Partial Face Matching between Near Infrared and Visual Images in MBGC Portal Challenge

Dong Yi, Shengcai Liao, Zhen Lei, Jitao Sang, and Stan Z. Li⋆

Center for Biometrics and Security Research,
Institute of Automation, Chinese Academy of Sciences,
95 Zhongguancun Donglu, Beijing 100190, China
{dyi,scliao,zlei,jtsang,szli}@cbsr.ia.ac.cn

Abstract. The latest multi-biometric grand challenge (MBGC 2008) sets up a new experiment in which near infrared (NIR) face videos containing partial faces are used as a probe set and the visual (VIS) images of full faces are used as the target set. This is challenging for two reasons: (1) it has to deal with partially occluded faces in the NIR videos, and (2) the matching is between heterogeneous NIR and VIS faces. Partial face matching is also a problem often confronted in many video based face biometric applications.

In this paper, we propose a novel approach for solving this challenging problem. For partial face matching, we propose a local patch based method to deal with partial face data. For heterogeneous face matching, we propose the philosophy of enhancing common features in heterogeneous images while reducing differences. This is realized by using edge-enhancing filters, which at the same time is also beneficial for partial face matching. The approach requires neither learning procedures nor training data. Experiments are performed using the MBGC portal challenge data, comparing with several known state-of-the-arts methods. Extensive results show that the proposed approach, without knowing statistical characteristics of the subjects or data, outperforms the methods of contrast significantly, with ten-fold higher verification rates at FAR of 0.1%.

Keywords: Multiple biometric grand challenge (MBGC), MBGC portal challenge, video based face recognition, near infrared (NIR), heterogeneous face biometrics.

1 Introduction

The latest multiple biometric grand challenge (MBGC 2008) [1] sets up a series of experiments to evaluate recent advances of biometric technologies. In the MBGC portal challenge, near infrared (NIR) face videos are used as one of the probe sets whereas the target set consists of full faces in visual (VIS) images, as illustrated in fig.1. The NIR videos contain partial faces due to out-of-view during video capturing. Such partial occlusions of faces are often confronted in video surveillance and other face based applications.

Many methods have been proposed to deal with the occlusion problem in face recognition, which mainly based on patch or local feature strategies. For example, in [2], local nonnegative matrix factorization is used to learn local feature subspaces for face

---

⋆ Corresponding author.

Fig. 1. Examples of full VIS face targets (upper) and partial NIR face probes (lower) from MBGC portal challenge

recognition with occlusions. In [3], a face image is divided into local parts, and each part is modeled by a Gaussian distribution. The probability of a given image is calculated with the sum of all Gaussian models. In [4], instead of dividing face image into patches, the authors create six subparts of images through masking some regions (eg. left eye, right eye, nose or mouth) and learn a PCA subspace for each subset. A given face image is projected to these six PCA subspaces separately, then the final score is fused by weighted sum rule.

As face recognition technology advances, active near infrared (NIR) imaging has been used to effectively solve the illumination problem. On the other hand, many applications require visual (VIS) image based enrollment. To take advantage of latest technology while meeting requirements from applications, heterogeneous face biometrics by matching face images from different sources has become a new direction.

Recent years have seen developments in non-visual image based face biometrics, including 3D [5], thermal infrared (TIR) [6], and NIR [7] based methods. In many face based biometric applications, such as access control and E-Passport, the enrollment of face templates is done using VIS face images. Such images are normally acquired in controlled environment where the lighting is approximately homogeneous and frontal.

However, face biometrics has to deal with changes in lighting conditions. In dark environment, the quality of VIS face images can be too poor to be used for face recognition. Recent NIR technology well overcomes problems arising from lighting changes and has been successfully applied to face recognition in normal or dark environments [7]. So many face recognition systems mount NIR imaging device to adapt dark environment. While the NIR based method requires that enrollment face images be acquired using the NIR as well, how to take advantage of the new technology of NIR face to match against visual (VIS) face as required in many applications becomes an interesting problem.

