2014 22nd International Conference on Pattern Recognition

Domain Adaptive Self-Taught Learning for Heterogeneous Face Recognition

Cheng-An Hou
Research Center for IT Innovation
Academia Sinica, Taipei, Taiwan
Email: c.a.andyhou@gmail.com

Min-Chun Yang
Research Center for IT Innovation
Academia Sinica, Taipei, Taiwan
Email: kaikaii@citi.sinica.edu.tw

Yu-Chiang Frank Wang
Research Center for IT Innovation
Academia Sinica, Taipei, Taiwan
Email: ycwang@citi.sinica.edu.tw

***Abstract*—Recognizing image data across different domains has been a challenging task. For biometrics, heterogeneous face recognition (HFR) deals with recognition problems in which training/gallery images are collected in terms of one modality (e.g., photos), while test/probe images are observed in the other (e.g., sketches). In this paper, we present a domain adaptation approach for solving HFR problems. By utilizing external face images (i.e., those collected from the subjects *not* of interest) from both source and target domains, we propose a novel *Domain-independent Component Analysis (DiCA)* algorithm for deriving a common subspace for relating and representing cross-domain image data. In order to introduce improved representation ability, we further advance the *self-taught learning* strategy for learning a domain-independent dictionary in our DiCA subspace, which can be applied to both gallery and probe images of interest to improve representation and recognition. Different from some prior domain-adaptation approaches, we do not require the data correspondences (i.e., data pairs) when collecting external cross-domain image data, nor the label information is needed for learning the common feature space when associating different domains. Thus, our method is practical for real-world cross-domain classification problems. In our experiments, we consider sketch-to-photo and near-infrared (NIR) to visible spectrum (VIS) face recognition problems for evaluating the performance of our proposed approach.**

I. Introduction

Conventional pattern recognition algorithms typically assume that the training and test data for addressing the associated classification task exhibit the same distribution, or are drawn from the same feature space or data domain [1], [2]. However, when the data distribution changes, or when training and test data are collected from different domains (e.g., images captured by different camera views, or features observed in different modalities), one cannot expect that the classifiers learned from training data can be applied for recognizing test data successfully.

Heterogeneous face recognition (HFR) is an example of the aforementioned cross-domain classification problems. It has been an emerging task in biometrics, since face images to be recognized in real-world scenarios are often acquired in terms of different modalities (e.g., visible spectrums (VIS), near-infrared (NIR), or even sketch images). For example, the face of a suspect might be described by witnesses at a crime scene, and thus only a sketch facial image can be obtained. Since the database of criminals typically consists of images in visible spectrums (i.e., photos), it would be a difficult task to apply standard matching or recognition algorithms on such images for determining the identity of the suspect [3].

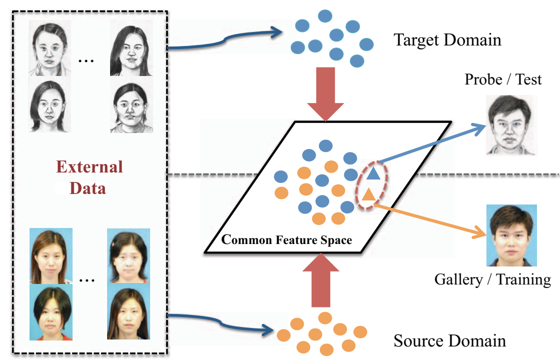


Fig. 1. Illustration of domain adaptation for heterogeneous face recognition. Note that the external face images applied for associating source and target domains are collected from subjects *not* of interest (in circles), while the gallery and probe images are from the subjects to be recognized (in triangles).

Recently, several approaches have been proposed for HFR. In [4], [5], [6], [7], [8], HFR is viewed as an image *synthesis* problem. With the focus on synthesizing the face image from source to target modalities (e.g., sketch to photo), these approaches apply existing classification techniques for recognition after the image of interest is produced. On the other hand, approaches like [9], [10], [11], [12] aim at determining a common feature space for associating cross-modality face images. To be more specific, these methods consider HFR as a *cross-domain image classification* problem, in which training and test data are collected from source and target domains, respectively. Once the common feature space for associating cross-domain data is derived, one can project training (gallery) face images into this space for recognizing projected test (probe) images accordingly. For example, Yi *et al.* [9] utilized cross-domain data and applied canonical correlation analysis (CCA), aiming at determining a common feature space for NIR-to-VIS face recognition. Sharma *et al.* [11] proposed to solve cross-domain face recognition problem based on a Partial Least Square (PLS) framework. Klare and Jain [13], [14] solved HFR problems by learning discriminative projections using multiple visual features from data observed from both image domains. Huang and Wang [12] presented a coupled-dictionary learning algorithm using cross-domain data to derive a common feature space and a dictionary pair for solving sketch-to-photo face recognition.

