An Improved Coupled Spectral Regression for Heterogeneous Face Recognition

Zhen Lei[1] Changtao Zhou[1] Dong Yi[1] Anil K. Jain[2] Stan Z. Li[1*]
[1]Center for Biometrics and Security Research & National Laboratory of Pattern Recognition,
Institute of Automation, Chinese Academy of Sciences,
95 Zhongguancun Donglu, Beijing 100190, China.
[2]Michigan State University, East Lansing, MI 48824, U.S.A.
{zlei,ctzhou,dyi,szli}@cbsr.ia.ac.cn, jain@cse.msu.edu

Abstract

Coupled spectral regression (CSR) is an effective framework for heterogeneous face recognition (e.g., visual light (VIS) vs. near infrared (NIR)). CSR aims to learn different projections for different face modalities respectively to find a common subspace where the samples of different modalities from the same class are as close as possible. In original CSR, the projection for one modality is supposed to be represented by the data from the same modality. In this paper, we show that not only the samples of the same modality, but also all samples from different modalities are useful to learn the projection. Based on this assumption, we propose an improved coupled spectral regression (ICSR) approach which assumes the projections are linearly represented by all samples. Moreover, in order to improve the generalization capability, the locality information among samples is considered during the ICSR learning. Experiments on PIE, Multi-PIE and CASIA-HFB face database show that the proposed ICSR enhances the heterogeneous face recognition performance compared with the original CSR and validates the effectiveness of the proposed method.

1. Introduction

Face recognition has attracted much attention due to its potential value for applications and its theoretical challenges. Although face recognition under controlled environments has been well addressed, its performance in real world application is still far from satisfactory. One of the main problems is that the quality of probe images and gallery ones differs too much so that the face recognition performance decreases dramatically. For example, in video surveillance scenario, the gallery images are usually high resolution photos, while the probe ones are of low resolution or even near infrared (NIR) images. In the law enforcement case, the sketch images are usually compared with the photos. All of these introduce a number of challenges for face recognition.

Up to now, many face representation approaches have been introduced. In this paper, we mainly focus on subspace learning method, which is one of the most widely used methods [24, 11]. Typical subspace methods include the well known Principal Component Analysis (PCA) [17], Linear Discriminate Analysis (LDA) [1], Independent Component Analysis (ICA) [3] etc. PCA provides an optimal linear transformation from the original image space to an orthogonal eigenspace with reduced dimensionality in terms of the least mean square reconstruction error. LDA seeks a linear transformation by maximizing the ratio of between-class variance and within-class variance. ICA is a generalization of PCA, which is sensitive to the high-order correlation among the image pixels. Recently, Yan et al. [20] re-interpret the subspace learning from the graph embedding view so that various methods, such as PCA, LDA, LPP [6], etc. can all be interpreted under this framework.

However, traditional subspace learning methods do not perform well on heterogeneous face recognition problem, where the quality of gallery and probe images varies too much. There are mainly two category of methods to deal with this problem. The first category of methods first tries to reduce the gap in appearance of face images and then applies the traditional face recognition methods. The second one learns a common discriminant subspace where the samples of different qualities are well classified. In the former category, Tang and Wang [16] develop eigen-transform method to synthesize sketch image from target photo and perform recognition between pseudo-sketch image and real probe sketch. Liu et al. [14] propose a local linear preserving method to synthesize sketch images from photos and a nonlinear discriminant analysis is used to recognize the sketches. Wang and Tang [19] utilize MRF modeling to synthesize sketch/photo from photo/sketch images. In the latter

*Corresponding author.

category, Lin and Tang [13] propose a common discriminant feature extraction (CDFE) method to transform query faces captured using near infrared or sketch images and target faces of visible spectrum into a common discriminant feature subspace, where the ratio of between scatter matrix and within scatter matrix are maximized. Although CDFE achieves high recognition rate on training set, its generalization performance is poor. Yi et al. [22] utilize canonical correlation analysis (CCA) to exploit the essential correlations in PCA [17] and LDA [1] subspaces of VIS and NIR images and Yang et al. [21] propose regularized kernel CCA to learn the relationship between VIS and 3D data spaces. However, their method does not consider class label information in CCA process and thus it does not utilize important information helpful for classification. Recently, Jain et al. [8, 9] have presented systematic solutions to match heterogeneous faces.