Heterogeneous face biometrics by matching face images from different image sources has become a new direction for face recognition research. Several methods have

been proposed to solve this problem. In [8], Tang and his colleagues developed a PCA based method for face sketch recognition against face photos. In [9], the authors propose a unified framework for inter-modality face matching called common discriminant feature extraction (CDFE). CDFE can be considered as a modified linear discriminant analysis (LDA) method for heterogeneous face matching, in which scatter matrices are defined on two different types of images, and local consistency of manifold is imposed to regularize the dimension reduction. CDFE is applied to sketch-photo image matching and NIR-VIS face image matching, and obtains good results. In [10], a method is proposed for NIR-VIS face image matching problem based on canonical correlation analysis (CCA) [11]. Given that NIR-VIS image pairs of subjects are available for training, a CCA based correlation mechanism can be learned from corresponding NIR-VIS face images. The CCA learning is performed between features in PCA or LDA subspaces, rather than between images, by which better classification performance is gained.

In this paper, we propose an approach for solving the new challenging problem of matching between partial NIR faces and VIS faces. This is challenging for two reasons: (1) it has to deal with partially occluded faces in the NIR videos, and (2) the matching is between heterogeneous NIR and VIS faces. We adopt the philosophy of enhancing common features while reducing differences between heterogeneous image sources. A Laplacian of Gaussian (LoG) filter is applied to NIR and VIS face images. Edge points or low level feature points are extracted and converted into binary images. This significantly reduces heterogeneities between the NIR and VIS face images. The filtered images are segmented into small patches, and then template matching is done between these patches. Because the proposed method is local patch and edge-enhancing feature based, it is able to solve the problem of partial matching in heterogeneous images. Another advantage of the method is that it is non-learning based and basically requires no training data in the design of the classifiers. Hence, it largely overcomes the over-fitting problem in learning based methods.

Experiments are performed using the MBGC data sets. The proposed method is compared with several state-of-the-art methods, such as CDFE [9], PCA-CCA and LDA-CCA [10], on the same basis of patch segmentation. Extensive results show that the proposed approach, without knowing statistical characteristics of the subjects or data, outperforms the methods of contrast significantly, with ten-fold higher verification rates at FAR of 0.1%.

The rest of the paper is organized as follows. Partial face processing and analysis are discussed in Section 2. Local, edge-enhancing features are described in Section 3 for partial and heterogeneous face matching. The NIR-VIS face matching algorithm is described in Section 4. Section 5 describe experimental set up and extensive results.

2   Partial Face Processing

MBGC portal challenge includes two kinds of face data: the full VIS face images as the target set, and the NIR videos containing full or partial faces as the probe set. Some partial NIR faces and full VIS faces are shown in Fig. 1. We treat the problem as still face matching since temporal information is difficult to explore with such data sets, that is, departing the NIR videos into frames and then converting the problem into

matching between sets of still frames. Before matching, all face images need to be properly aligned.

In conventional methods with full faces, face images are usually aligned by four steps: (1) face Detection; (2) face landmark point localization, usually of two eyes, and sometime with nose and mouth; (3) face alignment according to the eye centers and sometime with references to nose and mouth; (4) cropping of aligned face.

However, for partial face images, as in MBGC NIR portal videos, not all landmarks are present. Face detection becomes difficult and landmark localization become inaccurate. To solve this problem, we developed an automated algorithm for landmark localization and thereby face alignment with full faces. According to located landmark points, a full or partial NIR face is cropped; this also gives the corresponding mask of the valid face area. An example is shown in Fig.2.
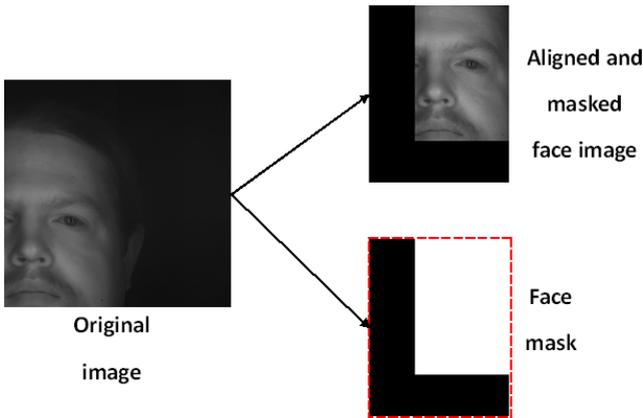

Fig. 2. A partial NIR face image, its corresponding aligned and masked version, and the face mask

While this paper is focused on the the problem of partial face matching in heterogeneous images, we leave the presentation of automated localization of landmarks in partial faces in a separate paper. However, we will evaluate partial face recognition performance with manual versus automated alignment.

