Face Photo-Sketch Recognition Based on Joint Dictionary Learning

Jixuan Liu[2] Seho Bae[2] Hanjae Park[2] Lei Li[2] Seongbeak Yoon[2] Juneho Yi[1,2]
Signal Capturing and Processing Lab[1]
North University of China
Taiyuan, China
+86-351-3922007
jhyi@skku.edu

Computer Vision Lab[2]
Sungkyunkwan University
Suwon, Korea
+82-31-290-7142
{liujixuan, pr0411, kanje0602, lilei, beagii}@skku.edu

Abstract

Face recognition technology is widely used in law enforcement agencies. Face photo-sketch recognition is one of possible ways to identify suspects. We propose a method using joint dictionary learning for face photo-sketch recognition. Our method bypasses the image synthesis procedure used by previous joint dictionary learning based methods. Compared with other methods such as coupled dictionary learning which projects features from two different modalities into a common space for recognition, our method does not need extra projections, and avoids the expensive optimization of coupled dictionary learning. By using the cosine distance nearest neighbor classifier, our method performs equally well as coupled dictionary learning based method with much less computation. In the experiments on a popular face photo-sketch database, our method achieves recognition rates higher than or comparable to that of the state-of-art methods.

1. Introduction

Face recognition technology has achieved remarkable success in both theory and applications. Yet in most applications, probe and gallery images are from the same modality. Usually, probe and gallery images are both visual images. However, recognition of faces between different modalities is essential in certain circumstances for law enforcement agencies. Matching face images from two different modalities are called heterogeneous face recognition. The most common cases of heterogeneous face recognition are matching faces between near infrared surveillance camera images and mug shots and between photos and sketches. This work describes our research on face photo-sketch recognition that matches a face sketch with its corresponding visual photo i. e., mug shot in police databases. Refer to Figure 1.

Methods solving photo-sketch recognition can be divided into two categories. Methods [1-4] in the first category generate synthesized images from one modality to the other. Then, with probe and gallery images in the same modality, standard face recognition algorithms are deployed. Methods [5, 10] in the second category project features of images in two different modalities into a common space where nearest neighbor classifiers are employed for recognition.

Most photo-sketch recognition approaches belong to the first category. Markov random fields (MRF) model is used to generate synthesized images i. e., synthesized sketch or synthesized photo [1, 2]. [3] uses patch-wise local linear embedding (LLE) to get an initial guess of the synthesized images and then apply multi-dictionary based sparse representation to generate the high frequency and detailed information for the synthesized images. [4] employs sparse neighbor selection (SNS) to obtain an initial estimate of the synthesized image, then apply sparse representation based enhancement (SRE) to further improve the quality of the synthesized images. In these methods, their recognition accuracy completely depends on the quality of the synthesized images. Thus, unless the quality of the synthesized images is warranted for the purpose of recognition, it would be better to bypass the synthesis procedure that otherwise only costs additional computation.

Figure 1. Examples of photo and sketches from CUHK Face Sketch database used in this paper. Images in each column belong to same subject.

As typical of the second category, [5] proposed a method called coupled dictionary learning, which performs very well for face photo-sketch recognition. Coupled dictionary learning refers to learning of a pair of joint dictionaries which is coupled trained and a pair of projection matrices for the coefficients from each modality. The projection matrices project the coefficients from two different modalities into a common space. Recognition can be carried out in this common space by nearest neighbor classifiers. However, the highly non-convex nature of the joint optimization is still a concern. The simplified optimization steps are not appealing to produce an accurate solution.

Though joint dictionary learning related methods [3, 4, 11] perform very well for synthesizing images, its direct application to recognition without image synthesis is not satisfactory because they are all patch-based, thus the dimension of the coefficients concatenated is too

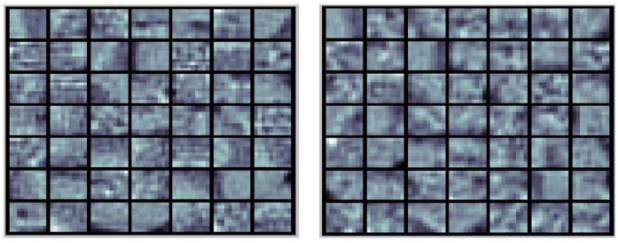

Figure 2. Dictionaries for photos and sketches. The dictionaries are trained jointly. Similarity and difference can be seen between the atoms in corresponding positions.

high for recognition. Experimental results show that direct application of joint dictionary learning to face photo-sketch recognition has only lower performance than coupled dictionary learning. On the other hand, [5, 10] use holistic images rather than image patches, and achieve recognition performance higher than or comparable to that of the patch-wise approaches.

