition under controlled conditions, we aim to determine whether a face pair depicts the same person or not in the scenario that two images in each negative pairs are very similar in appearance. To our best knowledge, this is the first study on the problem of fine-grained face verification in the wild.

2. Dataset and Protocol

This section will detail a fine-grained face verification (FGFV) dataset and benchmark protocol on this dataset.

2.1. Dataset

To advance the fine-grained face verification under uncontrolled environments, we collected a new dataset named fine-grained face verification (FGFV) dataset. The FGFV dataset contains 1820 face images, namely, 455 negative face pairs and 455 positive face pairs, respectively. The 455 negative face pairs are from 455 pairs of identical twins (910 subjects) collected from the internet without restrictions in lighting, expression, pose background, and partial occlusion. A significant characteristic of the negative pair is that two images of this pair are from the same photo with almost the same lighting. Figure 2 lists sample negative face pairs of the FGFV dataset. For 455 positive pairs of FGFV dataset, they were selected from positive (or matched) pairs of the LFW dataset [9], where few people have more than one face pair. Figure 3 shows some positive face pairs in the FGFV dataset. In addition, to reduce background information and easily use this dataset, we also offer a well-aligned version by aligning and cropping each face image to size of 64×64 according to the manually labeled eye coordinates. Figures 4 and 5 list some cropped samples of negative and positive pairs on FGFV dataset, respectively.

Figure 2. Some negative face pairs of the FGFV dataset. A negative pair consists of two neighbor twin images in each row, and two images of each pair pose very similar appearance.

Figure 3. Some positive face pairs of the FGFV dataset. A positive pair consists of two neighbor images of same person in each row.

2.2. Benchmark Protocol

Following the widely used benchmark protocol on the LFW dataset [9], we designed two different settings to evaluate the performance of different verification algorithm-

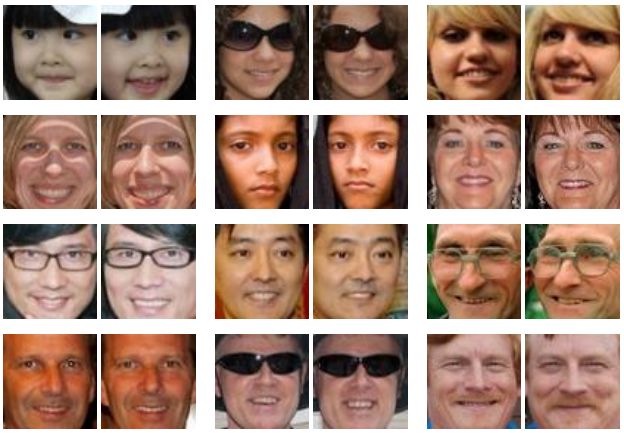


Figure 4. Cropped negative face pairs used in our experiments.

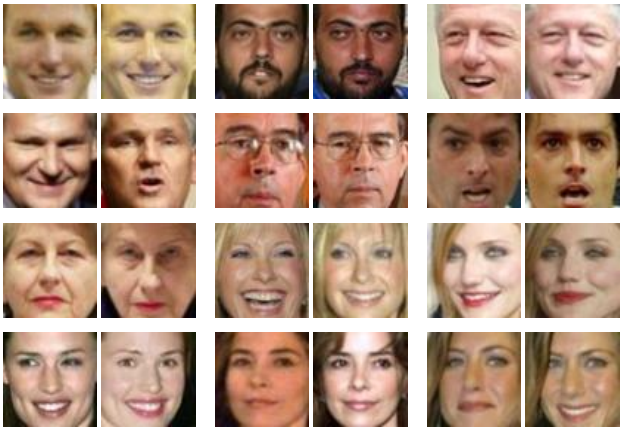


Figure 5. Cropped positive face pairs used in our experiments.

s: Unsupervised setting and Restricted setting. As a benchmark for comparison, we randomly partitioned all face pairs of the FGFV dataset into 5 independent folds, where each fold contains 91 negative pairs and 91 positive pairs. We also recommend reporting performance as 5-fold cross validation using this pre-specified training/testing splits, namely, 4 folds for training and the rest one for testing each time under the following two settings.

Unsupervised setting: Under this setting, a verification method cannot have access to any label information of face images, such as pairwise label (*positive* or *negative*), unique identifiers of persons, and the distribution of labels. The receiver operating characteristic (ROC) curve, the area under the ROC curve (AUC), equal error rate (EER) and verification accuracy are used for performance evaluation.

