Face Matching between Near Infrared and Visible Light Images

Dong Yi, Rong Liu, RuFeng Chu, Zhen Lei, and Stan Z. Li

Center for Biometrics Security Research & National Laboratory of Pattern Recognition
Institute of Automation, Chinese Academy of Sciences, Beijing 100080, China
http://www.cbsr.ia.ac.cn

Abstract. In many applications, such as E-Passport and driver's license, the *enrollment* of face templates is done using visible light (VIS) face images. Such images are normally acquired in controlled environment where the lighting is approximately frontal. However, *Authentication* is done in variable lighting conditions. Matching of faces in VIS images taken in different lighting conditions is still a big challenge. A recent development in near infrared (NIR) image based face recognition [1] has well overcome the difficulty arising from lighting changes. However, it requires that enrollment face images be acquired using NIR as well. In this paper, we present a new problem, that of matching a face in an NIR image against one in a VIS images, and propose a solution to it. The work is aimed to develop a new solution for meeting the accuracy requirement of face-based biometric recognition, by taking advantages of the recent NIR face technology while allowing the use of existing VIS face photos as gallery templates. Face recognition is done by matching an NIR probe face against a VIS gallery face. Based on an analysis of properties of NIR and VIS face images, we propose a learning-based approach for the different modality matching. A mechanism of correlation between NIR and VIS faces is learned from NIR→VIS face pairs, and the learned correlation is used to evaluate similarity between an NIR face and a VIS face. We provide preliminary results of NIR→VIS face matching for recognition under different illumination conditions. The results demonstrate advantages of NIR→VIS matching over VIS→VIS matching.

Key words: Face Recognition, Near Infrared (NIR), Visible Light (VIS) Images, Dimension Reduction, Canonical Correlation Analysis (CCA).

1 Introduction

This work is developed for *cooperative user applications*, such as E-passport, machine readable traveling document (MRTD), ATM ,and driver's license. In such applications, a user is required to cooperate with the camera to have his/her face image captured properly, in order to be permitted for some access. However, even in such applications, face recognition is compromised by environmental illumination changes. Much effort has been made to model and correct illumination changes on faces in visible light (VIS) images, eg [2–4]; however, this challenge remains unsolved problems due to formidable difficulties, especially uncontrolled illumination directions [5].

As a recent advance, the use of active near infrared (NIR) imaging provides an effective approach to solving the illumination problem. Face images are acquired with frontal lighting using active NIR lights mounted coaxal to the camera. This active NIR imaging method provides a good basis for face recognition regardless of environmental lighting. When incorporated with effective algorithms and system work, it leads to truly illumination invariant face recognition and excellent results for indoor, cooperative user applications [6, 1].

However, the exiting NIR based face recognition method [1] requires that the face templates of *enrollment* be also created from NIR face images, in addition to the use of NIR images for *authentication*. However, in many applications, templates of users were produced from VIS images taken in the visible light spectrum, such as passport and driver license photos. The ICAO (International Civil Aviation Organization) and ISO/IEC standards [7] have made recommendations for taking VIS photos.

In this work, we propose a new problem for face biometric, that of matching a face in an NIR image taken on authentication against a face in a VIS images taken on enrollment. This is for those applications where the use of VIS face images on enrollment is administratively required, such as E-passport and driver's license; the use of NIR face images on authentication is to take advantages of NIR face imaging in the latest NIR face recognition technology to combat illumination variations.

A solution to this new problem is to extract facial feature templates in the NIR and VIS face images, respectively, and then match them against each other. However, a straightforward matching between the two types of templates is not effective mainly because of different spectral properties in the two types of images, in addition to differences in the lighting conditions.

Therefore, we propose a more effective method for solving this new problem. We adopt a learning approach: Given that NIR→VIS image pairs of subjects are available for training, a canonical correlation analysis (CCA) [8] based correlation mechanism is learned from corresponding NIR→VIS face images for the NIR→VIS face matching. Here, the CCA learning is performed between features in linear discriminant analysis (LDA) subspaces, rather than between images. This leads to better classification performance. We provide preliminary results for NIR→VIS matching based face recognition. The results confirms advantages of NIR→VIS matching over VIS→VIS matching under varying illumination conditions.

