DESIGNING FACE VERIFICATION SYSTEMS BASED SYNTHESISED NEAR INFRARED INFORMATION

***Mohammad SadeghTaghizadeh[1], Seyed Mohammad SadeghTabatabaeifar[2],HosseinDehghan Tezerjan[3]**
[1,2]Department of Computer Engineering, Mehriz branch, Islamic Azad University, Mehriz, Iran
[3]Department of Electrical Engineering, Yazd branch, Islamic Azad University, yazd, Iran
*Author of Correspondence: mohammad.sadegh.t@gmail.com

ABSTRACT:*Changes in illumination conditions can cause drastic variations in face appearance and affect the performance of a face authentication system. Near infrared (NIR) face imaging systems have been proposed as a promising way towards illumination invariant face verification. In this paper, it is shown that when IR face images cannot be observed, learning the relationship between IR information and the corresponding visible images can provide useful source of complementary information about visible light image data. In particular, Canonical Correlation Analysis (CCA) is used to synthesise the NIR eigenfaces from their corresponding visible ones. The CCA-based synthesising algorithm is developed first. It is then shown that by fusing the visible and synthesised NIR information at the score level, the performance of the face verification system considerably improves.*

Keywords*: Face Verification, Fusion, Infrared Imaging, Canonical Correlation Analysis*

1. INTRODUCTION

With increased need for reliable authentication schemes, the use of automatic identity verification systems based on biometrics has become wide spread. Biometric products are currently used in e.g. airports, log-on devices for networked PCs, e-commerce, e-banking and health monitoring. Using this type of systems, individuals are recognized by behavioural or physiological traits such as finger prints, face, geometrical features of the hand, palm print, speech, iris and signature. Using these traits result in improved performance and efficiency of the security system, as compared with traditional methods. Face verification systems are widely used in recognition systems, mainly due to the passive nature, high acceptability, accessibility, relatively high precision and simplicity in obtaining the input data(image).Several challenges are encountered as face verification systems are used, e.g. illumination conditions[1], changes in facial expressions and the face angle. In general the image variations of different faces captured with the same conditions are smaller than that of the same face taken in a variety of environments. External factors such as pose and illumination can cause significant changes in the image plane. It has been shown that illumination causes larger variation in face images than pose [2]. Thus, for a reliable system the illumination issue has to be resolved.

In the past decades different design methods for designing reliable face verification systems with respect to environment illumination have been proposed. These methods are generally divided into active and passive methods [3]. For passive methods, imaging properties in the visible illumination spectrum(VIS) and are used in conjunction with different algorithms, to reduce the effects of problems involved with different environmental illuminations. Whereas for active methods the properties of different parts of the electromagnetic spectrum are considered, and by using imaging equipment appropriate for this non-visible spectrum an image is obtained. This image is independent of differences in environmental illumination. High energy electromagnetic waves with wavelengths less than that of visible light is harmful for humans. Thus most face verification research and systems are based on the VIS and infrared (IR)spectrum. The VIS extends from 0.4 to 0.7 µm and the IR spectrum extends from 0.7 µm to 10 mm. For face verification systems near infrared (NIR) (0.7 to 0.9µm) spectrum is used. The main reasons for adopting IR imaging in face verification systems are that the face is more easily detected, the differences between different images of one individual are less, IR imaging is almost not affected by environmental illumination and they are less sensitive to disguises and imposters as compared to VIS imaging. Adopting face verification systems based on all NIR and VIS spectra is associated with high equipment cost and complexities in equipment setup, as compared to solely VIS imaging. Investigations have shown that the NIR spectrum, has advantages as compared with VIS.A part of NIR waves radiated at objects are reflected. Thus it is possible to use NIR illumination sources for imaging in this spectrum. As NIR illumination is invisible this spectrum is invariant to visible light. NIR waves can easily penetrate through glasses and are invariant to ambient temperature [4].In 2003, Pan et al. designed a face verification system based on NIR spectrum, is shown that the NIR spectrum for different individuals' skins are distinctly different [4].Zhao and Grigat have proposed a face verification system based on NIR images. They adopted discrete cosine transform (DCT) and support vector machine (SVM) in their analysis [5]. Stan et al. have designed a NIR face verification system using NIR LEDs as the illumination source [6]. They adopted local binary pattern to extract ambient illumination independent features. In this paper the synthesis of NIR based on VIS images is investigated. Canonical Correlation analysis (CCA) is adopted to determine the relationship between these spectra [7].With the proposed method IR image information can be obtained without the use of IR cameras, so that the complexities and costs associated with IR imaging can be overcome. As each of the electromagnetic spectra show unique properties, it is expected that the accuracy and efficiency of face verification systems should improve. In the next section, the face verification system explained in detail. In section 3,the CCA algorithm and linear regression based on CCA are described. Later, the different data fusion techniques are introduced. The experimental results are presented in section 5. Finally, conclusions is presented.