Restricted setting: In the restricted setting, only the pairwise labels (*positive* or *negative*) of face pairs are used for training verification models. Finally, the mean accuracy with standard error and ROC curve are adopted for performance evaluation.

3. Baseline Approaches

In this section, we briefly elaborate several baseline methods for fine-grained face verification problem in the perspective of similarity or metric learning, these approaches include information theoretic metric learning (ITML) [5], KISS metric learning (KISSME) [12], side-information based linear discriminant analysis (SILD) [10], and cosine similarity metric learning (CSML) [16].

3.1. ITML

Given a training set $\mathbf{X} = [x_1, x_2, \cdots, x_N] \in \mathbb{R}^{d\times N}$ consisting of N samples, the goal of conventional distance metric learning methods is to look for a positive semi-definite matrix $\mathbf{M} \in \mathbb{R}^{d\times d}$ so that the squared Mahalanobis distance between two samples $\mathbf{x}_i$ and $\mathbf{x}_j$ is calculated as:

$$d^2_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i - \mathbf{x}_j)^T\mathbf{M}(\mathbf{x}_i - \mathbf{x}_j), \qquad (1)$$

where d is the dimension of the input sample $\mathbf{x}_i$.

Information-Theoretic Metric Learning (ITML) [5] utilizes the relationship between the multivariate Gaussian distribution and the set of Mahalanobis distances to generalize the standard Euclidean distance. The idea of ITML is to seek a distance metric $\mathbf{M}$ to close a prior metric $\mathbf{M}_0$ by minimizing the LogDet divergence between these two matrices under the constraints that the distance $d^2_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_j)$ between a positive pair is smaller than a upper bound τ_p while that of negative pair is larger than a lower bound τ_n, $\tau_n > \tau_p > 0$. Then, ITML is formulated as the following LogDet optimization problem:

$$\begin{aligned} \min_{\mathbf{M}} \; & D_{ld}(\mathbf{M}, \mathbf{M}_0) = \mathrm{tr}(\mathbf{M}\mathbf{M}_0^{-1}) - \log\det(\mathbf{M}\mathbf{M}_0^{-1}) - d \\ \text{s.t.} \; & d^2_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_j) \le \tau_p \qquad \forall\, \ell_{ij} = 1 \\ & d^2_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_j) \ge \tau_n \qquad \forall\, \ell_{ij} = -1, \end{aligned} \qquad (2)$$

where the prior $\mathbf{M}_0$ is set to the identity matrix for the Euclidean distance; $\mathrm{tr}(\mathbf{Z})$ is the *trace* of matrix $\mathbf{Z}$; and ℓ_{ij} denotes pairwise label between a pair of samples $\mathbf{x}_i$ and $\mathbf{x}_j$, which is set as $\ell_{ij} = 1$ for a positive pair and $\ell_{ij} = -1$ for a negative pair. To optimize this problem (2), Bregman projections are iteratively computed to project the current solution onto a single constraint by the update:

$$\mathbf{M}_{t+1} = \mathbf{M}_t + \beta\, \mathbf{M}_t(\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T\mathbf{M}_t, \qquad (3)$$

where β is the projection parameter dominated by pairwise label and learning rate.

3.2. KISSME

KISS Metric Learning (KISSME) [12] considers learning a simple and straightforward distance metric from a statistical inference perspective, which makes the statistical decision whether a sample pair $\mathbf{x}_i$ and $\mathbf{x}_j$ is negative or not

by a likelihood ratio test. It states the hypothesis $\mathcal{H}_0$ that a pair is negative, and the $\mathcal{H}_1$ that this pair is positive. The log-likelihood ratio is given as:

$$\delta(\mathbf{x}_i, \mathbf{x}_j) = \log\left(\frac{p(\mathbf{x}_i, \mathbf{x}_j|\mathcal{H}_0)}{p(\mathbf{x}_i, \mathbf{x}_j|\mathcal{H}_1)}\right), \tag{4}$$