2   NIR vs. VIS Images

We assume that enrollment is done using VIS images in controlled environment where the lighting is nearly frontal, following the Standard [7], whereas the authentication is done using NIR images in environment of arbitrary lighting but with active frontal NIR illumination [1]. Fig.1 shows example VIS and NIR face images taken in the two conditions. In the following, we give an analysis of these images.

According to the Lambertian reflectance model, an image $I(x,y)$ under a point light source is formed according to the following

$$I(x,y) = \rho(x,y)\mathbf{n}(x,y)\mathbf{s} \tag{1}$$

Fig. 1. Face image examples. Row 1: VIS face images (two on the left) taken on enrollment with frontal lighting and pose, and two NIR face images (two on the right) taken at the same time. Row 2: Two VIS image (left) and two NIR face images taken on authentication with lighting and pose variations.

where $\rho(x,y)$ is the albedo at point (x,y), $\mathbf{n} = (n_x, n_y, n_z)$ is the surface normal (a unit row vector) in the 3D space, and $\mathbf{s} = (s_x, s_y, s_z)^T$ is the lighting direction (a column vector, with magnitude). The albedo $\rho(x,y)$ function reflects the photometric properties, such as facial skin and hairs, and $\mathbf{n}(x,y)$ is the geometric shape (2.5 map) of the surface.

The product of $\rho(x,y)\mathbf{n}(x,y)$ is the intrinsic property of the face at a fixed imaging viewpoint, and is the only thing needed for face detection and recognition. Assume $\mathbf{s} = \kappa\mathbf{s}^0$, where κ is a multiplying constant due to the strength of the lighting, and $\mathbf{s}^0 = (s_x^0, s_y^0, s_z^0)^T$ is a unit column vector of the lighting direction. Equ.1 can be expressed as

$$I(x,y) = \kappa\rho(x,y)\mathbf{n}(x,y)\mathbf{s}^0 \qquad (2)$$

We see that the face image changes as the lighting direction changes, given albedo $\rho(x,y)$ and 3D shape $\mathbf{n}(x,y)$ fixed.

The photo taking Standards [7] suggest to use controlled lighting for face enrollment. The suggested lighting consists of mainly a frontal light, a little above the eyes (biased toward y direction), and some other auxiliary lights symmetric left-right. For cooperative users standing in front of camera, the overall lighting direction may be approximated by $\mathbf{s}^0 = (0, S_y^0, S_z^0)^T$ where $S_y^0 \ll S_z^0$ and S_y^0 is nonzero due to slight lift of the lighting above the eyes. Given these, the Lambertian equation can be written as

$$I(x,y) = \kappa\rho(x,y)\left[S_y^0 n_y(x,y) + S_z^0 n_z(x,y)\right] \qquad (3)$$

where $n_y(x,y)$ and $n_z(x,y)$ are the y and z components of the surface normal.

On authentication, the NIR images are taken by a camera with active NIR lighting. The active NIR lights, mounted on the camera panel, is from the frontal direction, *i.e.* $\mathbf{s}^0 = (0,0,1)^T$. An NIR image is depicted by

$$I(x,y) = \kappa\rho(x,y)n_z(x,y) \qquad (4)$$

Assuming VIS and NIR images are both taken in (nearly) frontal lighting, then both may be modeled by Equ.(4). However, the albedo $\rho(x,y)$ is different for the two

imaging processes due to different spectral properties of the facial surface in VIS and NIR spectra. Moreover, the assumption that the lighting is a point source and is from the frontal direction is an approximation. These cause matching from NIR to VIS a more challenging problem than the pure NIR to NIR matching [1].

