2014 IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP)

PROBABILISTIC LINEAR DISCRIMINANT ANALYSIS FOR INTERMODALITY FACE RECOGNITION

M. K. Shaikh[1], *M. A. Tahir*[2] *and A. Bouridane*[1]

[1]Computer and Electronics Security Systems
Department of Computer Science and Digital Technologies, Northumbria University, UK
{muhammad.shaikh,ahmed.bouridane}@northumbria.ac.uk
[2] College of Information and Computer Sciences
Al Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh, KSA
{mtahir}@ccis.imamu.edu.sa

ABSTRACT

Intermodality face matching or Heterogeneous face recognition involves matching faces from different modalities such as infrared images, sketch images and low/high resolution visual images. This problem is further alleviated due to inherit problems in face recognition such as pose, expression, illumination, occlusion etc. Existing face recognition algorithms fail to address the existing feature gap exist between images of different modalities. To solve this problem, we propose a new method inspired from Probabilistic Linear Discriminant Analysis (PLDA). PLDA is a generative probabilistic method which models the face into signal and noise components. This method reports outstanding results when compared to other contemporary approaches. But PLDA is designed to apply the image data in only one modality. In this paper, its efficacy has been extended to more generic problem of handling faces captured in different modalities. Experiments conducted on HFB (VIS-NIR) , Biosecure (Low-High or Webcam-Digitalcam) face databases validate its robustness and superiority over other methods.

Index Terms— Probabilistic Linear Discriminant Analysis (PLDA) , subspace learning , Latent Identity variable (LIV)

1. INTRODUCTION

Face recognition is a difficult problem due to the intrinsic similarity of the classes, the wide range of perturbations and changes in imaging conditions. These include variations in illumination and facial expression, occlusion, and pose or view angle. These challenges are manifested in large variability in facial appearance of the same person. The problem is further aggravated by intermodality face matching involving matching faces from different modalities such as infra-red images, sketch images and low/high resolution visual images.

Recent trends have shown that the researchers in Biometric area are trying to tackle this problem by minimizing the feature gap of the same image captured using different modalities. There are three major and broad categories where researchers can handle this issue. Analysis by synthesis Methods i.e. face samples of one modality are first transformed to another modality so that the appearance difference is minimized. The representative work in this area include Eigen transform Method [1] ,Local linear preserving Method [2], MRF modeling [3]. Extraction of Consistent Features i.e. proper texture descriptors are designed to reduce the feature gap between modalities. Difference of Gaussian (DOG) filter [4] is used to reduce appearance difference and extract Multi-block LBP. Using HOG and LBP , applying sparse representation classifier [5], Sift and multiscale LBP [6] is also employed in this area. Subspace learning Methods are developed to find a common discriminant subspace to classify heterogeneous data. Some of the representative works in this area are Regularized Discriminative CSR [7] , CDFE [8] etc.

In this paper, we propose a new method inspired from Probabilistic Linear Discriminant Analysis (PLDA) for heterogeneous face recognition. PLDA is a generative probabilistic method which models the face into signal and noise components. It seeks to maximize the discrimination probabilistically by maximizing the inter-class variation and minimizing the intra-class variance. Further, it is a Bayesian generative approach, thus, brings quite favorable characteristics e.g. allowing careful modeling of noise, ignoring variables of least interest by marginalizing over them and providing a coherent way of comparing models using Bayesian model comparison. Following are the main contributions in this paper

- The proposed method provides a theoretical foundation for intermodality face matching using probabilistic linear discriminant analysis (PLDA). Due to its probabilistic nature, information from different modalities can easily be combined and priors can be applied over the possible matching. To the best of our knowledge, this is first study that aims to apply PLDA for intermodality face recognition

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

- The proposed system is evaluated on two challenging benchmarks of intermodality face matching: Biosecure (Low vs High) and HFB (VIS vs NIR). The proposed technique has produced better and comparative results when compared with the existing techniques.

The remainder of the paper is organized as follows. In Section 2 , we describe Probabilistic linear discriminant analysis , its training and recognition stages. Experiment set-up and results are presented in Section 3. Section 4 concludes this paper.

2. PROBABILISTIC LINEAR DISCRIMINANT ANALYSIS(PLDA)

The use of generative probabilistic approaches have been applied in quite wider domain of object recognition [9], image segmentation [10], object tracking [11] etc. The main theme of these approaches lies under the notion that observations are indirectly created from set of underlying variables with some noise associated with it.

