

Multi-feature Canonical Correlation Analysis for Face Photo-Sketch Image Retrieval

Dihong Gong¹, Zhifeng Li¹, Jianzhuang Liu^{1,2,3}, and Yu Qiao¹

¹Shenzhen Key Lab of Computer Vision and Pattern Recognition
Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China

²Media Lab, Huawei Technologies Co. Ltd., China

³Department of Information Engineering, The Chinese University of Hong Kong, China
{dh.gong, zhifeng.li, yu.qiao}@siat.ac.cn, liu.jianzhuang@huawei.com

ABSTRACT

Automatic face photo-sketch image retrieval has attracted great attention in recent years due to its important applications in real life. The major difficulty in automatic face photo-sketch image retrieval lies in the fact that there exists great discrepancy between the different image modalities (photo and sketch). In order to reduce such discrepancy and improve the performance of automatic face photo-sketch image retrieval, we propose a new framework called multi-feature canonical correlation analysis (MCCA) to effectively address this problem. The MCCA is an extension and improvement of the canonical correlation analysis (CCA) algorithm using multiple features combined with two different random sampling methods in feature space and sample space. In this framework, we first represent each photo or sketch using a patch-based local feature representation scheme, in which histograms of oriented gradients (HOG) and multi-scale local binary pattern (MLBP) serve as the local descriptors. Canonical correlation analysis (CCA) is then performed on a collection of random subspaces to construct an ensemble of classifiers for photo-sketch image retrieval. Extensive experiments on two public-domain face photo-sketch datasets (CUFS and CUFSF) clearly show that the proposed approach obtains a substantial improvement over the state-of-the-art.

Categories and Subject Descriptors

I.4.9 [IMAGE PROCESSING AND COMPUTER VISION]:

Applications; I.5.4 [PATTERN RECOGNITION]: Applications

General Terms

Algorithms

Keywords

Image retrieval, face photo-sketch matching, canonical correlation analysis.

1. INTRODUCTION

Matching face sketch to face photo has been a challenging problem in the face recognition community. It refers to matching a given face sketch drawn by artist to a gallery of photos in the

reference databases. It has important applications in law enforcement. Based on a drawn face sketch, the police can use it to automatically retrieve the face photos in police mug-shot databases to narrow down potential suspect quickly. The major challenge of face photo-sketch image retrieval is the great discrepancy between the photo and the corresponding sketch, which is conventionally referred to as modality gap. As illustrated by Figure 1, in the sketches, much texture information is missing, and the outlines are also distorted. Thus, it is expected that the matching performance cannot be good if we match a face sketch directly to a face photo.

Due to its important applications in the law enforcements, a variety of studies have been carried out on this topic, which can be summarized as the following three categories: 1) Convert images from one modality to the other by synthesizing a pseudo-image from the query image such that the matching process can be done within the same modality [1-3]. 2) Design an appropriate representation that is insensitive to the modalities of images. For example, Klare et al [4] used SIFT feature descriptors and multi-scale local binary patterns to represent both the sketch and photo images; Zhang et al [5] proposed a learning-based algorithm to capture discriminative local face structures and to effectively match photo and sketch. 3) Compare the heterogeneous images on a common subspace where the modality difference is believed to be minimized. Tenenbaum et al [6] applied the Bilinear Model (BLM) by Singular Value Decomposition (SVD) to develop a common content (associated to identity) space for a set of different styles (corresponding to modalities); In [7], Yi et al used the Canonical Correlation Analysis (CCA) technique to construct a common subspace where the correlations between infrared and optical images can be maximized; In [8-9], authors applied the CCA on cross-pose and 2D-3D face recognition; In [10], Sharma et al applied the Partial Least Squares (PLS) method to derive a linear subspace in which cross-modality images are highly correlated.