This work focuses on methods in the second category, especially subspace related methods. Recently, Lei and Li [10] have proposed coupled spectral regression (CSR) framework to deal with heterogeneous face recognition problem. Coupled projections for different modalities are learnt respectively to project heterogeneous data onto a common discriminant subspace. The method is both effective and efficient. However, there are at least two shortcomings in original CSR. First, the projection for one modality is constructed based on the samples of the same modality. The samples from different modalities are not utilized. Second, in original CSR, the locality information among samples are not well explored. In this paper, we propose an improved coupled spectral regression (ICSR) approach to overcome these drawbacks. The projections of ICSR are represented based on all samples from all modalities (see difference of CSR and ICSR in Fig. 1). Therefore, more discriminative information hidden in samples between different modalities is expected to be explored. Furthermore, the local sample distribution is modeled and integrated into the ICSR learning to improve the generalization performance.

The remainder of this paper is organized as follows. Section 2 briefly reviews the original coupled spectral regression method. Section 3 details the improved coupled spectral regression algorithm. Experiments on PIE, Multi-PIE and CASIA-HFB face databases are illustrated in Section 4 and in Section 5, we conclude the paper.

2. Coupled Spectral Regression

Coupled spectral regression is an extension of spectral regression. It tries to learn projections for different modalities to project the data from different modalities respectively onto a common discriminant subspace. From the view of graph embedding and spectral regression, the traditional LDA can be formulated as a regression problem to find the mapping from the original data space to the low embeddings in subspace. In [2], it has proved that LDA is equivalent to spectral regression when the low embeddings $Y = [y_1; y_2; \cdots; y_c]$ are defined as

$$y_t = [\underbrace{0, \ldots, 0}_{\sum_{i=1}^{t-1} m_i}, \underbrace{1, \ldots, 1}_{m_t}, \underbrace{0, \ldots, 0}_{\sum_{i=t+1}^{c} m_i}]; t = 1, \ldots, c \quad (1)$$

where c is the number of classes and m_i is the number of i-th class samples. Since the samples of different modalities have different distributions in data space, their projections should be different. Coupled spectral regression aims to learn the projections W^g and W^p for modality g and p, respectively to satisfy $Y^g = {W^g}^T X^g$ and $Y^p = {W^p}^T X^p$, where X^g and X^p denote samples from two modalities and Y^g and Y^p are corresponding low embedding representations extracted from Y. By introducing appropriate regularization, both linear and kernel based nonlinear CSR solutions (LCSR and KCSR) can be derived in a closed form. Please refer to [10] for the details.

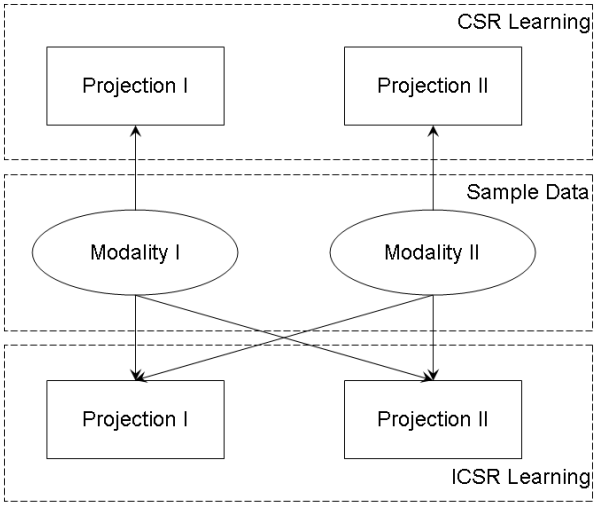


Figure 1. Data flow diagram for ICSR and CSR. In original CSR learning, the projections for modality I or II are represented based on the data from modality I or II, respectively. In ICSR, the coupled projections are constructed based on all data from both modality I and II.

3. Improved Coupled Spectral Regression

There are two shortcomings of original CSR. First, it assumes the projection for one modality is represented by the samples of the same modality. In fact, the samples from the other modality can also make contribution to this projection. Second, in original spectral regression, the sample locality information is not sufficiently explored.