1051-4651/14 $31.00 © 2014 IEEE
DOI 10.1109/ICPR.2014.529

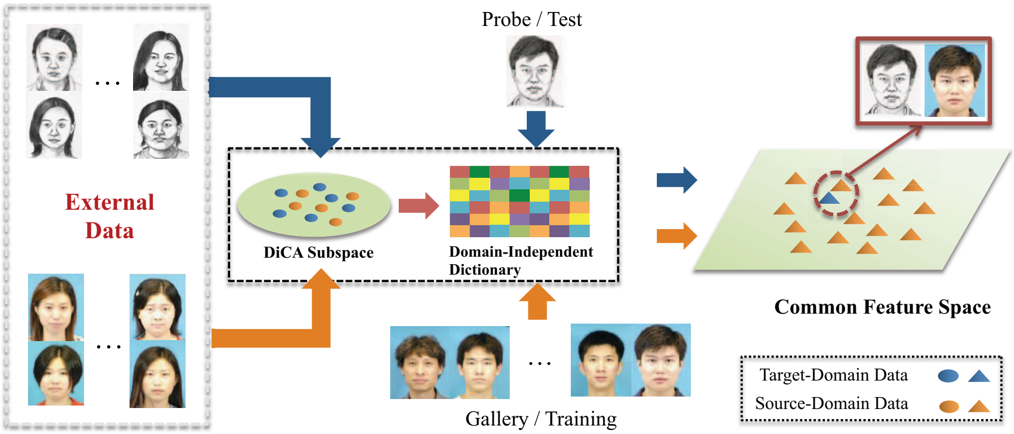


Fig. 2. Flowchart of our domain adaptive self-learning approach for HFR. By utilizing external cross-domain images (in circles) for observing our DiCA subspace, we learn a domain-independent dictionary for representing cross-domain images projected onto this subspace. As a result, the derived subspace and representation can be considered as a common feature space for relating, describing, and recognizing cross-domain images of interest (in triangles).

Although the aforementioned cross-domain approaches have reported promising results for HFR, their direct use of cross-domain *external data* (i.e., subjects *not* of interest and thus *not* to be recognized) for constructing the common feature spaces might be a concern. Take an extreme scenario for example (as depicted in Figure 1), the common feature space derived from cross-domain face images of females is not expected to exhibit sufficient representation abilities for describing those of males from either domain, and thus might produce poor generalization and recognition performance. Another concern of the above methods is their requirement of collecting external data *pairs* when learning the common feature space. This would also limit their uses for solving practical HFR problems.

Based on the above observations, we propose a novel domain adaptive self-taught learning approach for HFR. With the goal of constructing feature spaces for associating cross-domain face images, we propose *Domain-Independent Component Analysis (DiCA)* which does not require the same external face image to be observed in both domains. In order to improve the representation capability of the derived DiCA subspace, we further advance the strategy of *self-taught learning* and learn a domain-independent dictionary for describing cross-domain data in this space. Later in our experiments, we consider two HFR tasks: sketch-to-photo and NIR-to-VIS face recognition problems. Our empirical results will verify the effectiveness and robustness of our proposed method.

The remaining of this paper is organized as follows. Section II reviews prior works on domain adaptation and self-taught learning. Our proposed method for HFR will be detailed in Section III. Experimental results will be presented in Section IV, and finally Section V concludes this paper.

II. Related Works

A. Cross-Domain Recognition via Domain Adaptation

As noted earlier, cross-domain classification problems deal with training and test data collected from different domains or those exhibit different feature distributions. Since re-collecting training data at the feature domain of interest is generally not applicable due to time or complexity costs, transfer learning [15] is recently advanced by researchers in related fields, aiming at transferring the knowledge of the model (e.g., classifier) observed from the source domain to that in the target domain. As a result, the classification, synthesis, etc. tasks at the target domain can be addressed accordingly.

Among various scenarios of transfer learning, we particularly consider *domain adaptation* which focuses on solving the same learning problem across different domains. Take the HFR problem for example, recent approaches like [11], [10], [3] can be viewed as techniques of domain adaptation, which focus on observing a common feature space for cross-domain face images, so that recognition can be performed in the derived space. In particular, methods of [16], [17] focused on determining PCA-based projection matrices, which allow cross-domain data to be projected into a common subspace with preserved data distributions. However, as noted in [3], the aforementioned approaches require the data correspondences (i.e., data pairs) when collecting external cross-domain data, and thus might limit their practical uses.