Based on the aforementioned analysis, for the purpose of face photo-sketch recognition, we take the holistic image approach and propose a method using joint dictionary learning without the image synthesis procedure. In addition, we employ cosine distance instead of Euclidean distance to find neighbor relation. Euclidean distance has popularly been used to find the nearest neighbor, but the Euclidean distance may not be a proper measure in a high dimensional space to decide neighboring relation. Experimental results show that our method gives the performance equivalent to that of the coupled dictionary learning based method [5] with much less computational complexity.

The rest of our paper is organized as follows. Section 2 presents the joint dictionary learning framework for face photo-sketch recognition, along with discussions about the assumption for applying joint dictionary learning to face photo-sketch recognition. Section 3 describes the optimization methods of joint dictionary learning and our face photo-sketch recognition algorithm. Section 4 conducts the experiments and reports the experimental results.

2. Joint Dictionary Learning for Face Photo-Sketch Recognition

A joint dictionary learning approach assumes that the same subject from two different modalities share very similar coefficients. It is worth noting that this assumption is necessary for applying joint dictionary learning to cross modality problems. The assumption of joint dictionary learning was first made for image super resolution problems. In [11], for local areas, high and low resolution image patches share the same coefficients with jointly trained dictionaries. This assumption actually turns out to be working very well with other image super resolution works [6, 7, 8, 14].

Researchers further extended joint dictionary learning to heterogeneous face recognition. [9] assumes that near infrared images and visual images share similar sparse coefficients in local areas if the bases of the pair-wise dictionary are jointly trained. To avoid the mistake of blindly applying joint dictionary learning to

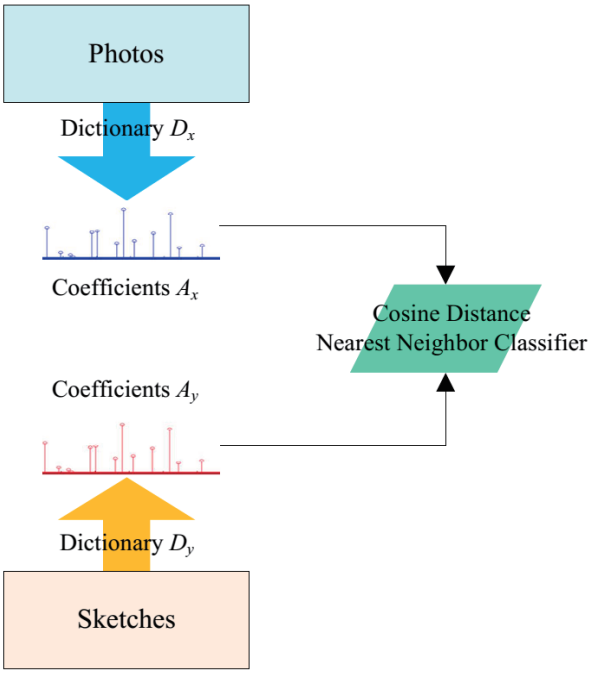

Figure 3. Framework of photo-sketch recognition based on joint dictionary learning. Images are represented by coefficients and dictionary. The two dictionaries are jointly trained. Recognition is done by applying cosine distance nearest neighbor classifier to the coefficients.

face photo-sketch recognition, it is necessary to verify that photos and sketches share similar coefficients for jointly trained dictionary pair.

For a more transparent look of the assumption made about photo and sketch, we carried out a small experiment as follows. We jointly trained two dictionaries with 88 pairs of face photos and sketches from CUHK Face Sketch database. From Figure 2, we can see the dictionaries are considerably coupled. The atoms in the corresponding positions have certain kind of similarity and difference. The difference between each pair of dictionary patches can be viewed as the style changes between photo and sketch. The similarity indicates that they could share the same coefficients by selecting the atoms in the dictionary with respect to the style.

[10] suggests that sketch captures the edges, or high frequency components of an image. The output of a filter such as Laplacian of Gaussian produces is similar to a sketch. Thus, certain transformations and correspondence exist between a photo and its corresponding sketch. So, we assume that in photo-sketch recognition, with a pair of jointly trained dictionaries, the coefficients from two modalities are almost the same.

Similar to dictionary learning, the joint dictionary can be addressed as a minimization problem. The cost function is written as follows.

$$\min_{D_x, D_y, A} \|X - D_x A\|_2^2 + \|Y - D_y A\|_2^2 + \lambda \|A\|_1 \quad s.t.\ \|D_x(:,k)\|_2 \le 1,\ \|D_y(:,k)\|_2 \le 1 \tag{1}$$

where X and Y are the input from two modalities. For photo-sketch recognition, X could be face photo images

and Y sketch images. D_x and D_y are dictionaries for modalities X and Y, respectively. Vector A is the coefficients shared between X and Y. The first term is the reconstruction error term for modality X, and the second term is the reconstruction error term for modality Y. The third term is the sparsity constraint for the coefficients. The balance between reconstruction errors and sparsity constraint is adjusted by λ.