2. FACE VERIFICATION SYSTEM

Verification is the process of determining whether a claim to an identity is correct or false. In face verification, this process is carried out by comparing an image of the claimant, with a stored template representing the client to whom the claim is made. The face verification system parts are as below:

1. Image acquisition
2. Face detection/localisation
3. Image normalisation
4. Feature extraction
5. Decision making

The input image contains both the face of the claimant and a potentially cluttered background scene. In the detection stage, the face is localised within the image. Often the output of this stage is the location of the eye centres. Normalisation consists of two processes; geometric normalisation and photometric normalisation. Geometric normalisation is necessary because the size of the face within the input image can vary as a function of the distance between the camera and the claimant. The face must therefore be cropped from the image and geometrically transformed to a pre-determined fixed size. A common strategy is to set the location of the eye centres to a constant position within an image of fixed size. The photometric normalisation attempts to remove unwanted lighting effects from the input image [8]. In the present work histogram equalization is adopted to achieve this, as it has a reasonable performance, in terms of computer efficiency, as compared with other methods [8].After standardizing the face image, the features are extracted to obtain unique data features, which can be used to efficiently differentiate between different faces.

Principal component analysis is a popular method of reducing the dimensionality of a set of data [9].

As the features are extracted, decision can be made. In this step the input data is classifiedinto two groups, client and imposter. Many pattern recognition systems, including verification systems, do not have enough training data, thus classifications methods like Bayesian which needs lots of training data cannot be used. For solving this problem, similarity measurement criteria can be applied. Normalised correlation (NC) criterion can be adopted as a similarity measurement [8]. NC is an angle measurement criterion, in which the angle between the input sample and the claimed model is measured. The cosine of this angle is termed as score. Using this score, the decision on accepting or rejecting an individuals' claims made. The$\Lambda_I(x)$is considered as the similarity score between the input sample, *x*, and the identity, *I*. If this is more than or equal to the system level threshold, *T*, the claim is accepted, otherwise it is rejected. The decision threshold level is determined in the system evaluation level.

3. CANONICAL CORRELATION ANALYSIS

CCA is a powerful multivariate analysis tool with diverse applications [10].Considering two sets of variables, CCA is applied to construct a space in which the correlation between the sets of variables is maximised. In this investigation a CCA based linear regression is adopted to synthesise NIR spectrum from VIS data.

3.1. CCA algorithm

Given *N* pairs of centralised samples (x_i, y_i) of (*X*, *Y*), *i=1,2,…,N*, where $X \in R^m$ and $Y \in R^n$, the goal of CCA is to find a pair of directions w_x and w_y which maximise the correlation of the projected data, $w_x^T X$ and $w_y^T Y$, where T denotes the transpose operator. This can be achieved by maximising the correlation function in

$$\rho = \frac{w_x^T C_{xy} w_y}{\sqrt{w_x^T C_{xx} w_x w_y^T C_{yy} w_y}}$$

where C_{xx} and C_{yy} are the within-sets covariance matrices,and C_{xy} and C_{yx} are the between-sets covariance matrices.Let

$$A = \begin{pmatrix} 0 & C_{xy} \\ C_{yx} & 0 \end{pmatrix}, B = \begin{pmatrix} C_{xx} & 0 \\ 0 & C_{yy} \end{pmatrix}$$

At most *k* = min(*p*, *q*) factor pairs $\langle w_x^i, w_y^i \rangle, i = 1,2,\dots,k$can be obtained by solving

$$W^i = \left(w_x^{iT}, w_y^{iT}\right)^T = argmax\left(w_x^i, w_y^i\right)\{\rho\}$$

subjected to,

$$\rho\left(w_x^j, w_y^i\right) = \rho\left(w_x^i, w_y^j\right) = 0, j = 1, \dots, i-1$$