where $p(\mathbf{x}_i, \mathbf{x}_j|\mathcal{H}_0)$ is probability distribution function of a pair under the $\mathcal{H}_0$. The $\mathcal{H}_0$ is accepted if $\delta(\mathbf{x}_i, \mathbf{x}_j)$ is larger than a threshold, otherwise the $\mathcal{H}_1$ is validated and this pair is positive. By assuming the single Gaussian distribution of the pairwise difference $\mathbf{z}_{ij} = \mathbf{x}_i - \mathbf{x}_j$ and relaxing the problem (4), $\delta(\mathbf{x}_i, \mathbf{x}_j)$ is simplified as:

$$\delta(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i - \mathbf{x}_j)^T(\mathbf{C}_p^{-1} - \mathbf{C}_n^{-1})(\mathbf{x}_i - \mathbf{x}_j), \tag{5}$$

where covariance matrices $\mathbf{C}_p$ and $\mathbf{C}_n$ are estimated as:

$$\mathbf{C}_p = \sum\nolimits_{\ell_{ij}=1} (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T, \tag{6}$$

$$\mathbf{C}_n = \sum\nolimits_{\ell_{ij}=-1} (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T. \tag{7}$$

To obtain the Mahalanobis distance metric $\mathbf{M}$, KISSME projects $\hat{\mathbf{M}} = \mathbf{C}_p^{-1} - \mathbf{C}_n^{-1}$ onto the cone of the positive semi-definite matrix $\mathbf{M}$ by clipping the spectrum of $\hat{\mathbf{M}}$.

3.3. SILD

Side-Information based Linear Discriminant Analysis (SILD) [10] effectively exploits the weakly-supervised side-information of sample pairs to directly calculate the within-class scatter matrix $\mathbf{C}_p$ in Eqn. (6) using positive pairs and the between-class scatter matrix $\mathbf{C}_n$ in Eqn. (7) utilizing negative pairs. Similarly to fisher linear discriminant analysis (FLDA) [2], SILD seeks a discriminative linear projection $\mathbf{W} \in \mathbb{R}^{d\times m}$, $m \leq d$ by solving the optimization problem as:

$$\max_{\mathbf{W}} \frac{\det(\mathbf{W}^T\mathbf{C}_n\mathbf{W})}{\det(\mathbf{W}^T\mathbf{C}_p\mathbf{W})}. \tag{8}$$

By diagonalizing $\mathbf{C}_p$ and $\mathbf{C}_n$ as:

$$\mathbf{C}_p = \mathbf{U}\mathbf{D}_p\mathbf{U}^T, \quad (\mathbf{U}\mathbf{D}_p^{-1/2})^T\mathbf{C}_p(\mathbf{U}\mathbf{D}_p^{-1/2}) = \mathbf{I}, \tag{9}$$

$$(\mathbf{U}\mathbf{D}_p^{-1/2})^T\mathbf{C}_n(\mathbf{U}\mathbf{D}_p^{-1/2}) = \mathbf{V}\mathbf{D}_n\mathbf{V}^T, \tag{10}$$

the projection matrix $\mathbf{W}$ can be obtained as:

$$\mathbf{W} = \mathbf{U}\mathbf{D}_p^{-1/2}\mathbf{V}, \tag{11}$$

where $\mathbf{U}$ and $\mathbf{V}$ are orthogonal matrices, and $\mathbf{D}_p$ and $\mathbf{D}_n$ are two diagonal matrices. Then, the squared Euclidean distance is computed in the transformed subspace as:

$$\begin{aligned} d^2_{\mathbf{W}}(\mathbf{x}_i, \mathbf{x}_j) &= \left\|\mathbf{W}^T\mathbf{x}_i - \mathbf{W}^T\mathbf{x}_j\right\|_2^2 \\ &= (\mathbf{x}_i - \mathbf{x}_j)^T\mathbf{W}\mathbf{W}^T(\mathbf{x}_i - \mathbf{x}_j) \\ &= (\mathbf{x}_i - \mathbf{x}_j)^T\mathbf{M}(\mathbf{x}_i - \mathbf{x}_j), \end{aligned} \tag{12}$$

which is equivalent to the Mahalanobis distance in the original space, $\mathbf{M} = \mathbf{W}\mathbf{W}^T$.