Without loss of generality and for ease of representation, we take VIS vs. NIR faces matching as an example to describe our algorithm. Suppose the sample

data of two modalities are $X^g = [X_1^g, X_2^g, \cdots, X_{N_1}^g]$ and $X^p = [X_1^p, X_2^p, \cdots, X_{N_2}^p]$, where g and p are indicators of different modalities and N_1, N_2 are the number of samples. We can first perform linear or nonlinear mapping to transform the data to a more separable space and denote $\phi(X)^g = [\phi(X)_1^g, \phi(X)_2^g, \cdots, \phi(X)_{N_1}^g]$ and $\phi(X)^p = [\phi(X)_1^p, \phi(X)_2^p, \cdots, \phi(X)_{N_2}^p]$ as the representations of transformed data. From linear algebra [4], the projections W^g and W^p for VIS and NIR faces can be linearly represented as $W^g = \sum \alpha_i^g \phi(X)_i^g + \sum \alpha_i^p \phi(X)_i^p$ and $W^p = \sum \beta_i^g \phi(X)_i^g + \sum \beta_i^p \phi(X)_i^p$. Note that it is different from the original CSR, where the projections of VIS/NIR faces are supposed to be linearly represented by the face data belonging to VIS/NIR, respectively. That is $W^g = \sum \alpha_i^g \phi(X)_i^g$ and $W^p = \sum \beta_i^p \phi(X)_i^p$. In our formulation, all samples from VIS and NIR make contribution to the coupled projections. Therefore, more information hidden in samples between VIS and NIR is expected to be utilized. Denoting $\phi(X) = [\phi(X)^g, \phi(X)^p]$, $A = [\alpha_1^g, \cdots, \alpha_{N_1}^g, \alpha_1^p, \cdots, \alpha_{N_2}^p]^T$, $B = [\beta_1^g, \cdots, \beta_{N_1}^g, \beta_1^p, \cdots, \beta_{N_2}^p]^T$, we have

$$\begin{aligned} W^g &= \phi(X)A \\ W^p &= \phi(X)B \end{aligned} \tag{2}$$

The objective function of ICSR can then be formulated as

$$\begin{aligned} J =& \frac{1}{N_1}||Y^g - W^{gT}\phi(X)^g||^2 + \frac{1}{N_2}||Y^p - W^{pT}\phi(X)^p||^2 \\ =& \frac{1}{N_1}||Y^g - A^T\phi(X)^T\phi(X)^g||^2 \\ &+ \frac{1}{N_2}||Y^p - B^T\phi(X)^T\phi(X)^p||^2 \end{aligned} \tag{3}$$

Locality constraint. In original spectral regression method, the samples from the same class are projected onto the same point in target subspace. The sample locality information which is important for classification [7] is ignored. In this work, we model the sample locality information in transformed space and impose it onto the ICSR learning. Different from LPP which tries to preserve the sample locality structure in the reduced subspace, we impose the locality constraint onto the elements of combination coefficient vectors. In our formulation, A and B are coefficient vectors that combine the samples to form the projective directions. The basic idea is that if the two samples are close, their corresponding coefficients should be similar; otherwise the coefficients are independent to some extent. In this way, we successfully transfer the sample locality information to the elements of coefficient vectors. Defining the similarity S_{ij} between two samples as

$$S_{ij} = \begin{cases} \phi(X)_i^T \phi(X)_j & \text{if } \phi(X)_i, \phi(X)_j \text{ are neighbors} \\ & \text{and from the same class} \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

the objective function of locality constraint can be formulated as

$$J_l = A^T L A + B^T L B \tag{5}$$

where $L = S - D$ is the Laplacian matrix over the samples and D is a diagonal matrix in which $D_{ii} = \sum_{j \neq i} S_{ij}$. It is easy to check that by minimizing J_l, the difference of coefficients whose corresponding samples are similar would be small, which is consistent with our motivation.

Consistency constraint. As introduced in [10], although the projections for VIS and NIR are different, they are generated and used to describe the same object (e.g., face), and hence the projections for different modalities should not differ too much. In this work, we impose a regularization term in the ICSR learning to penalize the difference of projections. That is,

$$J_c = ||W^g - W^p||^2 = ||\phi(X)A - \phi(X)B||^2 \tag{6}$$

ICSR learning. Introducing J_l and J_c in Eq. 3, the objective function of improved CSR (ICSR) can be reformulated as

$$\begin{aligned} J =& \frac{1}{N_1}||Y^g - A^T\phi(X)^T\phi(X)^g||^2 + \frac{1}{N_2}||Y^p - B^T\phi(X)^T\phi(X)^p||^2 \\ &+ \lambda(A^T L A + B^T L B) + \eta||\phi(X)A - \phi(X)B||^2 \end{aligned} \tag{7}$$