The factor pairs W^i can be obtained as solutions (i.e. eigenvectors) of a generalized eigenproblem (for details see e.g. [11]). The extreme values $\rho(W^i)$, which are referred to as *canonical correlations*, are obtained as corresponding eigenvalues. By employing CCA, regression is performed on only a small number of linear features (compared to the original dimensionality of the data), i.e. derived linear combinations of the original response variables. Thus, CCA can be used to compute the (reduced) rank-*n* regression parameter matrix, *R*, by using only *n* < *k* factor pairs. Thereby, in contrast to standard multivariate regression, CCA takes advantage of the correlations between the response variables to improve predictive accuracy [12].

3.2. CCA based multivariate linear regression

Multivariate linear regression (MLR) is used to model the predictive relationships of multiple related responses on a set of predictors. In multivariate linear regression, there are *n* observations on q responses, $y = (y_1, y_2, \dots, y_q)$, and *p* illustrative variables, $x = (x_1, x_2, \dots, x_p)$. The multivariate linear regression can be expressed as

$$Y = RX + E$$

where $Y = [y_1 y_2 \dots y_n]_{q\times n}$, $X = [x_1 x_2 \dots x_n]_{p\times n}$ and R is a $q\times p$ regression parameter matrix. $E = [e_1 e_2 \dots e_n]_{q\times n}$is the regression noise drawn independently from N(0,Σ) [13].By assuming $\tilde{x} = W_x^T x$ and that the variables y and $\tilde{x}$have a linear relationship as

$$y = R\tilde{x} + \varepsilon$$

where$\varepsilon$is the regression error. According to [10] R can be determined by

$$R = \left(\tilde{X}\right)^\tau Y^T = \tilde{X}^T\left(\tilde{X}\tilde{X}^T\right)^{-1} Y^T$$

where$\tilde{X}^\tau = \tilde{X}^T\left(\tilde{X}\tilde{X}^T\right)^{-1}$is the pseudo inverse of $\tilde{X}$.

Given a new input vector x_{new}, $\tilde{x}_{new}$ is computed by usingthe CCA transformation matrix,

$$\tilde{x}_{new} = W^T x_{new}$$

And *y* is predicted by:

$$y_{new} = R\tilde{x}_{new}$$

In the present work, *X* and *Y* are sets of visible and infrared images, respectively, which are used for training CCA and computing MLR parameter. y_{new} isthe synthesised NIR image obtained from the visible image,x_{new}.

4. Fusion of multispectral information

In multimodal or multi sensor biometric systems, the presented information can be integrated at four levels: 1) Signal level 2) Feature level 3) Score level and 4) Decision level. Investigations have shown that signal and score level fusion improves verification system efficiency noticeably [14]. In the present study, the effects of different score level fusion rules are investigated for three classifiers which are based on the visible and synthesised NIR image features. For this purpose, different untrained and trained fusion rules are examined.

4.1. Untrained score level fusion rules

In these methods, the final score, S, is achieved by integrating the classifiers scores, S_i. Some untrained fusion rules are presented in Table 1.

Table 1.Untrained rules for score level fusion.

max	min	average	Product
$S = \max\{S_i\}$	$S = \min\{S_i\}$	$S = \frac{1}{N}\sum_{i=1}^{N} S_i$	$S = \frac{1}{N}\prod_{i=1}^{N} S_i$

As previous investigations have shown that the average method is more efficient than the other methods presented in Table 1 [15], this is adopted in the analysis here.

4.2. Trained score level fusion rule

In this case, the fusion rule is learnt using a SVM. The aim of SVM is to find the condition in which the structural risk of the classifier is minimised. For score level fusion, a set of scores S = [S1, S2, …, SM] is used as the input feature vector for training the SVM. At the test level, scores are classified using the designed SVM [16].

5. EXPERIMENT

To determine the efficacy of the proposed system, two face datasets have been considered. The first dataset containing both visible and NIR images (VIS+NIR) is used for training the synthesis system. The performance of the proposed system is then evaluated by considering the third independent dataset which contains only visible images (XM2VTS). In this section, first the adopted datasets are introduced and then the experimental results are presented.