B. Utilizing External Data for HFR

In order to deal with an insufficient amount of training data for practical recognition problems, a common solution is to collect additional unlabeled data but from the *same* categories of interest, so that techniques of semi-supervised learning can be applied for learning the classifiers. However, if the above data collection strategy is not possible, one typically needs to utilize *external* unlabeled data (i.e., data not of interest, or the distribution of such data is different from that of existing training data), with the goal of learning proper feature representations or classifiers [14], [18], [19].

Take the HFR problem for example, Klare and Jain [14] proposed to utilize external data pairs as a set of prototypes, so that the source-domain training and target-domain test images can be represented in a unified formulation. However, as discussed earlier, the need of collecting external data *pairs* across different domains might not be applicable. Raina *et al.* [18]

proposed *self-taught learning* which applied external data and performed dictionary learning, which allows one to describe existing training and test data for improved recognition.

While the self-taught learning strategy applied in [18] was not particularly proposed for addressing cross-domain recognition problems (i.e., their external data was collected from the same domain as the training/test data was at), we advance this strategy for solving HFR in this paper. To be more specific, we first advance external cross-domain data for observing a proper subspace which associates the source and target domains, without the constraint of data correspondences. We then utilize the self-taught learning strategy for learning an domain-independent dictionary at this space, which can be applied to represent and discriminate between face images of different subjects. In the following section, we will detail our proposed algorithm for associating and representing cross-domain image data, and discuss how to apply our proposed method for solving HFR problems.

III. Our Proposed Domain Adaptive Self-Taught Learning Method

Figure 2 illustrates our proposed method for HFR, which advances a novel domain adaptive self-taught learning for relating and representing cross-domain data. As highlighted in this figure, our method can be viewed as a domain-independent subspace learning approach for associating images in different domains, followed by the learning of a domain-independent dictionary for representing such data. We now detail our proposed method in this section.

A. Motivation for Domain-Independent Component Analysis

Based on the observation of [20] that face images of different subjects are highly correlated in the *same* domain (compared to those of the same subject but across different domains), we propose *Domain-independent Component Analysis* (DiCA) for HFR. The goal of our DiCA is to describe the distribution of cross-domain data in the same feature subspace, while the domain dependency can be automatically disregarded based on *Maximum Mean Discrepancy* (MMD) [21], [16].

Since we apply the self-taught learning strategy when solving HFR problems, we utilize external cross-domain face images for learning the DiCA subspace. In addition to the disregard of data correspondence constraints, another advantage of our proposed method is that *no* class (label) information of cross-domain external data is needed when performing DiCA.

B. DiCA Algorithm

We now detail our DiCA algorithm. As noted above, we require external data for constructing a subspace for relating cross-domain image data for solving HFR. Let $\mathbf{X} = [\mathbf{x}_1, ..., \mathbf{x}_{n_{Ext}}] \in \mathbb{R}^{d \times n_{Ext}}$ as the external data matrix in a d-dimensional space, in which each instance $\mathbf{x}_i$ is collected from either the source and target domain. We note that $n_{Ext} = n_s + n_t$ indicates the total number of external images, where n_s is the number of images in the source domain D_s (e.g., photo images), and n_t is that in the target domain D_t (e.g., sketch or near-infrared images). With the centering matrix $\mathbf{C} = \mathbf{I} - \frac{1}{n_{Ext}}\mathbf{1}$ ($\mathbf{1}$ indicates the matrix of ones), we can calculate the covariance matrix of $\mathbf{X}$ as $\mathbf{XCX}^\top$.

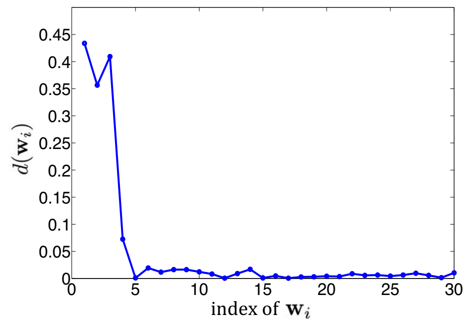


Fig. 3. Distances $d(\mathbf{w}_i)$ between projected cross-domain data using different projection vectors $\mathbf{w}$ determined in (1) using the CASIA NIR-VIS 2.0 dataset.