In this paper, we propose a new framework called multi-feature canonical correlation analysis (MCCA) to reduce the modality gap between the sketch and photo images. The key idea is to learn a series of canonical bases that map the features of both a sketch and a photo into a subspace where the correlation between their features is maximized. To improve the matching performance, we also incorporate two random sampling techniques (bagging [18] and random subspace [17]) with multiple features, histograms of oriented gradients (HOG) [12] and multi-scale local binary pattern (MLBP) [11], into our framework. Extensive experiments on two large photo-sketch databases clearly reveal that our framework can notably improve the matching performance over the state-of-the-art.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

MM '13, October 21 - 25 2013, Barcelona, Spain

Copyright 2013 ACM 978-1-4503-2404-5/13/10 \$15.00.

<http://dx.doi.org/10.1145/2502081.2502162>

The contributions of this paper are summarized as follows: (i) We propose a new framework called MCCA to effectively address the face photo-sketch image retrieval task. (ii) We observe a substantial performance improvement over published algorithms using our new approach on two public domain face photo sketch databases (CUFS [2] and CUFSF [5]).



Figure 1. Examples of photos and corresponding sketches from CUHK Face Sketch FERET Database [5].

2. PROPOSED APPROACH

The proposed framework consists of two major components: a patch-based local feature representation and canonical correlation analysis (CCA) based feature classification. We will describe each component of the framework in the following subsections.

2.1 Local Feature Representation

Local facial features have been shown to be more effective than global facial features in representing face images at diverse scales and orientations [14]. Considering that the entire face image (which has high structural complexity) is difficult to be characterized by a single image descriptor, we use a patch-based local feature representation scheme in this paper. We first divide the whole face images into a set of overlapping patches with size 12x12 pixels (overlapping factor = 0.5) and then apply local image descriptors to each patch. The extracted features from these patches are concatenated together to form a long feature vector for further analysis.

Among the existing local feature descriptors, histograms of oriented gradients (HOG) [12] and multi-scale local binary pattern (MLBP) [11] are among the most successful ones. Hence, we will use both of them as feature description in our study. The HOG feature descriptor quantizes the gradient orientations into 8 directions with 45 degrees for each, and compute a histogram within the local patch such that each bin corresponds to the total gradient magnitude of that orientation. For the MLBP, we compute the MLBP descriptors with radii {1, 3, 5, 7} using u2 coding [11].

In our study, we will fuse the two extracted local features (HOG and MLBP) as feature representation for each face photo or face sketch.

2.2 Multi-feature Canonical Correlation Analysis

2.2.1 Canonical Correlation Analysis (CCA)

In statistics, the Canonical Correlation Analysis (CCA) is an effective technique to maximize the correlation between two random variables. Let us suppose that we have $d_1 \times N$ and

$d_2 \times N$ data matrices $X = \{\vec{x}_1, \vec{x}_2, \dots, \vec{x}_N\}$ and $Y = \{\vec{y}_1, \vec{y}_2, \dots, \vec{y}_N\}$, consisting of N observations with dimension d_1 and d_2 , respectively. Then the CCA seeks one pair of vectors $\vec{a} \in R^{d_1 \times 1}$ and $\vec{b} \in R^{d_2 \times 1}$ such that the correlation between random variables $\vec{a}^T X$ and $\vec{b}^T Y$ can be maximized. Formally,

$$\arg\max_{\vec{a}, \vec{b}} \{\text{corr}(\vec{a}^T X, \vec{b}^T Y)\} \quad (1)$$

Then the $u = \vec{a}^T X$ and $v = \vec{b}^T Y$ are the first pair of canonical variables. One can further seek another pair of vectors \vec{a}' and \vec{b}' maximizing the same correlation, subject to the constraint that they are to be uncorrelated with the first pair of canonical variables, which gives the second pair of canonical variables. This procedure may be continued up to $k = \min\{\text{rank}(X), \text{rank}(Y)\}$ times, yielding k pairs of mutually uncorrelated canonical variables $U = \{u_1, u_2, \dots, u_k\}$ and $V = \{v_1, v_2, \dots, v_k\}$, along with the corresponding projection vector pairs $A = \{\vec{a}_1, \vec{a}_2, \dots, \vec{a}_k\}$ and $B = \{\vec{b}_1, \vec{b}_2, \dots, \vec{b}_k\}$, which we will refer to as the k -correlation matrix.