5.1. VIS+NIR database

This dataset contains pairs of visible and near infrared images of 40 subjects' faces. The face images were captured in two sessions, over a period of several weeks. In each session, 4 different illumination configurations were used (light sources directed individually from left, bottom, right and top) and for each illumination condition, 6 recordings were acquired. Therefore, there are 1920 visible and NIR face images. The original resolution of the visible and NIR images is 382×288 pixels [16]. Exemplary images from this database are shown in Figure1.

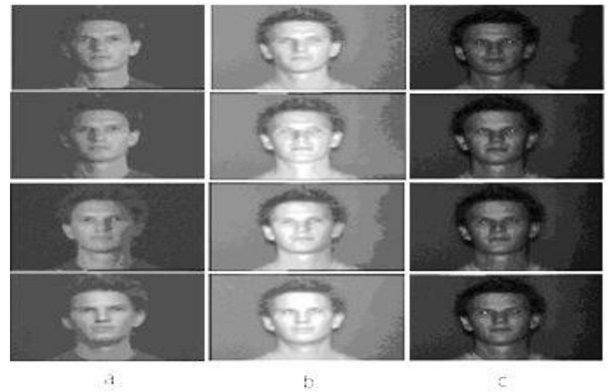

Figure1. Exemplary images of VIS+NIR face database, a.VIS images, b. VIS+NIR images, c. NIR images.

5.2. XM2VTS database

The XM2VTS database and its associated experimental protocols have been considered for evaluating the performance of the proposed face verification system. The XM2VTS database is a multi-modal database consisting of face images, video sequences and speech recordings taken of 295 subjects with one-month intervals. This database contains 4 sessions with two shots at each session. The size of the images are720x596. For the task of face verification, a standard protocol for performance assessment has been defined. There exist two configurations that differ by the selection of particular shots of people into the training, evaluation and test sets. These configurations are presented in Tables 2 and 3, respectively. Some samples taken from this database are shown in Figure2[17].

Table 2.Configuration1 of XM2VTS database.

session	Shot	200	25	70
1	1	Train	Imposter-Eval.	Imposter-Test
	2	Client-Eval.		
2	1	Train		
	2	Client-Eval.		
3	1	Train		
	2	Client-Eval.		
4	1	Client-Test		
	2			

Table3.Configuration 2 of XM2VTS database.

session	Shot	200	25	70
1	1	Train	Imposter-Eval.	Imposter-Test
	2			
2	1	Train		
	2			
3	1	Client-Eval.		
	2			
4	1	Client-Test		
	2			

Figure2.Sample gray level images from the XM2VTS database.

5.3 EXPERIMENTAL RESULTS First, all images are normalised based on the eye positions and the face region is extracted. This results in an image size reduction to 64x48 pixels. As the available image size is larger than the number of images, the covariance matrix becomes singular. By adopting the PCA algorithm to reduce the image dimensions this types of problems can be solved. In the present study, the principal component analysis is used to preserve 95% of data variations. The overall procedure involves two phases:

Phase1: Using VIS+NIR face database, the MLR parameters (w_x, w_y, R) are learnt. The factor pairs($\mathrm{w_x}$, $\mathrm{w_y}$) and R are calculated using (4) and (7), respectively. As the relationship between the VIS and NIR images is determined, eigenfaces corresponding to VIS images (69 eigenfaces) and NIR (222 eigenfaces) are extracted by adopting PCA. And the projected data to PCA space are used as input data to the CCA algorithm.

It is worth noting that, the transformation matrix obtained by adopting the CCA algorithm expresses the relationship between the projected visible and NIR images to PCA space(not the relationship between VIS and NIR images). As stated in section 3, determining the optimum size of factor pairs is essential. For this purpose, after applying the CCA algorithm and determining factor pairs, regression parameters are determined for the different factor pair sizes. Then, by applying the transformation matrices on the VIS image, the corresponding NIR data is synthesised. By comparing the synthesised NIR data with the real NIR data (in the related PCA space) the mean square error (MSE) for different factor pairs sizes are determined. It is noted that as the number of canonical factors increase, initially the error is drastically reduced, but then the error is almost constant. The point, after which the error is almost constant, is chosen as the optimum factor pairs' size.