As pointed out in [20], face images of different subjects are highly correlated in the same domain, while the correlation between images of the same subject but across different domains is much smaller. In other words, when performing the standard Principal Component Analysis (PCA) on $\mathbf{X}$, the first few dominant eigenvectors would correspond to the domain changes instead of describing the data distributions in either domain. Thus, we first derive an orthogonal projection matrix $\mathbf{W} = [\mathbf{w}_1, ..., \mathbf{w}_k] \in \mathbb{R}^{d \times k}$ for $\mathbf{X}$ ($k < d$):

$$\max_{\mathbf{W}^\top\mathbf{W}=\mathbf{I}} tr(\mathbf{W}^\top\mathbf{XCX}^\top\mathbf{W}), \tag{1}$$

It can be seen that the above optimization process effectively performs PCA on the collected cross-domain data, aiming at determining the first k dominant eigenvectors in the derived subspace. In other words, one can apply standard eigen-decomposition techniques and reformulate the above optimization problem as:

$$\mathbf{XCX}^\top\mathbf{W} = \mathbf{WV}, \tag{2}$$

where $\mathbf{V} = diag(v_1, v_2, ..., v_k) \in \mathbb{R}^{k \times k}$ is a diagonal matrix, in which the diagonal elements indicate the associated eigenvalues in a descending order.

As discussed earlier and suggested in [20], we do not expect the direct use of the PCA subspace is sufficient for relating cross-domain data, since the presence of cross-domain information (i.e., domain differences) can be easily observed in the dominant dimensions of the resulting subspace. Therefore, once all k PCA eigenvectors for $\mathbf{X}$ are obtained from (2), our next step is to identify the most dominant ones among these eigenvectors, which correspond to the domain variations instead of data distributions. In our work, we advance the distance measurement based on *Maximum Mean Discrepancy* (MMD) [16], [21] in the resulting PCA subspace, so that the similarity between data samples collected from different domains at each PCA subspace dimension can be determined accordingly. To be more specific, for the ith dimension in the PCA subspace (i.e., projected by $\mathbf{w}_i$), we measure the distance between the means of the projected data from source and target domains as follows:

$$d(\mathbf{w}_i) = \| \frac{1}{n_s} \sum_{x_j \in D_s} \mathbf{w}_i^\top \mathbf{x}_j - \frac{1}{n_t} \sum_{x_k \in D_t} \mathbf{w}_i^\top \mathbf{x}_k \|. \quad (3)$$

With $d(\mathbf{w}_i)$ calculated for each dimension, one can simply identify the dominant dimensions which correspond to the domain changes. Take Figure 3 for example, the separation between projected cross-domain data in the first four dimensions of the deriving PCA subspace is clearly larger than that of the remaining dimensions. This indicates that the first four dimensions in this subspace mainly describe the domain variations, instead of the distribution of face images. As a result, we can simply apply a threshold T for identifying and disregarding the eigenvectors $\mathbf{w}_i$ which are associated with the largest $d(\mathbf{w}_i)$ values. As for the remaining eigenvectors corresponding to the data distribution itself (i.e., with smaller $d(\mathbf{w}_i)$ values), we consider them as *Domain-independent Components* (DiC), which will be applied to relate and represent cross-domain face images. Based on the above discussions, we determine DiC for cross-domain data by:

$$\begin{aligned} &DiC(\mathbf{X}) = \{\mathbf{w}_j | \mathbf{w}_j \subseteq \mathit{columns}(\mathbf{W}), d(\mathbf{w}_j) \leq T\}, \\ &\mathbf{W}_{DiCA} = [\hat{\mathbf{w}}_1, ..., \hat{\mathbf{w}}_{\tilde{k}}] \in \mathbb{R}^{d \times \tilde{k}} \ \mathit{and} \ \hat{\mathbf{w}}_i \in DiC(\mathbf{X}), \forall i. \end{aligned} \quad (4)$$

where $\tilde{k}$ is the number of eigenvectors $\mathbf{w}_i$ whose corresponding $d(\mathbf{w}_i)$ are smaller than threshold T. Once all $\tilde{k}$ DiC components are constructed, the derivation of our DiCA is complete, and external cross-domain data $\mathbf{X}$ can be projected as $\mathbf{Y} = [\mathbf{y}_1, ..., \mathbf{y}_n] = \mathbf{W}_{DiCA}^\top \mathbf{X} \in \mathbb{R}^{\tilde{k} \times n}$ accordingly. We note that, due to the ability of disregarding dominant eigenvectors associated with domain differences, we are able to relate, visualize, and describe cross-domain face images in this final $\tilde{k}$-dimensional DiCA subspace. Figure 4 shows example comparisons between the standard PCA and our DiCA subspaces for cross-domain face images.