where the first two terms are data fitting items and the last two terms are the locality and consistency constraints that help to improve the generalization performance of the solution. λ and η are parameters controlling the trade-off between the data fitting accuracy and the generalization capability. By setting the derivatives of objective function with respect to A and B to zero, we have

$$\begin{aligned} \Theta_1 A &= \phi(X)^T\phi(X)^g Y^{gT} + N_1\eta\phi(X)^T\phi(X)B \\ \Theta_2 B &= \phi(X)^T\phi(X)^p Y^{pT} + N_2\eta\phi(X)^T\phi(X)A \end{aligned} \tag{8}$$

where

$$\begin{aligned} \Theta_1 &= \phi(X)^T\phi(X)^g\phi(X)^{gT}\phi(X) + N_1\lambda L + N_1\eta\phi(X)^T\phi(X) \\ \Theta_2 &= \phi(X)^T\phi(X)^p\phi(X)^{pT}\phi(X) + N_2\lambda L + N_2\eta\phi(X)^T\phi(X) \end{aligned} \tag{9}$$

With proper matrix manipulation, we can obtain the solution A and B as

$$\begin{aligned} A =& \Omega_1^{-1}(\phi(X)^T\phi(X)^g Y^{gT} \\ &+ N_1\eta\phi(X)^T\phi(X)\Theta_2^{-1}\phi(X)^T\phi(X)^p Y^{pT}) \\ B =& \Omega_2^{-1}(\phi(X)^T\phi(X)^p Y^{pT} \\ &+ N_2\eta\phi(X)^T\phi(X)\Theta_1^{-1}\phi(X)^T\phi(X)^g Y^{gT}) \end{aligned} \tag{10}$$

where

$$\begin{aligned} \Omega_1 &= \Theta_1 - N_1N_2\eta^2\phi(X)^T\phi(X)\Theta_2^{-1}\phi(X)^T\phi(X) \\ \Omega_2 &= \Theta_2 - N_1N_2\eta^2\phi(X)^T\phi(X)\Theta_1^{-1}\phi(X)^T\phi(X) \end{aligned} \tag{11}$$

After obtaining A and B, one can get the projections W_g and W_p for different modalities via Eq. 2.

4. Experiments

As shown in Eqs. 10 and 11, the solution to ICSR can be represented as a series of inner product of sample vectors. Therefore, we can use kernel trick as in SVM [18] to represent the data transformation ϕ implicitly. The RBF kernel $k(X_i, X_j) = exp(-||X_i - X_j||^2/\sigma)$ is utilized in the following experiments.

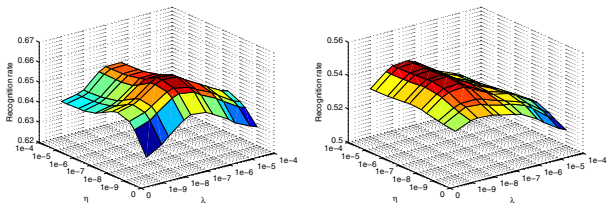
Two heterogeneous face recognition cases (*i.e.*, high resolution vs. low resolution and VIS vs. NIR) are considered in the experiments. For high resolution vs. low resolution, we conduct comparison experiments on Multi-PIE face databases [5]. For VIS vs. NIR, the experiments are conducted on CASIA-HFB face database [12], where the VIS and NIR images from the same persons are recorded.

4.1. Parameter selection

There are three parameters in ICSR. One is σ in RBF kernel and the other two are the trade-off parameters λ and η in Eq. 7. For σ, we set it to $5e7$ empirically according to the average distance of samples [23]. For λ and η, we select the optimal values for high resolution vs. low resolution face recognition on PIE database and the same values are also applied to VIS vs. NIR situation.

The PIE database [15] consists of $41,368$ images from 68 people under different poses, illumination conditions and expressions. Five near frontal poses (C05, C07, C09, C27, C29) and all the images under different illuminations and expressions are selected. There are 170 images for each individual. The images are randomly partitioned into gallery and probe sets. Specifically, 5 images for each person are selected to construct the gallery set and the remaining images are used to construct the probe set. For every image, the high resolution image is cropped into 32×32 size and the low resolution image is obtained by first downsampling the high resolution one to the low resolution size and then upsampling it to the 32×32 size. Two low resolution sizes 16×16 and 8×8 are tested in the experiment. In training phase, the high and low resolution images in gallery set are used. In testing phase, the high resolution images in gallery set are registered and low resolution images in probe set are tested. The random split is conducted 10 times and the mean recognition rate is reported.