Phase2: Using the images of XM2VTS Database, by considering the MLR parameters derived in phase 1,the features related to the NIR data is synthesised, and a face verification system based on fusion of the scores at the score level, is made.

Performance of the face verification system is determined by considering the half total error rate, $HTER = \frac{FAR+FRR}{2}$.FAR and FRR stand for the false acceptance rate and false rejection rate, respectively. FAR occurs when an imposter claims to be a client and the claim is accepted. FRR occurs when a client identity claim is rejected. At system evaluation stage the threshold level is determined. In this study the threshold is set by assuming equal error rate (EER), thus FAR and FRR have almost the same values. In Figure3 a receiver operating characteristic (ROC) curve sample showing the relationship between the two error types discussed as function of different threshold levels are shown. The point on this curve where FAR and FRR are equal is known as the EER. In the studies that follow, the threshold corresponding to EER is found on an evaluation set. This threshold is then applied to a test set to give FAR and FRR and hence the HTER.

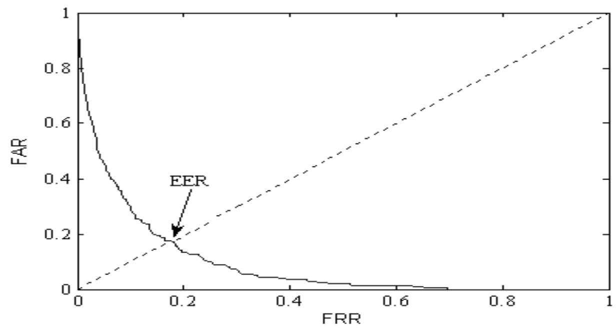

Figure3.The ROC curve and determining the threshold using EER.

In order to evaluate the performance of the proposed face verification system, the system based on visible images of the XM2VTS database for the first configuration (Config1) and second configuration (Config2) in PCA feature space at the first stage of the test are investigated. The performance of the system for synthesised NIR images is also determined. The FAR, FRR and HTER errors obtained in PCA space for this spectrum are presented in Tables 4 and 5.

Table4.Face verification system error(%) based on visible images from XM2VTS

	Visible					
	Evaluation			Test		
	FAR	FRR	HTER	FAR	FRR	HTER
Config1	4.66	6.83	5.74	5.53	4.75	5.14
Config2	4.46	3.75	4.10	5.45	4.50	4.97

Table5. Face verification system error(%) based on synthesised NIR images from XM2VTS

	Synthesised NIR					
	Evaluation			Test		
	FAR	FRR	HTER	FAR	FRR	HTER
Config1	8.50	8.33	8.41	9.48	7.00	8.24
Config2	5.67	5.75	5.71	5.89	8.00	6.94

It is seen that as contrary to what is expected, the performance of face verification systems based on synthesised NIR data does not give improved results as compared to visible image data.

In fact, at CCA learning level, registration of pair images are very important. The miss registration of VIS and corresponding NIR spectrum images has a great effect on NIR spectrum synthesising algorithm efficiency. Additionally, in this work it was seen that the covariance matrix is assumed to be static in PCA and CCA algorithms.

The synthesised NIR data is not accurate enough, as this is not entirely true. It will be shown later that, in fusion result of visible and synthesised NIR data, synthesised NIR data can be used as source of complementary information in a multispectral face verification system.

At the second stage of the test level, fusion of VIS and synthesised NIR is investigated by adopting average and SVM fusion rules. The results are presented in Table 6. It is worth noting that, as seen in Tables 4 and 5,HTERE with HTERT are corresponding for most cases. Thus only the HTERT are presented here.