PLDA [12] is a generative probabilistic method which models the face into signal component and noise component. The signal component represents the identity of an individual as hidden variable called as latent identity variable (LIV) while the noise component reflects any remaining variation of the face that is not attributed towards identity.

PLDA is very closely related to Linear discriminant Analysis [13] as it seeks to maximize the discriminability probabilistically by maximizing the inter-class variation and minimizing the intra-class variance. PLDA being a Bayesian generative approach brings quite favorable characteristics like posterior probabilities give more flexility in adjusting / deferring the final decision if uncertainty is quite big.

The other obvious advantage of PLDA as stated in [12] that a probabilistic solution means that we can easily combine information from different measurement modalities and apply priors over the possible matching. For this reason, we extend the utilization of PLDA in intermodality face matching problem.

2.1. Latent Identity Subspace (LIV)

PLDA assumes that there exits a multidimensional variable in a new subspace which represents the identity of an individual regardless of the modality. This variable is termed as latent identity variable (LIV) which resides in a subspace called latent identity space as opposed to observed space where the images are captured.

The key property of LIV is that if two LIVs take the same values then it corresponds to an identity of same individual and vice versa. PLDA never measures the LIVs directly but through observed images generated from latent variable with its associated noise. Figure 1 reflects latent identity approach

2.2. PLDA Model Description

PLDA model is of the form

$$x_{ij} = \mu + Fh_i + Gw_{ij} + \epsilon_{ij} \tag{1}$$

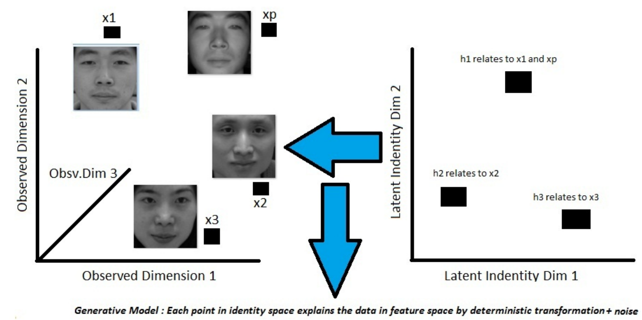

Fig. 1. Representation of Observed and Identity space showing each point in latent space is different individual while each position in observed space is reflecting different image.

It denotes the jth image of an ith individual by x_{ij}. The term μ represents the overall mean of the training dataset. F denotes the basis function for between individual variance with its associated LIV h_i (remain constant for every person) that corresponds to individual's position in the LIV subspace. G denotes the basis function within individual variance with its associated w_{ij} that corresponds to position in this subspace for jth image of ith individual. ϵ_{ij} is a residual noise term defined as Gaussian with diagonal covariance Σ.

The signal component in this model $\mu + Fh_i$ depends only on the identity of the person (only i) as there is no image dependence is present (no j) describing between-individual variance while noise component $Gw_{ij} + \epsilon_{ij}$ which depends on both i and j describing within-individual variance of the same images of an individual. Formally, the PLDA model can be described using conditional probabilities as

$$P_r(x_{ij}|h_i, w_{ij}, \theta) = g_x[\mu + Fh_i + Gw_{ij}, \Sigma] \tag{2}$$
$$P_r(h_i) = g_h[0, I] \tag{3}$$
$$P_r(w_{ij}) = g_w[0, I] \tag{4}$$

where $g_a[b, C]$ describes a gaussian in a with mean b and covariance C. Equations 3 and 4 define simple priors on h_i and w_{ij}.

Figure 2 shows the components of PLDA including signal and noise subspace components.

2.3. Learning PLDA parameters : Training Stage

In PLDA model, the only known parameters are the observed images while the rest $\theta = \mu, F, G, \Sigma$ are all unknown. If we know h_i and w_{ij} , then the learning parameters F and G will be quite easier. But, unfortunately, all the right hand side parameters of our PLDA model in Equation 1 are unknown.

Fortunately, for this chicken-egg problem, one can take advantage of Expectation and Maximization Algorithm [14] which iteratively maximizes the likelihood of parameters alternately in each iteration. The E step finds the unknown identity variables h_i and w_{ij} by calculating posterior probabilities over fixed parameter values. In M step, the algorithm maximizes the lower bound on the parameters $\theta = \mu, F, G, \Sigma$.