In comparison, in the traditional VIS to VIS face matching, the images on enrollment and authentication are of homogeneous type. The main difficulties there are due to significant variation in lighting direction, *i.e.* $\mathbf{s}$ in Equ.(1). Such variation has been causing formidable difficulties for VIS→VIS face recognition.

3   Matching by Multivariate Correlation

Critical to the success of matching between NIR face and VIS face is the formulation of a correlation score for the evaluation of matches. The present method learns such a score function from a set of training examples. Let $\{(\mathbf{X}_1, \mathbf{Y}_1), \cdots, (\mathbf{X}_K, \mathbf{Y}_K)\}$ be a training set of K examples where $\mathbf{X}_k$ (NIR) and $\mathbf{Y}_k$ (VIS) are cropped NIR→VIS face pairs $W \times H$ pixels. These face images are aligned by the same fixed eye coordinates, and properly preprocessed to be zero mean and of unit length. Assume that these faces come from L subjects, denoted $\{C_1, C_2, \cdots, C_L\}$ where C_ℓ consists of subscripts $k \in \{1, \cdots, K\}$ of the training examples $(\mathbf{X}_k, \mathbf{Y}_k)$ belonging to subject ℓ.

The most naive method for NIR→VIS matching would be to compare the images $\mathbf{X}_k$ and $\mathbf{Y}_k$ directly. A better way could be to formulate the problem as correlational regression between $\mathbf{X}$ to $\mathbf{Y}$. However, it is difficult to learn a regression between image pairs directly to achieve a reasonable performance, because of different properties of $\mathbf{X}$ and $\mathbf{Y}$, the high dimensionality $W \times H$, and the insufficient sample size K.

Therefore, we propose the following three steps for the NIR→VIS face matching :

1. Transform $\mathbf{X}$ and $\mathbf{Y}$, respectively, into features spaces of lower dimensionality, as $\mathbf{X} \to \mathbf{X}' = \mathbf{P}_X(\mathbf{X})$ and $\mathbf{Y} \to \mathbf{Y}' = \mathbf{P}_Y(\mathbf{Y})$, where $\mathbf{P}_X$ and $\mathbf{P}_Y$ are dimension reductions for images $\mathbf{X}$ and $\mathbf{Y}$, respectively.
2. Perform multivariate regression between features in the reduced feature spaces, so that the corresponding feature vectors $(\mathbf{X}'_k, \mathbf{Y}'_k)$ are best correlated.
3. Evaluate the similarity between two types of features based on the feature regression using some correlation score.

In this paper, we will use baseline feature extraction and dimension reduction methods of PCA and LDA [9], and baseline correlational regression methods of CCA [8]. The process is further described below.

In the first step, two projection matrices $\mathbf{P}_X$ and $\mathbf{P}_Y$ are learned from the NIR images $\{\mathbf{X}_k\}$ and VIS images $\{\mathbf{Y}_k\}$, respectively, to reduce the dimensionality from $W \times H$ to two much smaller numbers. For PCA, the projections are learned based on total scatter matrices of $\{\mathbf{X}_k\}$ and $\{\mathbf{Y}_k\}$. For LDA, they are based on the intra-class and extra-class scatter matrices of the two image sets, where class labels of C_ℓ are used in the calculation.

In the second step, CCA is used to compute best correlational regression between the two sets of corresponding feature vectors $(\mathbf{X}', \mathbf{Y}')$. CCA here finds two linear projection matrices $\mathbf{W}_X = [\mathbf{w}_X^1, \mathbf{w}_X^2, \cdots, \mathbf{w}_X^d]$ and $\mathbf{W}_Y = [\mathbf{w}_Y^1, \mathbf{w}_Y^2, \cdots, \mathbf{w}_Y^d]$ from $\{(\mathbf{X}'_k, \mathbf{Y}'_k)\}$