2.2.2 Learning CCA bases

Let $\Sigma_{XX} = \text{cov}(X, X)$ and $\Sigma_{YY} = \text{cov}(Y, Y)$.

Optimizing (1) is equivalent to maximizing:

$$\rho = \frac{\vec{a}^T \Sigma_{XY} \vec{b}}{\sqrt{\vec{a}^T \Sigma_{XX} \vec{a}} \sqrt{\vec{b}^T \Sigma_{YY} \vec{b}}} \quad (2)$$

According to [8], the optimal correlation defined in (2) can be obtained if \vec{a} is the eigenvector with the maximum eigenvalue of the matrix:

$$\Sigma_{XX}^{-1} \Sigma_{XY} \Sigma_{YY}^{-1} \Sigma_{YX}$$

and \vec{b} is the eigenvector with the maximum eigenvalue of the matrix:

$$\Sigma_{YY}^{-1} \Sigma_{YX} \Sigma_{XX}^{-1} \Sigma_{XY}$$

The subsequent pairs of the correlation vectors can be found by using the eigenvectors corresponding to the eigenvalues of decreasing magnitudes. The uncorrelated constraint of canonical variables is guaranteed by the orthogonality of the eigenvectors.

2.2.3 Multi-classifier Ensemble

CCA has been shown to be very effective in cross-modality matching problem [8]. However, if we directly use it for face photo-sketch image retrieval, we will encounter the following problems. First, the data size will be very large. In our study, there are two kinds of local features (HOG and MLBP) with each one spanning a high-dimensional space feature vector. So the fused HOG and MLBP feature is of high dimension (as high as 115200 in our experiments). Directly applying CCA on such a high-dimensional feature vector is too costly. Second, a more serious problem is the overfitting problem because of the large feature dimension, which would degrade the matching performance. To solve these problems, we apply two popular random sampling methods: random subspace [17] and bagging [18]. In the random subspace method, multiple classifiers are constructed by randomly sampling the feature space. The final matching score is the sum of matching scores from individual sub-classifiers. In the Bagging, multiple training subsets are generated by randomly sampling the training set. A classifier is then constructed from each training subset, and the results of these multiple classifiers are integrated. In order to better address the overfitting problem, we utilize both