3 Face Descriptors for NIR-VIS Matching

For heterogeneous face matching, we adopt the philosophy of enhancing common features while reducing differences between heterogeneous image sources. By visual inspection (Fig. 3), we can see although all pixel values of NIR and VIS face images are different, the structure of points, edges or contours are very similar. These low level features (such as nose location, eye shape, mouth contour, etc.) could supply common and essential information for NIR-VIS matching. By experiments, we found that features common to NIR and VIS images are mostly contained near edges. Therefore, we
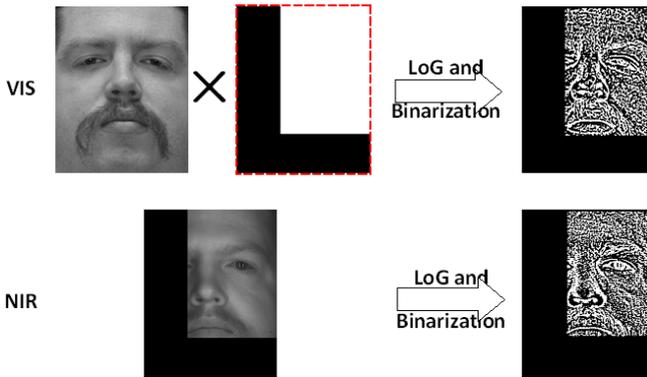
Fig. 3. NIR and VIS face images and their corresponding LoG filtered binary images

investigated into various edge enhancing features and found the Laplacian of Gaussian (LoG) is a good choice.

Laplacian filter is usually used to highlight the regions of rapid intensity change and is therefore often used for edge detection. Because Laplacian filter is approximating a second derivative measurement on a image, it is very sensitive to noise. To suppress noise, the image is smoothed by a Gaussian kernel before applying the Laplacian filter. This is equivalent to applying a LoG filter, which reduces high frequency noise components prior to the edge detection. The 2-D LoG function centered on zero and with Gaussian standard deviation σ has the following form:

$$LoG(x,y) = -\frac{1}{\pi\sigma^4}\left[1 - \frac{x^2+y^2}{2\sigma^2}\right]e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{1}$$

Assuming we use square 2D-LoG kernel to filter face images, there will be only two parameters of a 2D-LoG kernel: kernel size and Gaussian standard deviation. In practice, these two parameters are selected according the size of source image.

In this paper, LoG filter is used to detect the edge information of NIR and VIS face images. Firstly, using mask image to remove the redundant pixels of VIS image comparing to NIR image. Then NIR and VIS face images are filtered by a LoG filter and thresholded into binary images. Thresholded binary images are more robust to noise than gray images.

Fig. 3 illustrates a NIR and VIS face image and their corresponding LoG filtered binary images. We can see the difference between NIR-VIS image pair is remarkably reduced by binarized LoG filtering. This processing step provides a solid basis for the subsequent face matching. Surely, other filters also can be used to replace LoG filter, such as Canny operator, Sobel operator, and etc. Some kinds of local feature invariant to NIR-VIS spectrum are good choices too. Although this paper does not provide a thorough evaluation of all these filters (Gabor filters is proved not good), it directs to the choice of edge enhancing filters.

4   NIR-VIS Face Matching

The NIR-VIS face matching will be done based on the binary face features and the mask of the face area. Given a binary NIR-VIS image pair and mask, we need to compute a similarity score. Considering the structure of human face and possible occlusions, we divide the whole face into local patches of 16×16 (Fig. 4) and perform patch based matching. The following describes patch based matching algorithm for the face matching problem.
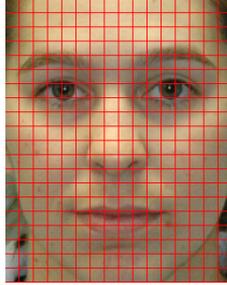