The first two moments of Expectation (E) steps and update rules for Maximization (M) step for this model are

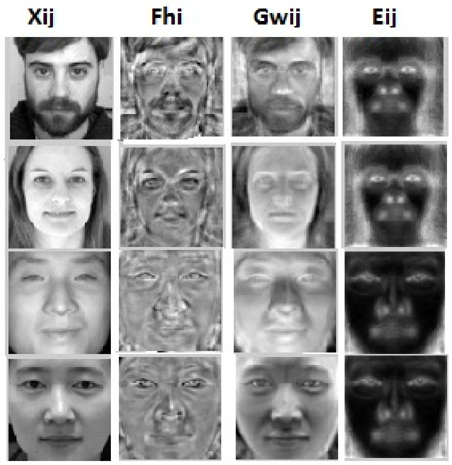

Fig. 2. Visualization of PLDA signal and noise components.

$$E[y_i] = \left(A^T\Sigma'^{-1}A + I\right)^{-1} A^T\Sigma'^{-1}(x_i - \mu') \quad (5)$$

$$E[y_i y_i^T] = \left(A^T\Sigma'^{-1}A^T + I\right)^{-1} + E[y_i]E[y_i]^T \quad (6)$$

$$\mu = 1/IJ \sum_{i,j} x_{ij} \quad (7)$$

$$x_{ij} = \mu + \begin{bmatrix} F & G \end{bmatrix} \begin{bmatrix} h_i \\ w_{ij} \end{bmatrix} + \epsilon_{ij} \quad (8)$$

$$x_{ij} = \mu + Bz_{ij} + \epsilon_{ij} \quad (9)$$

$$B = \Big(\sum_{i,j}(x_{ij-\mu}E[zi])^T\Big)\Big(\sum_{i,j}E[z_i z_i^T]\Big)^{-1} \quad (10)$$

$$\Sigma = 1/IJ \sum_{i,j} Diag\Big[(x_{ij}-\mu)(x_{ij}-\mu)^T - BE[z_i](x_{ij}-\mu)^T\Big] \quad (11)$$

where Diag represents only the diagonal elements of the matrix.

Figure 3 reflects the main idea of PLDA that images from different modalities of same subject share the same identity variable.

2.4. Recognition Stage

After learning model parameters, our next stage is to match two images sharing the same identity variable h. The recognition stage of PLDA compares the likelihood of the data under N different models which is denoted by $M_{1...N}$. In a closed set identification, the nth model represents the case where probe face x_p matches the nth gallery face so nth identity variable h_n is responsible of generating probe feature vector i.e. h_p = h_n while M_0 depicts the case where two faces belongs to different people having different identity variables.

The evidence for the model M_0 i.e non-match and M_n i.e. match can be given as

$$P_r(x_{1,...,N}, x_p|M_0) = \prod_{m=1}^{N} P_r(x_m)P_r(x_p) \quad (12)$$

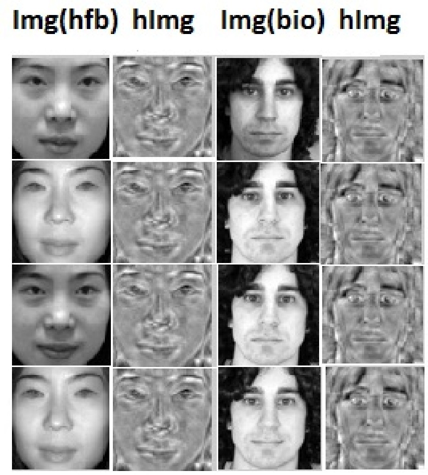

Fig. 3. The 1st and 3rd columns show images of individual from HFB (VIS-NIR) and Biosecure database and 2nd and 4th columns shows its learned identity variable.It is evident that different modalities images of an individual is represented by same LIV.

$$P_r(x_{1,...,N}, x_p|M_n) = \prod_{m=1, m\neq n}^{N} P_r(x_m)P_r(x_p, x_n) \quad (13)$$

where

$$P_r(x_m) = \iint P_r(x_m, h_m, w_m)dh_m dw_m \quad (14)$$

$$P_r(x_p) = \iint P_r(x_p, h_p, w_p)dh_p dw_p \quad (15)$$

$$P_r(x_p, x_n) = \iiint P_r(x_p, x_n, h_n, w_p, w_n)dh_n dw_p dw_n \quad (16)$$

The evaluation of above integrals is basically the evaluation of likelihood that N images share the same identity variable regardless of noise variables. We provide the generative equations for N images that share the same identity variable ,h, irrespective of the noise variables $w_1, ..., w_N$ and form a composite system

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} = \begin{bmatrix} \mu \\ \mu \\ \vdots \\ \mu \end{bmatrix} + \begin{bmatrix} F & G & 0 & \cdots & 0 \\ F & 0 & G & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ F & 0 & 0 & \cdots & G \end{bmatrix} \begin{bmatrix} h \\ w_1 \\ w_2 \\ \vdots \\ w_N \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_N \end{bmatrix} \quad (17)$$