(d is the dimension of CCA subspaces), such that $\mathbf{x} = \mathbf{w}_X^{i\ T}\mathbf{X}'$ and $\mathbf{y} = \mathbf{w}_Y^{i\ T}\mathbf{Y}$ are most correlated. This is done by maximizing the following correlation

$$\rho(\mathbf{w}_X, \mathbf{w}_Y) = \frac{E[\mathbf{x}\mathbf{y}^T]}{\sqrt{E[\|\mathbf{x}\|^2]E[\|\mathbf{y}\|^2]}} = \frac{\mathbf{w}_X^T\mathbf{C}_{XY}\mathbf{w}_Y}{\sqrt{\mathbf{w}_X^T\mathbf{C}_{XX}\mathbf{w}_X\mathbf{w}_Y^T\mathbf{C}_{YY}\mathbf{w}_Y}} \tag{5}$$

where $\mathbf{C}_{XY}$, $\mathbf{C}_{XX}$ and $\mathbf{C}_{YY}$ are the correlation matrices computed from the training features $\{(\mathbf{X}_k', \mathbf{Y}_k')\}$. This problem can be converted to the following generalized eigen problem:

$$\mathbf{A}\mathbf{w} = \lambda\mathbf{B}\mathbf{w}, \tag{6}$$

where

$$\mathbf{A} = \begin{pmatrix} 0 & \mathbf{C}_{XY} \\ \mathbf{C}_{YX} & 0 \end{pmatrix}, \mathbf{B} = \begin{pmatrix} \mathbf{C}_{XX} & 0 \\ 0 & \mathbf{C}_{YY} \end{pmatrix}, \mathbf{w} = \begin{pmatrix} \mathbf{w}_X \\ \mathbf{w}_Y \end{pmatrix}.$$

The solution $\mathbf{w}_X$ and $\mathbf{w}_Y$ can be found using singular value decomposition [9].

The third step evaluates correlation between $\mathbf{X}$ and $\mathbf{Y}$ in the learned subspaces. Features $\mathbf{X}'$ and $\mathbf{Y}'$ are extracted from $\mathbf{X}$ and $\mathbf{Y}$ using PCA or LDA projection learned in step 1. They are projected into the CCA subspaces as $\mathbf{x}$ and $\mathbf{y}$ using $\mathbf{w}_X$ and $\mathbf{w}_Y$ found in step 2. The correlation between $\mathbf{x}$ and $\mathbf{y}$, namely $\mathbf{x}\cdot\mathbf{y}/(\|\mathbf{x}\|\|\mathbf{y}\|)$, is calculated as the matching score.

CCA has been used for face recognition [10, 11]. There, both probe and gallery face templates are extracted from VIS images. Two other interesting related papers are [12, 13]. There, CCA-based methods are proposed for estimating face depth maps from color face images [12] and for generating NIR face images from RGB images [13]. Those are good examples of CCA based face analysis and synthesis. This paper explores another type of application of CCA, that for matching different face modalities.

4 Experiments

The experiments were aimed to evaluate the proposed NIR→VIS face matching method, as opposed to the conventional method of VIS→VIS matching. The following 6 algorithms were evaluated:

1. NIR→VIS / LDA+CCA;
2. NIR→VIS / PCA+CCA;
3. VIS→VIS / LDA+CCA;
4. VIS→VIS / PCA+CCA;
5. VIS→VIS / LDA;
6. VIS→VIS / LDA;

Methods using CCA directly on images without using LDA or PCA dimension reduction was not evaluated due to the memory limitation. We assumed that it was similar to the NIR→VIS / PCA+CCA and VIS→VIS / PCA+CCA algorithms. The VIS→VIS / PCA and VIS→VIS / LDA algorithms without CCA were included to provide baseline performances of the conventional algorithms. Note that pure PCA and LDA without CCA are inappropriate for NIR→VIS matching since NIR and VIS are images of different properties.

4.1   Data Description

A database of NIR and VIS face images of 200 persons was collected for the training and testing of the proposed method. The images were taken in 2 sessions, an enrollment session and an authentication session.