C. Learning of Domain-Independent Dictionary for Cross-Domain Data

Although our DiCA subspace can be utilized for describing cross-domain image data, this subspace is constructed by observing external data, which is collected from the subjects not of interest. As a result, it is not desirable to simply project gallery and probe images of the subjects of interest into this subspace for recognition. As discussed in Section II-B, another potential concern for HFR or other practical image recognition problems is that the number of gallery/training images is typically limited, and thus the direct use of the training images for deriving the features and classifiers might limit the performance. This is the reason why we advance the self-taught learning strategy in our method. More precisely, we advocate the learning of domain-independent dictionaries in our DiCA subspace, with the goal of better representing the face images (either from external data or those of interest).

As investigated in [22], [23], [24], dictionary learning has shown promising performance for face recognition problems. While most of prior approaches focused on dictionary learning with improved discriminative capabilities using label or locality information, these techniques are not applicable for HFR

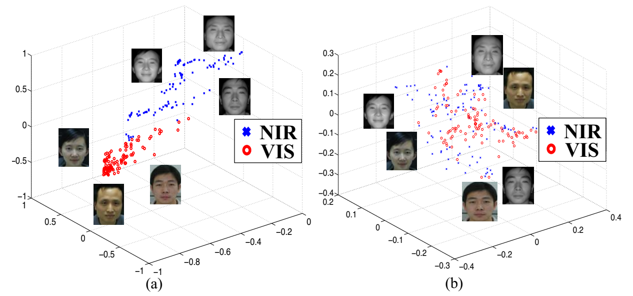


Fig. 4. Cross-domain face images of the CASIA NIR-VIS 2.0 dataset visualized in the subspace spanned by the top three dominant eigenvectors determined by (a) PCA and (b) our proposed DiCA, respectively. Note that the data variation due to domain changes is disregarded in our DiCA subspace.

due to the use of external data. As a result, after projecting cross-domain external images into our DiCA subspace, we focus on learning a domain-independent dictionary via sparse representation for describing such image data:

$$\min_{\mathbf{B}, \mathbf{A}_e} \|\mathbf{Y} - \mathbf{B}\mathbf{A}_e\|_F^2 + \lambda \|\mathbf{A}_e\|_1, \qquad \text{s.t.} \|\mathbf{b}_i\|_2 \leq 1, \forall i, \quad (5)$$

where $\mathbf{B} = [\mathbf{b}_1, ..., \mathbf{b}_m] \in \mathbb{R}^{\tilde{k} \times m}$ is the over-complete dictionary to be learned, in which each $\mathbf{b}_i$ represents the i-th column vector of $\mathbf{B}$. The matrix $\mathbf{A}_e$ indicates the sparse coefficients for each external face image in the DiCA subspace, using the observed dictionary $\mathbf{B}$. The parameter λ regularizes the sparsity of $\mathbf{A_e}$. Since the above objective function is not jointly convex to $\mathbf{B}$ and $\mathbf{A}_e$, one can solve this optimization problem by applying iterative optimization techniques, which update dictionary $\mathbf{B}$ and coefficient $\mathbf{A}_e$ during optimization, respectively. In our work, we directly apply the software tool of SPAMS [25] for solving this dictionary learning problem. Once the dictionary $\mathbf{B}$ is learned, we utilize this dictionary for representing gallery and probe images (collected from the subjects to be recognized) projected into our DiCA subspace. As a result, recognition can be performed in this subspace accordingly.

D. Recognition of Heterogeneous Face Images

From the above discussions, we first perform the proposed DiCA algorithm for deriving a common subspace for relating cross-domain face images, followed by a self-taught learning strategy for observing a domain-independent dictionary for describing cross-domain data. This not only alleviates the concern of using external data for domain adaptation, improved representation and recognition of cross-domain data can also be expected.

We now detail the recognition stage of our proposed method. Once the above learning procedures are complete, we project both source-domain training images (e.g., photos) and target-domain test images (e.g., sketches) in to the DiCA subspace. Using the derived domain-independent dictionary $\mathbf{B}$, we then calculate the corresponding sparse feature representations for the projected data. To be more precise, considering that we have n_g gallery images and n_p probe images for recognition, we extract the features in the DiCA subspace for face images $\mathbf{X}_g \in \mathbb{R}^{d \times n_g}$ and $\mathbf{X}_p \in \mathbb{R}^{d \times n_p}$ from D_s and D_t, respectively:

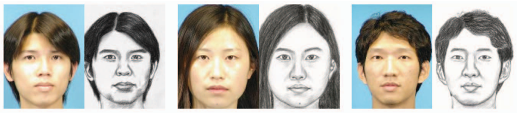


Fig. 5. Example sketch-photo image pairs in the CUFS dataset.