Fig. 4. The 256×320 pixel area of a face is divided into 16×16 local patches

First, segment the face from the background and obtain the mask of valid area of the face. Then compute the descriptor for each of the patches in the masked area. The binary features of corresponding patches are compared and distance computed using the Hamming distance. While the Hamming distance is suited for comparing two binary images, it's very sensitive to occlusion, expression and mis-alignment. Therefore, a local search is performed and the minimum Hamming distance is finally found to measure the difference between two patches. These are illustrated in Fig. 5. All the minimum distances are fused by the sum rule to obtain the overall distance.
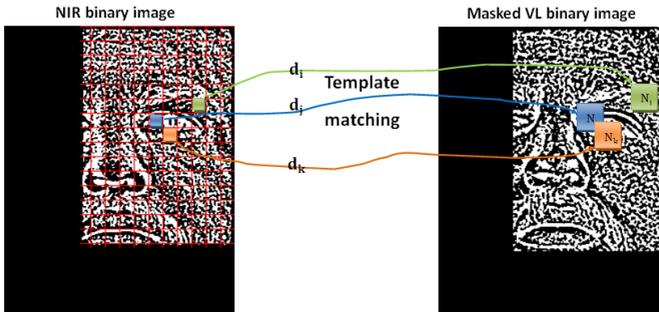

Fig. 5. The NIR-VIS face image matching strategy. The 256×320 pixel area of a face is divided into 16×16 local patches.

5   Experiments

The experiments are aimed to evaluate the performance of the proposed method in MBGC portal challenge, as compared to existing methods,including PCA-CCA, LDA-CCA, CDFE. Because PCA-CCA, LDA-CCA, CDFE cannot process partial data directly, for fair comparison, they are all implemented in the patch based way. That means these methods are applied to image patches and then patch matching scores are fused into a final score. Besides, in order to evaluate the influence of face alignment, all the experiments are conducted based on results of manual and automated face alignment. Therefore, the following 8 schemes are evaluated:

1. PCA-CCA-M: Patch based PCA-CCA on manually aligned face data;
2. PCA-CCA-A: Patch based PCA-CCA on automatically aligned face data;
3. LDA-CCA-M: Patch based LDA-CCA on manually aligned face data;
4. LDA-CCA-A: Patch based LDA-CCA on automatically aligned face data;
5. CDFE-M: Patch based CDFE on manually aligned face data;
6. CDFE-A: Patch based CDFE on automatically aligned face data.
7. Proposed-M: Proposed method on manually aligned face data;
8. Proposed-A: Proposed method on automatically aligned face data;

5.1   Database Description

For the portal challenge, MBGC only supply a database for testing. It contains 148 NIR videos and 148 VIS images of 114 persons. The still VIS face images are aligned and normalized into the size of 256×320 pixels. Frames are extracted from the NIR videos and masked and aligned in manual and automated way separately. Some aligned face images in MBGC database are shown in Fig.6. All the compared methods have the same basis of alignment.

Fig. 6. Examples of aligned VIS face images (upper) and NIR face images (lower) from the MBGC testing set

Because the CDFE and CCA based methods, unlike the proposed method, need a training set, a database of NIR and VIS face images of 200 persons was collected by ourselves. All the faces and eyes were detected in the images, aligned with eye coordinates and cropped into 256×320 pixels too, using an automated procedure, with results of acceptable quality. Some example training face images were shown in Fig.7.

Fig. 7. Examples of aligned VIS face images (upper) and NIR face images (lower) from the self-prepared training set

From these, we have the training set and testing set as follows:

1. NIR training set: Collected by ourselves, consisting of 200 persons, 6 NIR images per person;
2. VIS training set: Collected by ourselves, consisting of the same 200 persons with NIR training set, 6 VIS images per person;
3. NIR testing set: Coming from MBGC database, consisting of 114 persons not overlapping with training set, 1-2 NIR vidoes per person;
4. VIS testing set: Coming from MBGC database, consisting of the same 114 persons with NIR testing set, 1-2 VIS images per person.

Data sets (1) and (2) are used for training PCA-CCA, LDA-CCA and CDFE, whereas (3) and (4) are for testing. Apparently, the testing sets did not include any of persons in the training sets.

5.2   Performance Evaluation

The training of PCA-CCA, LDA-CCA and CDFE is performed as follows: PCA and LDA projection matrices are computed from the NIR and VIS patches of the training set, respectively. The projection reduces the dimensionality from 16×16 to 199. The CCA and CDFE projections further reduce the 199 to 100 dimensions to reduce overfitting.