Table 6.HTERT(%)for average and SVM fusion rules

	Average rule	SVM rule
	NIR+VIS	NIR+VIS
Config1	4.25	3.95
Config2	3.85	3.68

As previously seen in Table 4, HTERT(%) obtained from visible images for Config1 and Config2 are 5.14, 4.97, respectively. It is noted that, all state errors reduce as different spectra are fused by adopting average and SVM rules, as compared to face verification systems based on only visible images. Even though synthesised NIR images result contains more errors as compared with visible images, but fusion of these data leads to a face verification system with reduced error. It is seen that by adopting average and SVM fusion rules in Config1 and Config2, fusion of VIS and synthesised NIR gives the best system performance. Adopting SVM yields the least error. It is seen that the least fusion error(%)for Config1 is 3.95, which corresponds to a reduction of 1.19 units (almost 23%) as compared to the error(%) of visible images with 5.14. For config2 the least error(%) is 3.68, whereas the same for visible image error(%) is 4.97. The system error(%)for this configuration has reduced by 1.29units (almost 26%).This convey that fusion of synthesised NIR and visible data improves performance of the system significantly.

CONCLUSIONS

In this work NIR spectrum issynthesised from VIS data, for adoption in a faceverification system. The aim of this is to improve the face verification performance. The improvement in fusion of the VIS and synthesised NIR spectrum are investigated.CCA based Linear regression is adopted for synthesising NIR data from VIS images. It is expected that by fusing VIS with NIR face data, the system performance should improve. It is seen that fusing the VIS face data with synthesised NIR face data improves the proposed faceverification system performance by reducing the average error more than 23%.

REFERENCES

[1] Zhao, W.;Chellappa, R.; Rosenfeld, A.: Face recognition: A literature survey. ACM ComputingSurveys, Vol. 35, 399-458 (2003).

[2] Moses, Y.;Adani, Y.; Ullman, S.: Face Recognition: The problem of compensating for changes in illumination direction. in Proceedings of European Conference on Computer Vision, 286- 2996(1997).

[3] Zou, X.; Kittler, J.; Messer, K.: Illumination Invariant Face Recognition: A Survey. In Proceedings of IEEE Conference on Biometrics: Theory, Applications and Systems,1-8(2007).

[4] Pan, Z.; Healey, G.; Prasad, M.;Tromberg,B.:Face Recognition in Hyperspectral Images. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 25, No. 12, 1552–1560(2003).

[5] Zhao, S.;Grigat, R.: An Automatic Face Recognition System in the Near Infrared Spectrum. In Proc. of the Conference on Machine Learning and Data Mining in Pattern Recognition:437–444(2005).

[6] Stan, Z. Li.; Chu, R.; Liao, S.; Zhang, L.: Illumination Invariant Face Recognition using near-infrared images. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.29, No.4,627-639(2007).

[7] Hotelling, H.: Relations between two sets of variates. Biometrika, Vol.8,321-377(1936).

[8] Li, Y.: Linear Discriminated Analysis and its application to face Identification. PhD. Thesis, University of Surrey (2000).

[9] Jolliffe, T.: Principal Component Analysis.Springer(2002).

[10] Hardoon, D. R.;Szedmak, S. R.; Shawetaylor, J. R.: Canonical correlation analysis: An overview with application to learning methods. Neural Computation, Vol. 16, No. 12, 2639-2664(2004).

[11] Melzer, T.; Reiter, M.; Bischof,H.: Appearance models based on kernel canonical correlation analysis. Pattern Recognition, Vol. 36, No. 9,1961–1971(2003).

[12] Breiman, L.; Friedman, J.: Predicting multivariate responses in multiple linear regression. Journal of the Royal Statistical Society, 59(1):3–54(1997).

[13] Yuan, M.;Ekici, A.; Lu, S.;Monteiro, R.: Dimension reduction and coefficient estimation in multivariate linear regression. Journal of Royal Statistical Society, part 3, 329-346 (2007).

[14] Sadeghi, M.; Kittler, J.: A Comparative Study of Data Fusion Strategies in Face Verification. In the 12th European Signal Processing Conference, 45-56(2004).

[15] Sadeghi, M.;Samiei, M.; Kittler, J.: Selection and Fusion of Similarity Measure Based Classifiers Using Support Vector Machines. International conference on Structural and Syntactic Pattern Recognition,479-488(2008).

[16] Zou, X.: Illumination Invariant Face Recognition Based on Active Near-Infrared Differential Imaging. Ph.D. Thesis, University of Surrey(2007).

[17] Messer, K.;Matas, J.; Kittler, J.;Luettin, J.;Maitre, G.: XM2VTSDB: The extended M2VTS database. In Second International Conference on Audio and Video-based Biometric Person Authentication,72–77(1997)