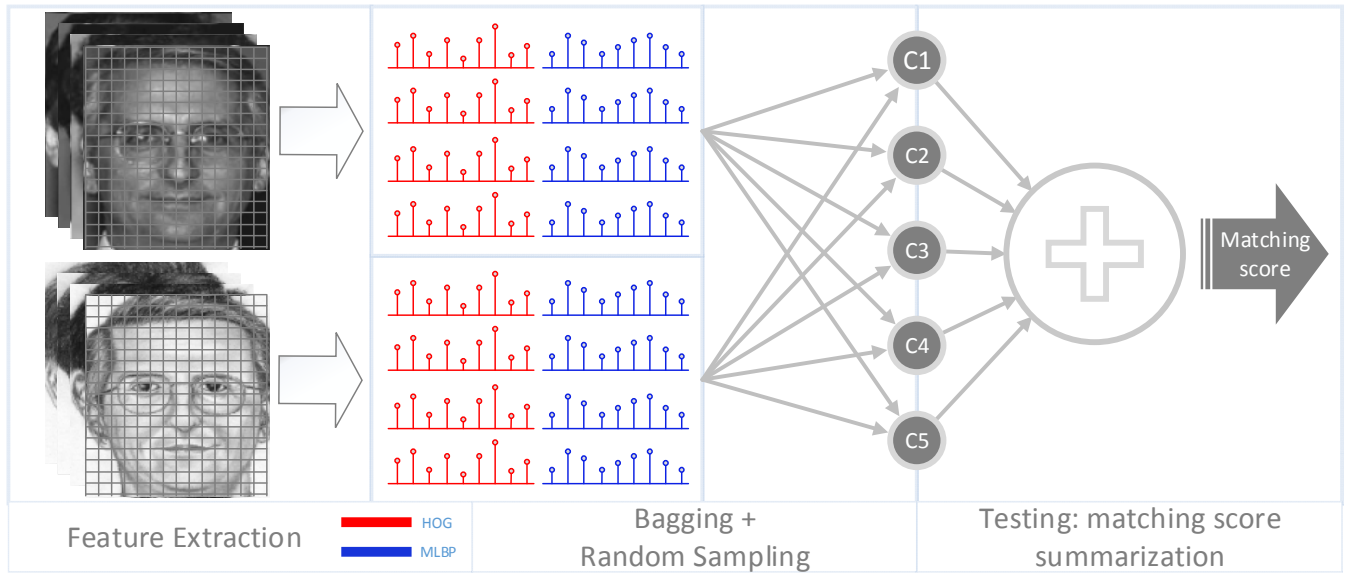


Figure 2. Illustration for the pipeline of the MCCA system. At the training stage, HOG and MLBP features of the face images are extracted and then concatenated to form a long feature. Secondly, for each of the classifier (we only show 5 classifiers in the figure), training samples are selected with replacement randomly and then sampled 5% of the components. Finally, the CCA is applied on each of the random training set to construct a sub-classifier. At the testing stage, features of the probe face are extracted and sampled using the same way as the training procedure, which are then fed into the sub-classifiers to compute the matching scores. The final matching score is the summarization of scores from all the sub-classifiers.

random subspace and bagging schemes. The detailed algorithm is as follows (illustrated in Figure 2).

Suppose N pairs of face training images are represented as $F_{HOG} = \{\vec{f}_H^1, \vec{f}_H^2, \dots, \vec{f}_H^N\}$ and $F_{MLBP} = \{\vec{f}_M^1, \vec{f}_M^2, \dots, \vec{f}_M^N\}$ for HOG and MLBP features respectively. Then the final feature descriptor is obtained by concatenating the two features into one for each face image: $F = \{\vec{f}^1, \vec{f}^2, \dots, \vec{f}^N\}$. The following two random sampling techniques are then applied on the F :

- 1) Randomly sample N pairs of training face images from F with replacement, forming a new training set $F' = \{\vec{f}^{i1}, \vec{f}^{i2}, \dots, \vec{f}^{iN}\}$ (there may be duplicated samples in F' since we are sampling with replacement).
- 2) Randomly sample p components for each of the features in F' , forming the final training set $Ft = \{\vec{f}_{[c1 \dots cp]}^{i1}, \vec{f}_{[c1 \dots cp]}^{i2}, \dots, \vec{f}_{[c1 \dots cp]}^{iN}\}$, where operator $[c1 \dots cp]$ means taking the corresponding components of that vector.

With the above random sampling methods, we can generate one training set at each run of the algorithm. For each of the generated training set, we first apply Principle Component Analysis (PCA) [15] to remove the redundant information, and then the CCA projection matrix is learned with the algorithm described in Section 2.2.2, which serves as a sub-classifier. In our experiments, 100 sub-classifiers are constructed and the matching score of each sub-classifier is the cosine distance between photo and sketch pairs computed in the CCA subspace. The final matching score is computed by summing the scores of all the sub-classifiers.

The MCCA algorithm is an improvement and extension of the CCA algorithm. From the experimental comparison in Section 3, we can see the significant performance improvement of our MCCA algorithm over the CCA algorithm.