The above formulation can be rewritten as

$$x' = \mu' + Ay + \epsilon' \quad (18)$$

The probabilistic form of this composite model is

$$P_r(x'|y) = g_{x'}[\mu + Ay, \Sigma'] \quad (19)$$

$$P_r(y) = g_y[0, I] \quad (20)$$

This now transforms into standard factor analyzer whose likelihood is well established i.e

$$P_r(x_{1...N}) = P_r(x') = g_{x'}[\mu' + AA^T + \Sigma'] \quad (21)$$

3. EXPERIMENTS AND RESULTS

The robustness of PLDA is tested on two different types of intermodality scenarios i.e Visual vs. NIR, Low resolution (Web Camera) vs. High resolution (Digital Camera). In testing , intensity and LBP features [15] are employed and rank one recognition rates are mentioned. Figure 4 shows different samples of HFB and Biosecure face database.

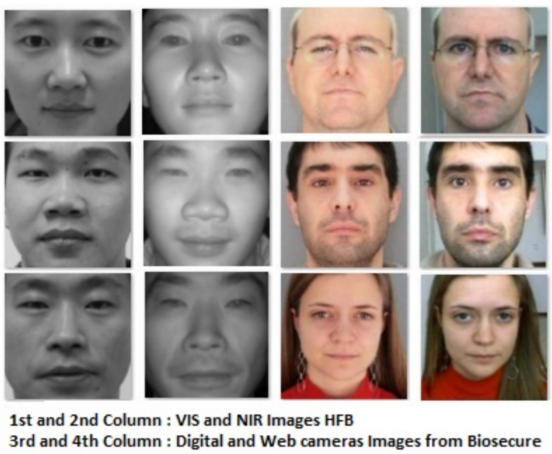

Fig. 4. Examples images from HFB and Biosecure Face database.

3.1. Protocol I and II of HFB VIS-NIR Face Database

In order to test the validity of PLDA in intermodality face matching problem, we employ the same protocols setup of [7]. The test is carried on VIS-NIR HFB face database [16]. The HFB contains 202 individuals with total 5097 images out of which 2095 and 3002 images are of visual and NIR modality, respectively. In protocol I, the training set comprises of 1062 VIS and 1487 NIR images of 202 subjects randomly selected while the test set is made up of gallery images from VIS and probe from NIR (test set is not used in training set). In protocol II, the training set is made up of 1438 VIS and 1927 NIR of 168 subjects while test set comprises of images from 174 persons using one gallery and probe images not included in training stage. The samples of VIS-NIR database are of the size 128 x 128 cropped using eye co-ordinates. For intensity and LBP features, all the images are resized to 32 x 32. Table 1 compares the recognition rates on HFB using our proposed PLDA and other methods. PLDA reports consistent results

Table 1. Results for the HFB Evaluation protocol 1 and 2.

Method	Intensity/P1	LBP/P1	Intensity/P2	LBP/P2
LDA [13]	98.01	98.74	64.51	79.03
CDFE [8]	97.21	99.73	54.87	62.82
LCSR [7]	97.48	99.40	75.65	93.84
LDSR [7]	97.54	99.80	73.96	94.04
KDSR [7]	98.34	99.73	77.04	**95.33**
Proposed PLDA	**99.50**	**100**	**78.74**	90.40

on both protocols with intensity and LBP based features. In protocol I and II on intensity and LBP features, it outperforms all the stat-of-art heterogeneous face methods except in PII LBP and reflects its effectiveness over other state-of-art approaches. PLDA validates the LBP feature superiority over other features as it reflects increase in recognition rates.

3.2. Protocol I and II of Biosecure Face Database

The Biosecure Database [17] contains 420 subjects with 12 samples taken in two sessions. Each session has 6 samples from each individual. Two samples has been captured with webcam while rest with digital camera consisting of flash and non-flash versions in each session. Biosecure database is regarded as multiscenario and multienvironment database. We normalize all images using eye co-ordinates. All images are resized to 32x32 for experiments using intensity and LBP feature vectors. In protocol I and II, four images out of six from each session are used to create the training dataset of 300 individuals. The leftover images one from digital camera and other from webcam make up gallery and probe datasets, respectively. Table 2 reports the rank-1 recognition rates by comparing PLDA with other methods on both protocols.

Table 2. Results for the Biosecure protocol 1 and 2.