TABLE I. RECOGNITION COMPARISONS ON CUFS.

Method	Recognition Rate (%)	Data Pairs for Ext. Data
CCA [26]	94.6	Y
BLM [27]	94.2	Y
PLS [11]	93.6	Y
Huang *et al.* [12]	97.2	Y
PCA+self-taught learning	97.4	N
Ours	**99.4**	N

$$\underset{\mathbf{A}_g}{\operatorname{argmin}} \|\mathbf{Y}_g - \mathbf{B}\mathbf{A}_g\|_F^2 + \lambda\|\mathbf{A}_g\|_1, \quad (6)$$

$$\underset{\mathbf{A}_p}{\operatorname{argmin}} \|\mathbf{Y}_p - \mathbf{B}\mathbf{A}_p\|_F^2 + \lambda\|\mathbf{A}_p\|_1, \quad (7)$$

where $\mathbf{Y}_g$ and $\mathbf{Y}_p$ are the gallery and probe images observed in the DiCA subspace (i.e., $\mathbf{Y}_g = \mathbf{W}_{DiCA}^\top \mathbf{X}_g$ and $\mathbf{Y}_p = \mathbf{W}_{DiCA}^\top \mathbf{X}_p$). The coefficient matrices $\mathbf{A}_g = [\mathbf{a}_g^1, ..., \mathbf{a}_g^{n_g}] \in \mathbb{R}^{m \times n_g}$ and $\mathbf{A}_p = [\mathbf{a}_p^1, ..., \mathbf{a}_p^{n_p}] \in \mathbb{R}^{m \times n_p}$ are the sparse feature representations derived by the dictionary $\mathbf{B}$. Once all the gallery and probe images are projected into the DiCA subspace and their sparse features calculated, we then apply nearest neighbor (NN) classifiers for performing recognition on the projected data.

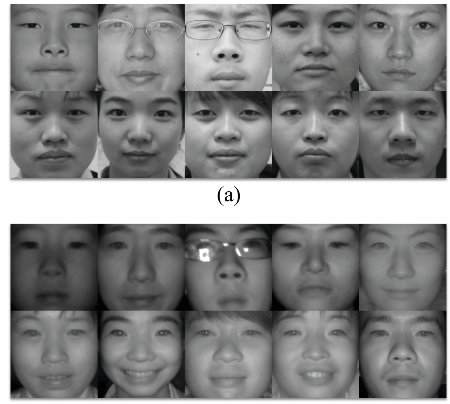
IV. Experiments

A. Sketch-to-Photo Face Recognition

We first consider sketch-to-photo face recognition as the first HFR application. We apply the same setting as [11], [12] and take a subset of CUFS database [7], which has sketch/photo image pairs for 188 persons (see examples in Figure 5). When performing recognition, 100 subjects are randomly selected to be recognized (i.e., $n_g = n_p = 100$), and the photo and sketches images are considered as the training (gallery) and test (probe) sets, respectively. The sketch and photo images of the remaining $n_s = n_t = 88$ subjects are thus viewed as external cross-domain images, and will be utilized for constructing the common feature space.

The face images in CUFS are of size 200×155 pixels, and thus the dimension d is 31000 for each image. We set the threshold $T = 0.05$ in (4) for constructing $\mathbf{W}_{DiCA}$, and we choose to learn the dictionary $\mathbf{B}$ with $m = 150$ basis vectors. We have $\lambda = 0.001$ in (5), (6), and (7). Once the DiCA subspace is observed, we project gallery and probe images into this space, and recognition is achieved by NN using the resulting sparse coefficient features.

To compare our proposed method with other HFR approaches (especially those aiming at determining common feature spaces), we consider Canonical Correlational Analysis (CCA) [26], Bilinear Model (BLM) [27], Partial Least Square (PLS) [11], and the method of Huang *et al.* [12]. For BLM [27], we select 70 PLS bases and 50 eigenvectors as suggested in [11]. In addition, we also consider the direct use of PCA for constructing the common subspace for cross-domain data, followed by our self-taught learning strategy. Table I lists and compares the recognition rates of the above methods. From this table, it is clear that our approach outperformed all other methods, while we do *not* require any data pair information when utilizing external data. We note that, we only consider the use of pixel-based features for deriving the DiCA subspace and the final feature representation in this paper. When more visual features are considerer, further improved recognition results can be expected (e.g., [28]).