The enrollment session was with normal frontal lighting and frontal pose. For the enrollment session, 20 VIS face images were taken for each person, using a VIS camera, which was to simulate passport photos. We name this set as $\mathbf{VIS}_{frt}$, where the subscript $_{frt}$ stands for frontal lighting and pose. $\mathbf{VIS}_{frt}$ corresponds to Y in (X,Y) in the aforementioned formulations.

The authentication session was carried out with varying illuminations and slightly varying pose. 20 VIS face images were taken for each person using a VIS camera, and 20 NIR images were taken in the same condition. This was to simulate the situation of on-site person authentication. We name these sets as $\mathbf{VIS}_{var}$ and $\mathbf{NIR}_{var}$, where the subscript $_{var}$ stands for varying lighting and pose. Either $\mathbf{NIR}_{var}$ or $\mathbf{VIS}_{var}$ corresponds to $\{\mathbf{X}\}$.

The faces were detected in the images, and then aligned with eye positions and cropped into 120×142 pixels using automatic programs. The detection and alignment results were of acceptable quality. Some example face images were shown in Fig.1.

5 data sets were constructed from the database, each comprising all the 200 persons:

1. Training set for NIR→VIS matching – $\{\mathbf{X}\}$ consisting of 10 samples per person from $\mathbf{NIR}_{var}$, and $\{\mathbf{Y}\}$ consisting of 10 samples per person from $\mathbf{VIS}_{frt}$;
2. Training set for VIS→VIS matching – $\{\mathbf{X}\}$ consisting of 10 samples per person from $\mathbf{VIS}_{var}$, and $\{\mathbf{Y}\}$ consisting of 10 samples per person from $\mathbf{VIS}_{frt}$;
3. Gallery set – consisting of the other 10 samples per person from $\mathbf{VIS}_{frt}$;
4. Probe set for NIR→VIS matching – consisting of the other 10 samples from $\mathbf{NIR}_{var}$;
5. Probe set for VIS→VIS matching – consisting of the other 10 samples from $\mathbf{VIS}_{var}$.

Data sets (1), (3) and (4) were for evaluating NIR→VIS matching, whereas (2), (3) and (5) were for VIS→VIS matching. Note that the gallery and probe sets did not include any of the training data.

4.2   Results

The training of the model was performed as follows: For algorithms 1-4, PCA and LDA projection matrices are computed from the $\{\mathbf{X}\}$ and $\{\mathbf{Y}\}$ parts of the training set, respectively. For algorithms 5-6, PCA and LDA projection matrices are computed from the $\{\mathbf{Y}\}$ parts of the training set, respectively. The projection reduces the dimensionality from 120×142 to 199. CCA projections further reduce the 199 to 100 dimensions. Figure 2 shows the final ROC curves for the 6 algorithms. The ranking of the 6 algorithms in terms of verification rates at FAR=0.001 and 0.0001 as 1, 2, 3, 5, 4, and 6, Algorithm 1 giving the highest verification rate of all.

Table 1 provides verification rates at FAR=0.1% and 0.01%. Note that the verification rate of algorithm 1, the best of the NIR→VIS matching algorithms, is about 11% higher than algorithm 3, the best of the VIS→VIS algorithms. This was owing to the fact the NIR imaging is more stable than VIS imaging under various illuminations.

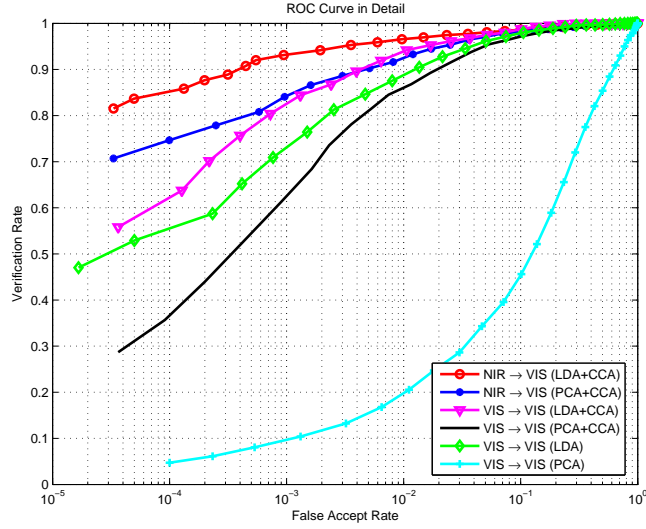**Fig. 2.** ROC curves for the 6 algorithms.