3.4. CSML

Cosine Similarity Metric Learning (CSML) [16] aims to learn a linear transformation $\mathbf{W} \in \mathbb{R}^{d\times m}$, $m \leq d$ to calculate cosine similarity in this transformed subspace:

$$\begin{aligned} cs_{\mathbf{W}}(\mathbf{x}_i, \mathbf{x}_j) &= \frac{(\mathbf{W}^T\mathbf{x}_i)^T(\mathbf{W}^T\mathbf{x}_j)}{\|\mathbf{W}^T\mathbf{x}_i\|\,\|\mathbf{W}^T\mathbf{x}_j\|} \\ &= \frac{{\mathbf{x}_i}^T\mathbf{W}\mathbf{W}^T\mathbf{x}_j}{\sqrt{{\mathbf{x}_i}^T\mathbf{W}\mathbf{W}^T\mathbf{x}_i}\sqrt{{\mathbf{x}_j}^T\mathbf{W}\mathbf{W}^T\mathbf{x}_j}}. \end{aligned} \tag{13}$$

To learn $\mathbf{W}$, CSML minimizes the cross validation error by formulating the objective function as:

$$\begin{aligned} \max_{\mathbf{W}}\ F(\mathbf{W}) = &\sum_{\ell_{ij}=1} cs_{\mathbf{W}}(\mathbf{x}_i, \mathbf{x}_j) \\ &- \alpha \sum_{\ell_{ij}=-1} cs_{\mathbf{W}}(\mathbf{x}_i, \mathbf{x}_j) - \beta\|\mathbf{W} - \mathbf{W}_0\|^2, \end{aligned} \tag{14}$$

where $\mathbf{W}_0$ is a predefined matrix; parameter α balances the contributions of positive and negative pairs to margin; and β controls the tradeoff between margin and regularization term $\|\mathbf{W} - \mathbf{W}_0\|^2$. Finally, a gradient based method is used to obtain the optimal $\mathbf{W}$, and more details of the optimization are provided in the work [16].

4. Baseline Results

Following the benchmark protocol in the FGFV dataset, we evaluate several methods for fine-grained face verification under the unsupervised setting and restricted setting.

4.1. Experimental Settings

For feature representation, we use the aligned version of the FGFV dataset, where each face image was aligned and cropped to size of 64×64 according to the manually labeled eye coordinates. Furthermore, these cropped images were converted to grayscale and we extracted three types of features for each face image as follows:

- Histogram of Oriented Gradients (HOG) [4]: each face image was first divided into 16×16 non-overlapping blocks, where the size of each block is 4×4. Then it was evenly partitioned into 8×8 non-overlapping blocks again, where the size of each block is 8×8. Lastly, we extracted a 9-bin HOG descriptor for each block and concatenated them to form a 2880-dimensional feature vector;

- Local Binary Patterns (LBP) [1]: we partitioned each image into 8×8 non-overlapping blocks of size 8×8, then computed a 59-bin uniform pattern LBP descriptor for each block and concatenated all descriptors to form a 3776-dimensional feature vector;

Table 1. The verification accuracy, AUC and EER of three features on the FGFV dataset under the unsupervised setting.

Feature	Dim.	Accuracy (%)	AUC (%)	EER (%)
HOG	2880	60.44	62.59	40.44
LBP	3776	59.67	62.92	41.32
SIFT	6272	58.79	61.23	42.45

- Scale-Invariant Feature Transform (SIFT) [14]: we densely sampled SIFT descriptors on each 16×16 patch with stepsize of 8 pixels and obtained 49 SIFT descriptors. Then extracted a 6272-dimensional feature by concatenating these SIFT descriptors.

For experimental settings, the cosine similarity is used to directly compute the similarity of a given face pair for evaluating the performance of above three features in the unsupervised setting. For the restricted setting, these three kinds of features are first reduced to 200 dimensions by PCA learnt on training set. Then four weakly supervised baseline methods (ITML, KISSME, SILD, and CSML) are utilized to learn distance metrics for each feature on the training set, and the learned distance metric is applied to compute similarity of each testing pair $\mathbf{x}_i$ and $\mathbf{x}_j$ as $sim(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-d^2_{\mathrm{M}}(\mathbf{x}_i, \mathbf{x}_j)/\gamma\right)$ for three baseline methods: ITML, KISSME and SILD, where γ is a positive scalar. We implement these methods with various features by carefully tuning the parameters of these algorithms, especially setting $\alpha = 1$ and $\beta = 0.1$ for CSML method. Lastly, the mean accuracy and ROC curves are used for performance evaluation.