(a)

(b)

Fig. 6. Examples (a) VIS and (b) NIR images of ten different subjects from the CASIA 2.0 dataset. It can be seen that the difference between face images of the same subject but from different domains is more significant than that in the CUFS dataset.

B. NIR-to-VIS Face Recognition

The second HFR problem to address is NIR-to-VIS face recognition, and we evaluate the performance of our method using the dataset of CASIA NIR-VIS 2.0 [20]. This dataset contains face images of a total of 725 persons (1-22 VIS face images and 5-50 NIR face images available for per subject). Four different types of facial image variations can be observed (i.e., pose, expression, eyeglasses, and scale), which make the recognition task more challenging. Example images are shown in Figure 6. The cropped face images of this dataset are of size 128×128 pixels, and thus we have $d = 16384$.

When evaluating the recognition performance using the CASIA NIR-VIS 2.0 dataset, we apply the same setting as [20] did. To be more specific, 367 (out of 725) persons are randomly selected as subjects not of interest (i.e., external data), and we randomly choose $n_s = n_t = 1500$ images from each image domain (which are *not* necessarily from the same persons) for constructing our DiCA subspace. As for the remaining $n_g = 725 - 367 = 358$ subjects of interest, we have one VIS gallery image per person, and about $n_p = 6200$ NIR images of these subjects as the test (probe) images. The above experiments are repeated 11 times, and we report the average recognition performance. As for the parameters for our method, we have $T = 0.1$, $m = 600$, and $\lambda = 0.0001$.

In addition to the use of PCA for determining the common feature space for cross-domain data, we consider the recent approach of Heterogeneous Component Analysis (HCA) proposed in [20] for comparisons. We note that, methods like CCA, BLM, PLS, etc. considered in Section IV-A cannot

TABLE II. THE RANK-1 RECOGNITION RATE COMPARISONS OF CASIA 2.0 DATABASE

Methods	Recognition Rate (%)	Standard Deviation (%)
PCA	7.16	0.52
HCA [20]	23.07	1.12
HCA + Sym. [20]	23.70	1.89
PCA+self-taught learning	26.20	1.31
HCA+self-taught learning	29.63	1.89
Ours	**32.58**	1.47

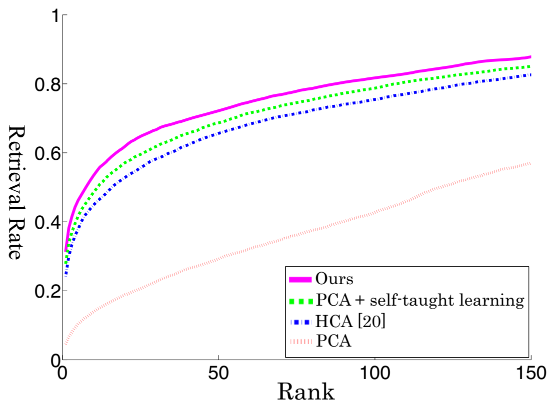


Fig. 7. CMC comparisons for the CASIA NIR-VIS dataset.

be applied, since they require the cross-domain data correspondence when constructing the common feature space. Table II lists and compares the recognition results of several approaches. We note that HCA+Sym was also proposed in [20], which advances facial symmetry properties for boosting the recognition performance. Moreover, we compare the recognition performance using the cumulative match characteristic (CMC) curves, which are shown in Figure 7. From the empirical results and comparisons, although the average performance for this NIR-to-VIS face recognition problem was lower than that reported in Section IV-A due to more challenging settings and scenarios, our method clearly outperformed baseline and state-of-the-art approaches. Therefore, the effectiveness of our proposed method on HFR problems can be successfully verified.

V. CONCLUSION

In this paper, we presented a domain adaptive self-taught learning framework for HFR. We proposed a DiCA algorithm which utilizes external source and target domain data for constructing a subspace, aiming at associating the two image domains. Different from some prior HFR or domain adaptation approaches, we do not require data correspondence nor label information when collecting cross-domain external data during the learning of DiCA subspaces. We further advocated the strategy of self-taught learning in the observed DiCA subspace, and learned a domain-independent dictionary for better representing cross-domain images projected into this subspace. This dictionary learning stage introduces additional image representation capabilities, and alleviates the concern of insufficient amounts of gallery images for recognition. We achieved very promising recognition results on both sketch-to-photo and NIR-to-VIS HFR problems, and we successfully verified that our method outperformed baseline and state-of-the-art HFR approaches.