By concatenating the reconstruction error terms and with new notations $\bar{X}$ and $\bar{D}$ as in (2), we can rewrite the cost function (1) as (3).

$$\bar{X} = \begin{bmatrix} X \\ Y \end{bmatrix}, \bar{D} = \begin{bmatrix} D_x \\ D_y \end{bmatrix} \tag{2}$$

$$\begin{aligned} &\min_{\bar{D},A} \left\| \bar{X} - \bar{D}A \right\|_2^2 + \lambda \left\| A \right\|_1 \\ &s.\,t.\ \left\| \bar{D}(:,k) \right\|_2 \le 1 \end{aligned} \tag{3}$$

For a joint dictionary learning framework, the dictionary pair, D_x and D_y, should only be optimal in the concatenated feature space of X and Y, but not in each modality individually.

As shown in Figure 3, in the training stage, training face photo-sketch pairs are used as the input to learn a jointly trained dictionary pair. In the testing stage, both photos and sketches are represented by dictionaries and coefficients. By finding the nearest cosine distance neighbor of the probe image's coefficient in the gallery, we can identify the subject.

3. The Optimization of Joint Dictionary Learning

The cost function (3) is not jointly convex for both dictionary $\bar{D}$ and coefficients A. We can optimize the cost function alternatively, because (3) is convex in each of $\bar{D}$ and A while the other is fixed.

For updating dictionary $\bar{D}$, the cost function can be rewritten as (4). Notice only $\bar{D}$ is variable, and A is fixed, so the sparsity constraint term is a constant and is no longer in the cost function. With the assumption of joint dictionary learning, we force the two cost functions share the same coefficients. The optimization of (4) is a Quadratically Constrainted Quadratic Programming which can be solved by the Lagrange duel techniques [13].

$$\begin{aligned} &\min_{\bar{D}} \left\| \bar{X} - \bar{D}A \right\|_2^2 \\ &s.\,t.\ \left\| d_{x,k} \right\|_2 \le 1 \end{aligned} \tag{4}$$

Updating the coefficients A is similar to updating $\bar{D}$. Notice the sparsity constraint for coefficients still exist. Thus the cost function can be rewritten as follows.

$$\min_{A} \left\| \bar{X} - \bar{D}A \right\|_2^2 + \lambda \left\| A \right\|_1 \tag{5}$$

The optimization of (5) can be solved by an optimization toolbox such as SPAMS [12].

Algorithm 1 Face Photo-Sketch Recognition

Input: Paired training data X_1 (photo) and Y_1 (sketch), un-paired testing data X_2 (photo) and Y_2 (sketch).

Part 1: Training of dictionaries

while not converged **do**

1. Initialize $\bar{D}^0$ and A^0 by [11].
2. Update $\bar{D}^{k+1}$ by (4) with other variables obtained from the previous iteration.
3. Update A^{k+1} by (5) with other variables obtained from the previous iteration.

end while

Part 2: Testing the recognition ability

1. Compute coefficients A_x and A_y for photo and sketch, respectively, using the following equations.

$$\min_{A_x} \left\| X - D_x A_x \right\|_2^2 + \lambda \left\| A_x \right\|_1$$

$$\min_{A_y} \left\| X - D_y A_y \right\|_2^2 + \lambda \left\| A_y \right\|_1$$

2. Compare the cosine distance between each A_x and each A_y, the nearest cosine distance neighbor from the other modality is the corresponding image (a match).
3. When there is a match, record the image name, verify if it is a correct match. The number of correct matches divided by the total number of probe images is the recognition rate.

Output: Dictionary D_x, D_y, and recognition rate.

4. Experiments

We examine the performance of joint dictionary learning for face photo-sketch recognition. The training and testing images are from CUHK Face Sketch database [2], which contains 188 face photo-sketch pairs of students. We take 88 photo-sketch pairs for training and the rest 100 photo-sketch pairs for testing. No testing sample is in the training samples. We have set the regularization parameters λ to 0.01. We compare the recognition performance of our method with that of the previous methods that include canonical correlation analysis (CCA) [15], partial least squares (PLS) [10], and coupled dictionary learning (CDL) [5] which is a method very related to our method.