Table 1. Comparison of verification rates for the 6 algorithms

Algorithm	1	2	3	4	5	6
FAR=0.1%	93.1%	84.1%	82.1%	62.4%	73.3%	9.8%
FAR=0.01%	85.5%	75.5%	61.5%	37.2%	54.3%	4.9%

5 Summary and Conclusions

Face matching from NIR to VIS images is a new problem in biometrics. Taking advantages of the recently developed NIR-based face recognition technology, the problem is presented as a new approach towards accurate face recognition in practical applications. The preliminary results show that NIR→VIS matching is more advantageous than VIS→VIS matching for biometric authentication in environments of varying illumination. The success would have a positive impact on biometric applications such as E-Passport and drivers license.

Currently, we are using CCA for matching faces in different types of images. However, the CCA learning can be prone to over-fitting when the training data size is small, which has been reported elsewhere. The future work include developing better learning methods for object matching in different types of images (such as NIR vs. VIS), performing training and testing using larger data sets, and evaluating the performance under various conditions.

Acknowledgments

References

1. 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**(Special issu e on Biometrics: Progress and Directions) (2007)
2. Georghiades, A.S., Belhumeur, P.N., Kriegman, D.J.: "From few to many: Illumination cone models for face recognition under variable lighting and pose". IEEE Transactions on Pattern Analysis and Machine Intelligence **23**(6) (2001) 643–660
3. Gross, R., Brajovic, V.: "An image preprocessing algorithm for illumination invariant face recognition". In: Proc. 4th International Conference on Audio- and Video-Based Biometric Person Authentication, Guildford, UK (2003) 10–18
4. Nayar, S.K., Bolle, R.M.: "Reflectance based object recognition". International Journal of Computer Vision **17**(3) (1996) 219–240
5. Adini, Y., Moses, Y., Ullman, S.: "Face recognition: The problem of compensating for changes in illumination direction". IEEE Transactions on Pattern Analysis and Machine Intelligence **19**(7) (1997) 721–732
6. Li, S.Z., and His Face Team: "AuthenMetric F1: A Highly Accurate and Fast Face Recognition System". ICCV2005 - Demos (2005)
7. ISO/IEC JTC 1/SC 37: "Proposed Draft Amendment to ISO/IEC 19794-5 Face Image Data on Conditions for Taking Pictures". ISO/IEC 19794-5:2005/PDAM 1 (2006)
8. Hotelling, H.: "Relations between two sets of variates". Biometrika **28** (1936) 321–377
9. Fukunaga, K.: Introduction to statistical pattern recognition. 2 edn. Academic Press, Boston (1990)
10. Sun, Q., Heng, P., Zhong, J., Xia, D., Huang, D., Zhang, X., Huang, G.: "Face recognition based on generalized canonical correlation analysis". In: Proceedings of International Conference on Intelligent Computing, Hefei, China, Springer (2005)
11. He, Y., Zhao, L., Zou, C., Lipo, W., Ke, C., Soon, O.Y.: "Face recognition based on pca/kpca plus cca". In: Proceedings of International Conference on Advances in Natural Computation, Changsha, China, Springer (2005)
12. Reiter, M., Donner, R., Georg, L., Horst, B.: "3D and infrared face reconstruction from RGB data using canonical correlation analysiss". In: Proceedings of International Conference on Pattern Recognition. (2006)
13. Reiter, M., Donner, R., Georg, L., Horst, B.: "Predicting near infrared face texture from color face images using canonical correlation analysis". In: Proceedings of the Workshop of the Austrian Association for Pattern Recognition. (2006)