3. EXPERIMENTS

In this section, we investigate the performance of our approach for the face photo-sketch image matching. Two public domain sketch-photo databases are used in our study. The first is the CUHK Face-Sketch FERET database (CUFSF) [5] which is the largest publicly available database for sketch-photo recognition. The dataset consists of 1194 pairs of sketch and photo images. In our study we use 300 pairs as training, and the rest 894 pairs as testing. Another is the CUFS database [2], which consists of 606 pairs of sketch and photo images. Following the configuration of training and testing split for CUFS in [2], we use 306 pairs as training and the rest 300 pairs as testing.

3.1 Comparison with subspace analysis methods in cross-modality face matching

Since the proposed MCCA approach is inherently a subspace analysis method, in this part we compare it against (i) PCA+LDA [16] (a popular subspace analysis method for face matching) and (ii) several newly developed subspace analysis methods for cross-modality face matching including PLS [10], LFDA [4], and CCA [8]. Here the CCA algorithm refers to applying CCA directly to the fused HOG and MLBP features, which is a regular practice in the literature. In this experiment, all the algorithms are using the same fused HOG and MLBP features for fair comparison. Also, all the algorithms are tuned to the best configurations according to their papers. The comparative results are illustrated in Figure 3. We illustrate the results of our approach for both the Bagging version (using both the random subspace and bagging techniques) and the one without bagging (using only the random subspace technique). It is very clear that the both of these two MCCA versions notably outperform the CCA algorithm, and the one with Bagging ensemble slightly outperforms the one without it. As discussed earlier, directly applying CCA to the fused HOG and MLBP feature may degrade the performance. The comparative results in Figure 3 clearly confirm this point. It is desirable to

propose the MCCA approach to improve the photo-sketch matching performance. Moreover, it is very encouraging to see that the proposed MCCA approach obtains a significant performance improvement over the existing subspace analysis methods for face photo-sketch matching. This shows the effectiveness of the MCCA approach.

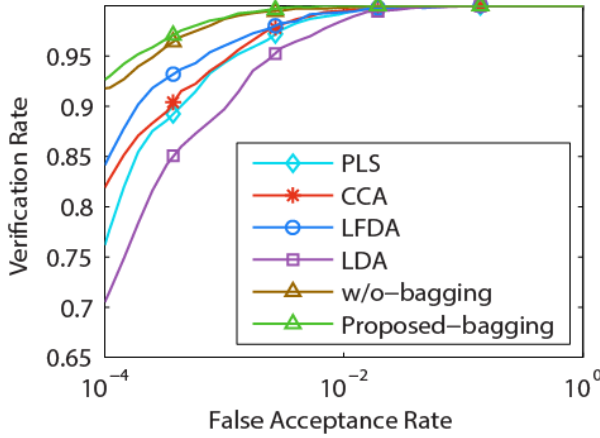


Figure 3. Comparison of verification accuracy on the CUFSS dataset [5].

3.2 Overall Benchmark Comparison

In addition to subspace analysis based methods, there are some other photo-sketch matching methods on the CUFS and CUFSS datasets. In this part, we compare our approach (with Bagging) to the state-of-the-art methods for photo-sketch matching on the CUFS and CUFSS datasets. The comparative results are reported in Table 1 and Table 2. From these results we can see that our approach also has superior performance over the state-of-the-art algorithms on the two datasets.

Table 1. Comparison of rank-1 identification accuracy on the CUFSS dataset [5].

| CCA [8] | PLS [10] | LFDA [4] | CITE [5] | Ours |
|---------|----------|----------|----------|---------------|
| 86.71% | 85.55% | 87.29% | 89.54% | 92.17% |

Table 2. Comparison of rank-1 identification accuracy on the CUFS dataset [2].

| MRF+RS-LDA [2] | LFDA [4] | CITE [5] | GS [19] | Ours |
|----------------|----------|----------|---------|-------------|
| 96.30% | 99.47% | 99.87% | 99.91% | 100% |

4. CONCLUSIONS

In this paper, we have proposed a new algorithm called MCCA for matching face sketch to face photo. Extensive experiments are conducted on two public domain face photo-sketch datasets to demonstrate the effectiveness of MCCA over the state-of-the-art. In the future work, we will adapt our system to general heterogeneous face matching problems, such as infrared-optical recognition or matching faces of different poses.