4.2. Results on Unsupervised Setting

We first compare the three types of features, HOG, LBP, and SIFT for fine-grained face verification under unsupervised setting. Table 1 shows the verification results of different types of features on the FGFV dataset, and Figure 6 plots the ROC curves of these three features. From this table, we see that the best result under unsupervised setting is only 60.44% in verification accuracy, and 62.92% in the AUC. We find these three hand-crafted features show similar ability of face verification. Therefore, the fine-grained unconstrained face verification is a very challenging task.

4.3. Results on Restricted Setting

We also compare several weakly supervised similarity learning based methods with different kinds of features under the restricted setting. Table 2 shows performance of the four weakly supervised methods under the this setting. ITML obtains 64.29% verification accuracy using SIFT feature; KISSME achieves 68.35% – 69.67% mean accuracy in three different features; SILD shows good performance using HOG and SIFT features than LBP; and CSML using

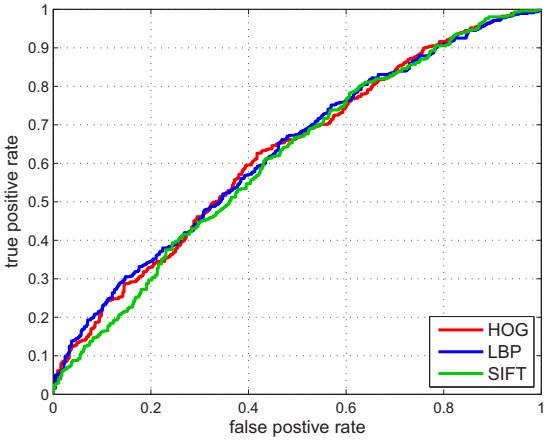


Figure 6. The ROC curves of different features on the FGFV dataset underthe unsupervised setting.

Table 2. The verification results in terms of accuracy (%) on the FGFV dataset under the restricted setting.

Method	HOG	LBP	SIFT
ITML	63.52 ± 4.41	62.86 ± 3.84	64.29 ± 4.29
KISSME	69.67 ± 3.37	68.35 ± 3.26	69.67 ± 3.40
SILD	70.00 ± 3.68	62.53 ± 3.17	68.57 ± 3.53
CSML	71.43 ± 1.94	72.31 ± 3.53	72.31 ± 3.24

LBP feature obtains the mean accuracy of 72.31%, which is better than other three compared baseline approaches. These methods consistently improve the performance of face verification than that under unsupervised setting. These results show that learning a distance metric can help the challenge of fine-grained face verification in the wild. In addition, Figure 7 also plots the ROC curves of these baseline methods with different features on the FGFV dataset under the restricted setting.

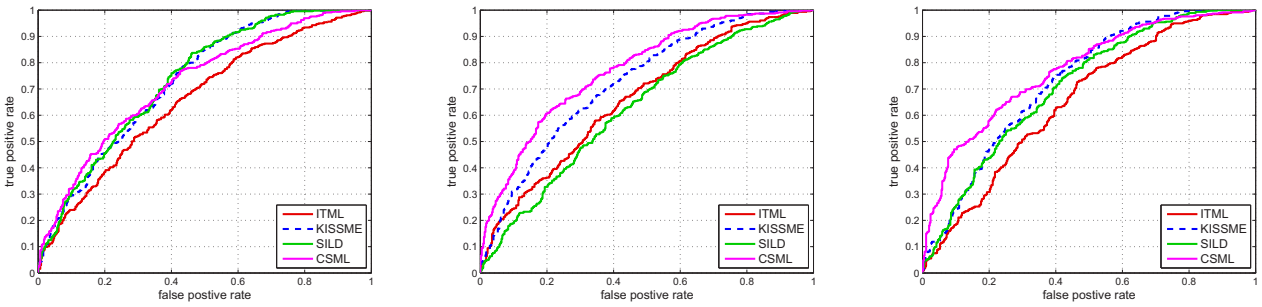
4.4. Discussion

From the above experimental results, we make the following three observations:

1) Fine-grained face verification under unsupervised setting is more challenging than the restricted setting. This is because some discriminative information has been exploited in the restricted setting.

2) HOG and SIFT slightly outperform LBP in the task of fine-grained face verification. However, the difference is not significant. The reason is that for fain-grained face verification, face images look very similar for both positive and negative pairs so that the texture feature descriptor such as LBP cannot perform well.