Method	Intensity/P1	LBP/P1	Intensity/P2	LBP/P2
PCA [18]	20.00	22.70	25.33	22.33
LDA [13]	23.00	5.33	30.33	9.00
KDA [19]	46.67	5.33	55.33	5.33
Proposed PLDA	**88.50**	**92.50**	**90.00**	**94.67**

PLDA method significantly generates very promising results on this database. It is evident from the results that this approach clearly outperforms all competing methods with very good margin. It is to be noted that none of the approaches in intermodality matching reports results on the Biosecure face database comparing webcam (low resolution) and digital camera (high resolution).

4. CONCLUSION

In this paper, we utilize the efficacy of PLDA method based on posterior probabilities to LDA in the intermodality face matching problem. PLDA shows some consistent results on two different types of intermodality problems involving visual vs NIR and digital camera vs web camera. It is also to be noted that many heterogenous face matching approaches use variety of statistical learning approaches but PLDA , a generative probabilistic approach is applied first time in this domain. Although, we have received comparable and competitive results but considerable gain can be attained by using local PLDA models i.e. separate PLDA model may be built on each manual or automatic annotations of the face. The possible future direction of this method will be to introduce Bayesian regularization and to apply it in different variants of NIR i.e. SWIR, MWIR etc. and source identification.

5. REFERENCES

[1] T. Xiaoou and X. Wang, "Face sketch recognition.," *IEEE Transaction, Circuits and Systems for Video Technology*, vol. 14, pp. 50–57, 2004.

[2] Q. Liu, X. Tang, H. Jin, H. Lu, and S. Ma, "A nonlinear approach for face sketch synthesis and recognition," *CVPR*, vol. 1, pp. 1005–1010, 2005.

[3] X. Wang and X. Tang, "Face photo-sketch synthesis and recognition," *PAMI*, vol. 31, pp. 1955–1967, 2009.

[4] S. Liao, D. Yi nad Z. Lei, R. Qin, and S. Z. Li, "Heterogeneous face recognition from local structures of normalized appearance," *Springer, In Advances in Biometrics*, pp. 209–218, 2009.

[5] B. Klare and A. K. Jain, "Heterogeneous face recognition: Matching nir to visible light images," *ICPR*, pp. 1513–1516, 2010.

[6] B. Klare and Z. Li andA. K. Jain, "Matching forensic sketches to mug shot photos," *PAMI*, vol. 33, pp. 639–646, 2011.

[7] X. Huang, Z. Lei, X. Wang, and S. Z. Li, "Regularized discriminative spectral regression method for heterogeneous face matching," *ITIP*, vol. 22, pp. 1513–1516, 2013.

[8] D. Lin and X. Tang, "Inter-modality face recognition," *ECCV*, pp. 13–26, 2006.

[9] R. Fergus, P. Perona, and A. Zisserman, "Object class recognition by unsupervised scale invariant learning," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition*, pp. 264–271, 2003.

[10] M. Isard and A. Blake, "Condensation conditional density propagation for visual tracking," *IntŠl J. Computer Vision*, vol. 29, pp. 5–28, 1998.

[11] C. Rother, V. Kolmogorov, and A. Blake, "Grabcut: Interactive foreground extraction using iterated graph cuts," *Proc. ACM SIGGRAPH*, pp. 309–314, 2004.

[12] P. Li, Y. Fu, U. Mohammed, J. H. Elder, and S.J.D. Prince, "Probabilistic models for inference about identity," *PAMI*, vol. 34, no. 1, pp. 144–157, 2012.

[13] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs fisherfaces: recognition using class specific linear projection," *PAMI*, vol. 19, no. 7, pp. 711–720, 1997.

[14] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Proc. Roy. Stat. Soc. B*, vol. 39, pp. 1–38, 1977.

[15] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns : Application to face recognition," *PAMI*, vol. 28, no. 12, pp. 2037–2041, 2006.

[16] S. Z. Li, Z. Lei, and M. Ao, "The hfb face database for heterogeneous face biometrics research," *Proc. 6th IEEE Workshop Object Track. Class. Beyond Visible Spectrum*, pp. 1–8, 2009.

[17] J. O. Garcia, J. Fierrez, F. A. Fernandez, J. Galbally, M. R. Freire, J. G. Rodriguez, ..., and Savran, "The multiscenario multienvironment biosecure multimodal database (bmdb)," *PAMI*, vol. 32, pp. 1097–1111, 2010.

[18] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[19] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K. R. Mullers, "Fisher discriminant analysis with kernels," *Proc. IEEE Neural Networks for Signal Processing*, pp. 41–48, 1999.