5. ACKNOWLEDGMENTS

This work was supported by grants from Natural Science Foundation of China (61103164, 61002042, 61070148), Science,

Industry, Trade, and Information Technology Commission of Shenzhen Municipality, China (JC201005270378A), and Guangdong Innovative Research Team Program (No. 201001D0104648280).

6. REFERENCES

- [1] X. Tang and X. Wang, "Face Sketch Recognition", IEEE Transactions on Circuits Systems for Video Technology, 14(1), 50-57, 2004.
- [2] X. Wang and X. Tang, "Face photo-sketch synthesis and recognition", IEEE Transactions on Pattern Analysis Machine Intelligence, 31(11), 1955-1967, 2009.
- [3] Q. Liu, X. Tang, H. Jin, H. Lu, S. Ma, "Nonlinear Approach for Face Sketch Synthesis and Recognition", Proceedings of CVPR, 1005-1010, 2005.
- [4] B. Klare, Z. Li and A. K. Jain, "Matching forensic sketches to mugshot photos", IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010.
- [5] W. Zhang, X. Wang, and X. Tang, "Coupled Information-Theoretic Encoding for Face Photo-Sketch Recognition," Proceedings of CVPR, 2011.
- [6] J. B. Tenenbaum, W. T. Freeman, "Separating style and content with bilinear models", Neural Computation 12 (6), 1247-1283, 2000.
- [7] Dong Yi, Rong Liu, RuFeng Chu, Zhen Lei, and Stan Z. Li, "Face Matching Between Near Infrared and Visible Light Images," Advances in Biometrics: International Conference, pp. 523-530, 2007.
- [8] Annan Li, Shiguang Shan, Xilin Chen1 and Wen Gao, "Maximizing Intra-individual Correlations for Face Recognition Across Pose Differences", Proceedings of CVPR, pp.605-611, 2009.
- [9] W. Yang, D. Yi, Z. Lei, J. Sang, and S. Z. Li, "2D-3D Face Matching Using CCA". Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition, 2008.
- [10] Abhishek Sharma, David W Jacobs, "Bypassing Synthesis: PLS for Face Recognition with Pose, Low-Resolution and Sketch," Proceedings of CVPR, pp. 593 - 600, 2011.
- [11] T. Mäenpää, M. Pietikäinen, "Multi-Scale Binary Patterns for Texture Analysis". Proceedings of the SCIA, Gothenberg, Sweden, pp. 885-892, 2003.
- [12] Dalal, N.; Triggs, B, "Histograms of Oriented Gradients for Human Detection", Proceedings of CVPR, vol.1, pp.886-893, 2005.
- [13] Johnson, Richard A.; Wichern, Dean W. (2007). "Applied Multivariate Statistical Analysis (Sixth ed.)". Prentice Hall. ISBN 0-13-187715-1, ISBN 978-0-13-187715-3.
- [14] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 27, no. 10, pp. 1615-1630, Oct. 2005.
- [15] M. Turk and A. Pentland, "Face recognition using eigenfaces," IEEE International Conference Computer Vision and Pattern Recognition, pp. 586-591, 1991.
- [16] P. Belhumeur, J. Hespanha, and D. Kiregeman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," IEEE Trans. on PAMI, Vol. 19, No. 7, pp. 711-720, July 1997.
- [17] T. K. Ho, "The random subspace method for constructing decision forests," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, no. 8, pp. 832-844, 1998.
- [18] L. Breiman, "Bagging predictors," Machine Learning, vol. 24, no. 2, pp. 123-140, 1996.
- [19] H. Galoogahi and T. Sim, "Face Photo Retrieval by Sketch Example," ACM MM 2012.