The ROC curves of the compared methods are shown in Fig. 8. At FAR=0.001, the proposed Proposed-M and The Proposed-M achieves impressive 77% verification rate at FAR=0.1%, more than ten-fold higher than the compared PCA-CCA, LDA-CCA and CDFE methods of PCA-CCA-M, PCA-CCA-A, LDA-CCA-M, LDA-CCA-A, CDFE-M, and CFFE-A. Understandably, the manual alignment leads better performance than the automated alignment, with about 5% higher verification rate at FAR=0.001 for the proposed method. Evidently, we can see CCA based methods and CDFE performed poorly in portal challenge, which may be due to their poor generalization ability while they did well on the training data. The proposed method is very stable even without training.
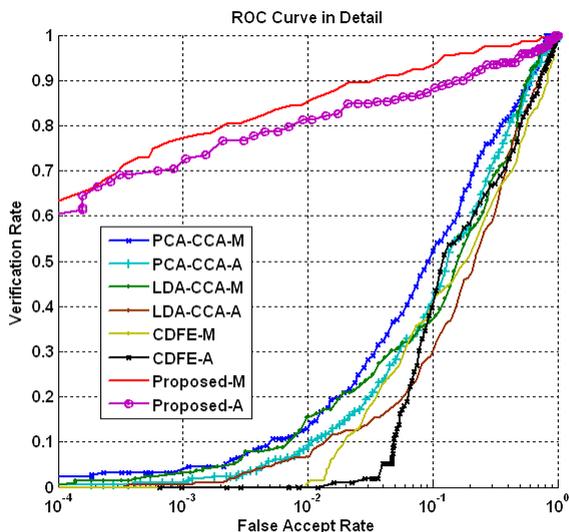

Fig. 8. ROC curves for 8 schemes in the proposed method

6   Conclusions

In this paper, we proposed a novel approach for partial face matching in NIR and VIS images for MBGC portal challenge. To deal with partial occlusion, local patch based matching is devised. To deal with NIR-VIS face matching, effective features are extracted using binarized LoG filters, which also help matching of partial faces. The experimental results on the MBGC portal challenge data illustrate the proposed method obtains significantly higher verification rate than the compared methods. The Proposed-M achieves impressive 77% verification rate at FAR=0.1%, more than ten-fold higher than the compared PCA-CCA, LDA-CCA and CDFE methods. These results support that the proposed approach is an effective solution for the new challenging problem.

References

1. NIST: Multiple Biometric Grand Challenge (MBGC) (2008),
   `http://face.nist.gov/mbgc`
2. Li, S.Z., Hou, X.W., Zhang, H.J.: Learning spatially localized, parts-based representation. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Hawaii, December 11-13, vol. 1, pp. 207–212 (2001)
3. Martinez, A.: Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(6), 748–763 (2002)
4. Tarres, F., Rama, A.: A novel method for face recognition under partial occlusion or facial expression variations. In: 47th International Symposium on ELMAR, June 2005, pp. 163–166 (2005)
5. Bowyer, K.W., Chang, Flynn, P.J.: A survey of 3D and multi-modal 3D+2D face recognition. In: Proceedings of International Conference on Pattern Recognition, August 2004, pp. 358–361 (2004)
6. Kong, S.G., Heo, J., Abidi, B., Paik, J., Abidi, M.: Recent advances in visual and infrared face recognition - A review. Computer Vision and Image Understanding 97(1), 103–135 (2005)
7. Li, S.Z., Chu, R., Liao, S., Zhang, L.: Illumination invariant face recognition using near-infrared images. IEEE Transactions on Pattern Analysis and Machine Intelligence 26 (April 2007) (Special issue on Biometrics: Progress and Directions)
8. Tang, X., Wang, X.: Face sketch recognition. IEEE Transactions on Circuits and Systems for Video Technology 14(1), 50–57 (2004)
9. Lin, D., Tang, X.: Inter-modality face recognition. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3954, pp. 13–26. Springer, Heidelberg (2006)
10. Yi, D., Liu, R., Chu, R., Lei, Z., Li, S.Z.: Face matching between near infrared and visible light images. In: Lee, S.-W., Li, S.Z. (eds.) ICB 2007. LNCS, vol. 4642, pp. 523–530. Springer, Heidelberg (2007)
11. Hotelling, H.: Relations between two sets of variates. Biometrika 28, 321–377 (1936)