To make sure that our method is not data driven, all training and testing samples are randomly chosen. We repeat the experiment five times with randomly chosen training and testing samples, and report the average of the recognition rates as the recognition performance. In order to have a better comparison with CDL [5], we set the number of dictionary atoms to 50. As shown in Table 1, our method shows the recognition performance better than CCA and PLS and comparable to CDL. The recognition performance of CDL is almost the same as that of our method while CDL employs additional projection of the coefficients.

Table 2 shows the effectiveness of cosine distance measure. When Euclidean distance is used for CDL and our method, they both show poor performance. With the use of cosine distance, the recognition rate for both methods boosts. This experimentally proves that cosine distance is a better measure than Euclidean distance in

high dimensional spaces.

Table 1. Comparison of recognition performance between our method and the other methods.

Method	CCA	PLS	CDL	Ours
Result	94.6	93.6	97.4	97.2

Table 2. Comparison of recognition performance between CDL and our method with respect to different distance measures.

Method	CDL		Ours	
Distance	Euclidean	Cosine	Euclidean	Cosine
Result	75.3	97.4	58.9	97.2

Table 3. Performance comparison of joint dictionary learning based methods based on image patches and holistic images, both using cosine distance nearest neighbor classifier.

Method	Patch-based	Holistic Image
Result	95.4	97.2

Furthermore, to demonstrate our method is better than the previous patch-based joint dictionary learning approaches which form feature vectors by concatenating coefficients of image patches, we give their performance comparison in Table 3.

In summary, our method avoids expensive optimization of CDL [5], and still has a recognition rate as high as CDL.

5. Conclusion

In this paper, we proposed a joint dictionary learning method for face photo-sketch recognition. In the training stage, dictionaries for photos and sketches are trained jointly with face photo and sketch image pairs. In the testing stage, images are represented by coefficients using learnt dictionaries. Both training and testing image samples are holistic images instead of image patches. By comparing the cosine distances between the coefficients of a probe image and that of the gallery images, we can find the gallery image that matches with the probe.

Compared with the previous joint dictionary learning based methods, our method not only outperforms them, but also achieves computational efficiency by bypassing the image synthesis procedure. Our method also performs equally well as the coupled dictionary learning method with much less computational complexity.

6. Acknowledgement

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2013R1A1A2006164).

References

[1] Yang, L., *et al.*: "Face sketch-to-photo synthesis from simple line drawing", Signal & Information Processing Association Annual Summit and Conference (APSIPA ASC), Asia-Pacific, 2012.

[2] Xiaogang, W. and T. Xiaoou: "Face Photo-Sketch Synthesis and Recognition", *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI),* 31(11): p. 1955-1967, 2009.

[3] Nannan, W., *et al.*: "Face Sketch-Photo Synthesis under Multi-dictionary Sparse Representation Framework", Sixth International Conference on Image and Graphics (ICIG), 2011.

[4] Xinbo, G., *et al.*: "Face Sketch-Photo Synthesis and Retrieval Using Sparse Representation", *IEEE Transactions on Circuits and Systems for Video Technology,* 22(8): p. 1213-1226, 2012.

[5] De-An, H. and Y.C.F. Wang: "Coupled Dictionary and Feature Space Learning with Applications to Cross-Domain Image Synthesis and Recognition", IEEE International Conference on Computer Vision (ICCV), 2013.

[6] Li, H., *et al.*: "Beta Process Joint Dictionary Learning for Coupled Feature Spaces with Application to Single Image Super-Resolution", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013.

[7] Di, Z. and D. Minghui: "Super-resolution image reconstruction via adaptive sparse representation and joint dictionary training", 6th International Congress on Image and Signal Processing (CISP), 2013.

[8] Lei, Z., *et al.*: "An improved joint dictionary training method for single image super resolution", International Conference on Computational Problem-Solving, 2012.

[9] Zeda, Z., *et al.*: "Face synthesis from near-infrared to visual light via sparse representation", International Joint Conference on Biometrics (IJCB), 2011.

[10] Sharma, A. and D.W. Jacobs: "Bypassing synthesis: PLS for face recognition with pose, low-resolution and sketch", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011.

[11] Jianchao, Y., *et al.*: "Image Super-Resolution Via Sparse Representation", *IEEE Transactions on Image Processing,* 19(11): p. 2861-2873, 2010.

[12] SPAMS (SPArse Modeling Software) optimization toolbox: http://spams-devel.gforge.inria.fr/

[13] H. Lee, *et al.*: "Efficient sparse coding algorithms", Neural Information Processing Systems Conference, 2006.

[14] Jianchao, Y., *et al.*: "Face hallucination VIA sparse coding", 15th IEEE International Conference on Image Processing (ICIP), p. 1264-1267, 2008.

[15] H.Hotelling: "Relation between two sets of variates", *Biometrika,* 28: p. 321-377, 1936.