3) CSML achieves the best verification rate than other



(a) HOG (b) LBP (c) SIFT

Figure 7. The ROC curves of the baseline method with three different features on the FGFV dataset under the restricted setting.

metric learning methods, which indicates that the cosine similarity is better than the Mahalanobis distance to learn the similarity for fine-grained face verification because face samples are better normalized by the cosine similarity, which can remove some face variations such as varying illuminations.

5. Conclusion

In this paper, we have investigated the problem of fine-grained face verification under unconstrained conditions. Unlike the conventional face verification task that a negative sample pair usually consists of two face images from different subjects with large appearance difference, we collected a new unconstrained face dataset which contains 455 pairs of identical twins to generate negative face pairs, and evaluated several baseline verification approaches for fine-grained unconstrained face verification under a benchmark protocol. Benchmark results on both the unsupervised setting and restricted setting show the grand challenges of the fine-grained face verification.

Acknowledgement

This work is partly supported by a research grant for the Human Centric Cyber Systems (HCCS) Program at the Advanced Digital Sciences Center (ADSC) from the Agency for Science, Technology and Research (A*STAR) of Singapore.

References

[1] T. Ahonen, A. Hadid, and M. Pietikäinen. Face description with local binary patterns: Application to face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(12):2037–2041, 2006.

[2] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720, July 1997.

[3] Z. Cui, W. Li, D. Xu, S. Shan, and X. Chen. Fusing robust face region descriptors via multiple metric learning for face recognition in the wild. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3554–3561, 2013.

[4] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 886–893, 2005.

[5] J. V. Davis, B. Kulis, P. Jain, S. Sra, and I. S. Dhillon. Information-theoretic metric learning. In *International Conference on Machine Learning*, pages 209–216, 2007.

[6] M. Guillaumin, J. J. Verbeek, and C. Schmid. Is that you? metric learning approaches for face identification. In *IEEE International Conference on Computer Vision*, pages 498–505, 2009.

[7] J. Hu, J. Lu, and Y. Tan. Discriminative deep metric learning for face verification in the wild. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1875–1882, 2014.

[8] G. Hua, M. Yang, E. G. Learned-Miller, Y. Ma, M. Turk, D. J. Kriegman, and T. S. Huang. Introduction to the special section on real-world face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(10):1921–1924, 2011.

[9] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst, October 2007.

[10] M. Kan, S. Shan, D. Xu, and X. Chen. Side-information based linear discriminant analysis for face recognition. In *British Machine Vision Conference*, pages 1–12, 2011.

[11] B. Klare, A. A. Paulino, and A. K. Jain. Analysis of facial features in identical twins. In *IAPR International Conference on Biometrics*, pages 1–8, 2011.

[12] M. Köstinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof. Large scale metric learning from equivalence constraints. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2288–2295, 2012.

[13] S. Liao, Z. Lei, D. Yi, and S. Z. Li. Deepface: Closing the gap to human-level performance in face verification. In *International Joint Conference on Biometrics*, pages 1–8, 2014.

[14] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[15] J. Lu, G. Wang, W. Deng, and K. Jia. Reconstruction-based metric learning for unconstrained face verification. *IEEE Transactions on Information Forensics and Security*, 10(1):79–89, 2015.

[16] H. V. Nguyen and L. Bai. Cosine similarity metric learning for face verification. In *Asian Conference on Computer Vision*, pages 709–720, 2010.

[17] J. R. Paone, P. J. Flynn, P. J. Phillips, K. W. Bowyer, R. W. V. Bruegge, P. Grother, G. W. Quinn, M. T. Pruitt, and J. M. Grant. Double trouble: Differentiating identical twins by face recognition. *IEEE Transactions on Information Forensics and Security*, 9(2):285–295, 2014.

[18] P. J. Phillips, P. J. Flynn, K. W. Bowyer, R. W. V. Bruegge, P. Grother, G. W. Quinn, and M. Pruitt. Distinguishing identical twins by face recognition. In *IEEE International Conference on Automatic Face and Gesture Recognition*, pages 185–192, 2011.

[19] Y. Sun, Y. Chen, X. Wang, and X. Tang. Deepface: Closing the gap to human-level performance in face verification. In *Advances in Neural Information Processing Systems*, 2014.

[20] Z. Sun, A. A. Paulino, J. Feng, Z. Chai, T. Tan, and A. K. Jain. A study of multibiometric traits of identical twins. In *Biometric Technology for Human Identification VII*, 2010.

[21] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf. Deepface: Closing the gap to human-level performance in face verification. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1701–1708, 2014.