We select the value of λ and η in the range $\{1e{-}9, 5e{-}9, 1e{-}8, 5e{-}8, 1e{-}7, 5e{-}7, 1e{-}6, 5e{-}6, 1e{-}5, 5e{-}4, 1e{-}4\}$. Fig. 2 shows the recognition rate trend with respect to λ and η for 16×16 and 8×8 resolutions, respectively. For 16×16 resolution, the best accuracy is achieved when both λ and η are chosen to be $1e{-}7$, while for 8×8 images, the optimal values of λ and η are $5e{-}7$ and $5e{-}8$, respectively. The optimal choice of parameters for 16×16 and 8×8 resolutions are similar and in the following experiment, the σ, λ, η for ICSR are fixed as $5e7$, $1e{-}7$ and $1e{-}7$, respectively.



(a) 16×16 resolution (b) 8×8 resolution

Figure 2. Recognition rate trend of ICSR with respect to λ and η.

4.2. Multi-PIE

In this part, we compare the proposed ICSR with LDA, LCSR and KCSR methods in the case of high resolution vs. low resolution face recognition [1] on Multi-PIE face database. For LCSR and KCSR, the parameters are adopted according to the recommended values in [10]. For LDA, we combine the heterogeneous data (high resolution and low resolution images) together and train a single projection matrix for high and low resolution data.

The Multi-PIE database is an extended version of PIE which contains 337 subjects from 4 sessions under different poses, illumination conditions and expressions. The frontal views with neutral expression under different illuminations are selected. There are $18,420$ images in total from 337 persons with $20-60$ images per person. In this experiment, we divide the database into three sets, namely training set, gallery set and probe set. In training set, 100 subjects with 20 images per person are selected. We select another 100 persons to construct the gallery and probe sets. For ease of representation, we use GN to denote that N images per person are selected to construct the gallery set and the remaining images are used as the probe set. There is no intersection between training set and gallery/probe set. In testing phase, the low resolution images in probe set are compared with high resolution ones in gallery set. The random split is run 10 times and the mean recognition rate along with the std. are reported.

In our experiments, all the images are cropped to 32×32 size. For the low resolution $M \times M$ images, the images are first downsampled to $M \times M$ and then they are upsampled to 32×32 size. Fig. 3 shows high resolution image examples with four low resolution sizes 16×16, 8×8, 6×6 and 4×4, respectively.

Table 1 lists the recognition rates of LDA, LCSR, KCSR and ICSR on Mutli-PIE database. From the results, we can see that the performances of LCSR, KCSR and ICSR

[1] Note that for high resolution vs. low resolution face recognition, there are many methods utilizing super-resolution technique to synthesize high resolution images, followed by face recognition between high resolution images. In this experiment, we take high resolution vs. low resolution as an example of heterogeneous face recognition and focus on subspace related methods.

Table 1. Performance comparison (mean accuracy±std%) of high vs. low resolution face recognition on Multi-PIE database. Four low resolutions 16×16, 8×8, 6×6 and 4×4 are tested.

Methods	G2	G5	G10
LDA	63.67±2.62	77.19±1.69	84.66±0.94
LCSR	82.52±1.41	92.58±0.69	96.18±0.20
KCSR	84.57±1.42	93.40±0.40	96.71±0.37
ICSR	**85.53**±1.47	**94.02**±0.38	**97.08**±0.30

(a) 16×16

Methods	G2	G5	G10
LDA	33.53±2.43	41.48±2.87	47.26±2.19
LCSR	64.54±1.11	77.09±1.25	82.45±0.70
KCSR	73.92±1.31	85.70±0.63	91.63±0.50
ICSR	**74.52**±1.37	**86.71**±0.90	**92.55**±0.47

(b) 8×8

Methods	G2	G5	G10
LDA	15.21±1.21	17.84±1.88	20.37±1.22
LCSR	39.96±1.59	48.66±1.15	51.51±1.18
KCSR	51.99±1.70	63.79±1.44	71.41±1.10
ICSR	**57.15**±1.57	**70.54**±1.10	**79.34**±0.87

(c) 6×6

Methods	G2	G5	G10
LDA	9.50±1.04	10.61±1.23	11.07±0.93
LCSR	17.83±1.03	20.15±0.99	20.81±1.21
KCSR	30.97±1.27	37.75±0.98	42.91±0.79
ICSR	**36.26**±0.86	**50.73**±0.94	**62.35**±0.87