ACKNOWLEDGEMENT

This work is supported in part by the Ministry of Science and Technology of Taiwan via NSC102-3111-Y-001-015 and NSC102-2221-E-001-005-MY2.

REFERENCES

[1] Q. Yang and X. Wu, "10 challenging problems in data mining research," *Int'l Journal of Information Technology & Decision Making*, vol. 5, no. 04, pp. 597–604, 2006.

[2] X. Wu, V. Kumar, J. R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A. Ng, B. Liu, S. Y. Philip *et al.*, "Top 10 algorithms in data mining," *Knowledge and Information Systems*, 2008.

[3] Z. Lei and S. Z. Li, "Coupled spectral regression for matching heterogeneous faces," in *IEEE CVPR*, 2009.

[4] X. Tang and X. Wang, "Face sketch recognition," *IEEE Trans. CSVT*, 2004.

[5] Q. Liu, X. Tang, H. Jin, H. Lu, and S. Ma, "A nonlinear approach for face sketch synthesis and recognition," in *IEEE CVPR*, 2005.

[6] X. Gao, J. Zhong, J. Li, and C. Tian, "Face sketch synthesis algorithm based on e-hmm and selective ensemble," *IEEE Trans. CSVT*, 2008.

[7] X. Wang and X. Tang, "Face photo-sketch synthesis and recognition," *IEEE Trans. PAMI*, 2009.

[8] W. Zhang, X. Wang, and X. Tang, "Lighting and pose robust face sketch synthesis," in *ECCV*, 2010.

[9] D. Yi, R. Liu, R. Chu, Z. Lei, and S. Z. Li, "Face matching between near infrared and visible light images," in *Advances in Biometrics*, 2007.

[10] A. Li et al., "Maximizing intra-individual correlations for face recognition across pose differences," in *IEEE CVPR*, 2009.

[11] A. Sharma and D. W. Jacobs, "Bypassing synthesis: PLS for face recognition with pose, low-resolution and sketch," in *IEEE CVPR*, 2011.

[12] D.-A. Huang and Y.-C. F. Wang, "Coupled dictionary and feature space learning with applications to cross-domain image synthesis and recognition," in *IEEE ICCV*, 2013.

[13] B. Klare and A. K. Jain, "Heterogeneous face recognition: Matching nir to visible light images," in *IEEE ICPR*, 2010.

[14] B. Klare and A. Jain, "Heterogeneous face recognition using kernel prototype similarities," *IEEE PAMI*, 2012.

[15] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE TKDE*, 2010.

[16] M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu, "Transfer feature learning with joint distribution adaptation," in *IEEE ICCV*, 2013.

[17] S. Shekhar, V. M. Patel, H. V. Nguyen, and R. Chellappa, "Generalized domain-adaptive dictionaries," in *IEEE CVPR*, 2013.

[18] R. Raina, A. Battle, H. Lee, B. Packer, and A. Y. Ng, "Self-taught learning: Transfer learning from unlabeled data," in *ICML*, 2007.

[19] A. Quattoni et al., "Transfer learning for image classification with sparse prototype representations," in *IEEE CVPR*, 2008.

[20] S. Li et al., "The CASIA NIR-VIS 2.0 Face Database," in *PBVS*, 2013.

[21] A. Gretton, K. Borgwardt, M. Rasch, B. Schlkopf, and A. Smola, "A kernel method for the two sample problem," in *NIPS*, 2007.

[22] M. Yang, L. Zhang, X. Feng, and D. Zhang, "Fisher discrimination dictionary learning for sparse representation," in *IEEE ICCV*, 2011.

[23] M. Yang, L. Zhang, J. Yang, and D. Zhang, "Metaface learning for sparse representation based face recognition," in *IEEE ICIP*, 2010.

[24] S. Kong and D. Wang, "A dictionary learning approach for classification: Separating the particularity and the commonality," in *ECCV*, 2012.

[25] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online dictionary learning for sparse coding," in *ICML*, 2009.

[26] H. Hotelling, "Relations between two sets of variates," *Biometrika*, vol. 28, no. 3/4, pp. 321–377, 1936.

[27] J. B. Tenenbaum and W. T. Freeman, "Separating style and content with bilinear models," *Neural computation*, 2000.

[28] D. Gong, Z. Li, J. Liu, and Y. Qiao, "Multi-feature canonical correlation analysis for face photo-sketch image retrieval," in *ACM MM*, 2013.