(d) 4×4

are significantly better than that of LDA, indicating that the coupled projection strategy for heterogeneous face recognition is effective. The KCSR and ICSR which adopt nonlinear technique perform better than LCSR. The RBF kernel is helpful to improve the separability of samples. Comparing ICSR with KCSR, their performances are similar for 16×16 and 8×8 resolutions. However, in the lower resolution of 6×6 and 4×4, ICSR outperforms KCSR significantly. ICSR improves the recognition rate by $6 \sim 20$ percent when the low resolution size is 4×4. In KCSR, the projections for high and low resolution data are derived from the high and low resolution samples, respectively. ICSR learns the coupled projections by utilizing high and low resolution images together. The good performance of ICSR, especially in lower resolution cases, validates that it is possible to explore more discriminative information between heterogeneous data. This along with the locality constraint are helpful to improve the generalization capability. Overall, ICSR achieves better recognition performance than the original CSR.

Figure 3. Cropped face image examples. From left to right: high resolution (32×32) followed by low resolution (16×16, 8×8, 6×6, 4×4) images.

4.3. CASIA-HFB

The CASIA-HFB database was collected by CBSR for heterogeneous biometric research. There are 5,097 images in total, including 2,095 VIS and 3,002 NIR images from 202 persons in the database. In this experiment, two protocols are designed to evaluate the performance of different methods. In protocol I, we randomly select 1,062 VIS and 1,487 NIR images from 202 subjects and the remaining images form the testing set. The persons in the testing set are all included in the training set. In protocol II, the training set contains 1,438 VIS and 1,927 NIR images from 168 persons and the testing set contains 657 VIS and 1,075 NIR ones from 174 persons. The persons in the testing set are partially included in the training set. In testing phase, the NIR images are compared with VIS images and the recognition rate and receiver operating characteristic (ROC) performance are reported. All images are cropped into 32×32 size according to the automatically detected eye coordinates. Fig. 4 shows some VIS and NIR face images from the database.

Figure 4. Cropped VIS and NIR face image examples. The left three columns are VIS images and the right three columns are NIR images.

We compare ICSR with LDA, CDFE [13], CCA+PCA, CCA+LDA [22], LCSR and KCSR methods. The parameters of these methods are set to the recommended values in their papers. For LDA method, the VIS and NIR images are combined together to learn a single projection for VIS and NIR images. The performance (rank-1 recognition rate, verification rate (VR) at false accept rate (FAR) of 0.001 and 0.01) of different methods is reported in Tables 2

and 3. In protocol I, where the subjects in testing set are all included in the training set, the performance of most methods is acceptable. The methods using coupled projections achieve better performance than the traditional ones. In protocol II, where the test subjects are partially included in the training set, the performance of CDFE, PCA+CCA and LDA+CCA is degraded dramatically, even worse than LDA. On the contrary, CSR related methods (LCSR, KCSR and ICSR) possess better generalization capability. The proposed ICSR, which sufficiently explores the discriminant information between different modalities and incorporates the sample locality information, achieves the highest recognition rate among all the methods.

Table 2. Performance comparison (%) on VIS-NIR database with protocol I.

Method	Rec. Rate	VR@FAR=0.01	VR@FAR=0.001
LDA	98.21	91.43	88.14
CDFE	97.21	97.47	95.38
PCA+CCA	95.42	95.56	91.96
LDA+CCA	97.74	95.37	91.99
LCSR	97.48	97.56	95.42
KCSR	97.34	97.94	96.53
ICSR	**98.54**	**99.49**	**99.00**

Table 3. Performance comparison (%) on VIS-NIR database with protocol II.

Method	Rec. Rate	VR@FAR=0.01	VR@FAR=0.001
LDA	65.41	38.02	25.80
CDFE	45.43	29.55	14.15
PCA+CCA	51.09	34.40	16.40
LDA+CCA	41.55	25.86	11.56
LCSR	75.65	**65.88**	44.71
KCSR	73.06	64.94	44.73
ICSR	**77.53**	65.59	**47.39**

5. Conclusions

This paper proposes an improved coupled spectral regression (ICSR) method. In original CSR, the coupled projections are derived from the corresponding modality data respectively. The mutual information between different modalities are not sufficiently explored. In ICSR, both coupled projections are supposed to be represented by all samples from different modalities. In this way, the discriminative information hidden in samples are sufficiently explored. Moreover, the sample locality information is integrated into ICSR learning to improve the generalization performance of ICSR. Experiments on high vs. low resolution and VIS vs. NIR heterogeneous face recognition cases validate that ICSR does improve the performance of original CSR.

6. Acknowledgement

This work was supported by the Chinese National Natural Science Foundation Project #61070146, #61105023, #61103156, #61105037, National IoT R&D Project #2150510, Chinese Academy of Sciences Visiting Professorship for Senior international Scientists, Grant No. 2011T1G18 and European Union FP7 Project #257289 (TABULA RASA http://www.tabularasa-euproject.org), and AuthenMetric R&D Funds.

References

[1] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *IEEE T-PAMI*, 19(7):711–720, July 1997. 1, 2

[2] D. Cai, X. He, and J. Han. Spectral regression for efficient regularized subspace learning. In *ICCV*, 2007. 2

[3] P. Comon. Independent component analysis - a new concept? *Signal Processing*, 36:287–314, 1994. 1

[4] G. H. Golub and C. F. van Van Loan. *Matrix Computations, Third Edition*. The Johns Hopkins University Press, 1996. 3

[5] R. Gross, I. Matthews, J. F. Cohn, T. Kanade, and S. Baker. Multi-pie. *Image Vision Comput.*, 28(5):807–813, 2010. 4

[6] X. He, S. Yan, Y. Hu, P. Niyogi, and H. Zhang. Face recognition using laplacianfaces. *IEEE T-PAMI*, 27(3):328–340, 2005. 1

[7] X. He, S. Yan, Y. Hu, P. Niyogi, and H. Zhang. Face recognition using laplacianfaces. *IEEE T-PAMI*, 27(3):328–340, 2005. 3

[8] B. F. Klare and A. K. Jain. Heterogeneous face recognition: Matching nir to visible light images. In *ICPR*, 2010. 2

[9] B. F. Klare, Z. Li, and A. K. Jain. Matching forensic sketches to mug shot photos. *IEEE T-PAMI*, 33(3):639–646, 2011. 2

[10] Z. Lei and S. Z. Li. Coupled spectral regressoin for matching heterogeneous faces. In *CVPR*, 2009. 2, 3, 4

[11] S. Z. Li and A. K. Jain (eds.). *Handbook of Face Recognition, Second Edition*. Springer-Verlag, New York, August 2011. 1

[12] S. Z. Li, Z. Lei, and M. Ao. The hfb face database for heterogeneous face biometrics research. In *OTCBVS*, 2009. 4

[13] D. Lin and X. Tang. Inter-modality face recognition. In *ECCV*, pages 13–26, 2006. 2, 5

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

[15] T. Sim, S. Baker, and M. Bsat. The CMU pose, illumination, and expression database. *IEEE T-PAMI*, 25(12):1615–1618, 2003. 4

[16] X. Tang and X. Wang. Face sketch recognition. *IEEE T-CSVT*, 14(1):50–57, 2004. 1

[17] M. A. Turk and A. P. Pentland. Face recognition using eigenfaces. In *CVPR*, pages 586–591, Hawaii, June 1991. 1, 2

[18] V. N. Vapnik. *Statistical learning theory*. John Wiley & Sons, New York, 1998. 4

[19] X. Wang and X. Tang. Face photo-sketch synthesis and recognition. *IEEE T-PAMI*, 31(11):1955–1967, 2009. 1

[20] S. Yan, D. Xu, B. Zhang, H. Zhang, Q. Yang, and S. Lin. Graph embedding and extensions: A general framework for dimensionality reduction. *IEEE T-PAMI*, 29(1):40–51, 2007. 1

[21] W. Yang, D. Yi, Z. Lei, J. Sang, and S. Z. Li. 2d-3d face matching using cca. In *FG*, 2008. 2

[22] D. Yi, R. Liu, R. Chu, Z. Lei, and S. Z. Li. Face matching between near infrared and visible light images. In *ICB*, pages 523–530, 2007. 2, 5

[23] J. Zhang, M. Marszalek, S. Lazebnik, and C. Schmid. Local features and kernels for classification of texture and object categories: A comprehensive study. *IJCV*, 73(2):235–249, 2007. 4

[24] W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Computing Surveys*, pages 399–458, 2003. 1