FACE SKETCH RECOGNITION BY LOCAL RADON BINARY PATTERN: LRBP

Hamed Kiani Galoogahi and Terence Sim

School of Computing, National University of Singapore, Singapore

ABSTRACT

In this paper, we propose a new face descriptor to directly match face photos and sketches of different modalities, called Local Radon Binary Pattern (LRBP). LRBP is inspired by the fact that the shape of a face photo and its corresponding sketch is similar, even when the sketch is exaggerated by an artist. Therefore, the shape of face can be exploited to compute features which are robust against modality differences between face photo and sketch. In LRBP framework, the characteristics of face shape are captured by transforming face image into Radon space. Then, micro-information of face shape in new space is encoded by Local Binary Pattern (LBP). Finally, LRBP is computed by concatenating histograms of local LBPs. In order to capture both local and global characteristics of face shape, LRBP is extracted in a spatial pyramid fashion. Experiments on CUFS and CUFSF datasets indicate the efficiency of LRBP for face sketch recognition.

Index Terms— Face sketch recognition, Radon Transform, Local Binary Pattern

1. INTRODUCTION

Face sketch recognition has received significant attention due to its vital role in law enforcement. In many cases, the only available information is the recollection of suspect's face provided by an eye-witness; which can be used by forensic artist to draw a face sketch. This sketch is very useful to automatically recognize or narrow down face photos of potential suspects from police mug-shot databases. The major challenge of face sketch recognition is matching images of different modalities which is referred as modality gap [1]. A face photo is captured by a digital camera, while a face sketch is drawn by an artist. Even for the same human subject, the face shape and texture might be distorted or lost in its corresponding face sketch.

To date, various approaches have been proposed to address the problem of face sketch recognition; which can be categorized into two classes: intra-modality and inter-modality approaches. Intra-modality approaches focus to eliminate or reduce the modality gap by transforming face photos and sketches into a same modality using photo/sketch synthesis. The proposed method of Tang and Wang [2,3] synthesized a pseudo sketch by applying Eigen-transformation on entire face photo which used for matching in sketch modality. Similarly, Liu et al. [4] improved the synthesis framework of [2] by applying Eigen-transformation on local patches instead of whole image. Later, Wang and Tang [5] improved the original idea of Liu et al. [4] by modeling spatial relation of local patches using multi scale Markov Random Field (MRF). However, the performance of all these methods is highly dependent on the results of photo/sketch synthesis. That means imperfect synthesis might lead to poor sketch recognition.

On the other hand, approaches in inter-modality class directly match face photos and sketches of different modalities. The first approach in this class proposed by Klare and Jain [6] by representing face photo and sketch using SIFT descriptors [7] integrated with a simple 1-NN classifier. Later, Klare et al. [8] proposed local feature based discriminative analysis (LFDA) to match forensic sketches to mug shot photos. Five different features based on SIFT and multi local binary pattern were fused in LFDA. However, the problem of modality gap was not mainly addressed by the above approaches. Moreover, none of the features used by these approaches was originally designed for face sketch recognition. Recently, Zhang et al. [1] introduced a coupled information-theoretic encoding to reduce the modality gap by maximizing the mutual information between encoded photos and sketches. However, this method suffers from parameter tuning and inefficient complexity. The method proposed by Kiani and Sim [9] achieved promising results on matching face photo and sketch without shape degradation. However, it is not robust in the presence of shape degradation in face sketches.

In this paper, we propose a new face descriptor called Local Radon Binary Pattern (LRBP). LRBP inspired by the fact that unlike face texture, the shape of face sketch and its corresponding photo is similar, even when the shape of sketch is distorted or exaggerated by the artist. Therefore, the local and global characteristics of face shape can be exploited to describe a face in both sketch and photo modalities. Given a face image, the LRBP descriptor is extracted as follows: First, the input face image is divided into non-overlapping local patches. Then, the shape of each local patch is modeled by Radon transform [10,11]. The transformed image in Radon space is encoded by Local Binary Pattern (LBP) operator [12]. Finally, the LRBP is constructed by concatenating spatial histograms of LBP codes computed from all transferred local patches. In order to represent both local and global characteristics of face shape, LRBP is extracted in a spatial pyramid fashion [13].

2. LOCAL RADON BINARY PATTERN

This section provides a brief introduction of Radon transform and LBP, followed by presenting the new Local Radon Binary Pattern (LRBP) framework.

2.1. Radon Transform

Radon transform computes projections of image intensity along tracing lines. Each line is characterized by its distance to the origin s and rotation angle from reference axes θ. Given an image $f:(x,y) \rightarrow g$ and a straight line $l(s,\theta)$, the projection of the image along line l is computed as :

$$\mathcal{R}_f(\theta,s) = \int_l f(x,y)\,dl \tag{1}$$

where all points on the line l satisfy the Equation 2:

$$x\sin(\theta) - y\cos(\theta) - s = 0 \tag{2}$$

therefore, Equation 1 can be rewritten as:

$$\mathcal{R}_f(\theta,s) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)\delta(x\sin\theta - y\cos\theta - s)dxdy \tag{3}$$

where $\delta(.)$ is Dirac delta function. The Radon transform of the given image is determined by a set of projections of the image along lines with different orientations and distances. Radon transform is illustrated in Figure 1 (a).

2.2. Local Binary Pattern (LBP)

The LBP operator was originally introduced by Ojala et al. [12] as a gray-scale invariant descriptor for texture analysis. Later, LBP has been applied for many computer vision applications due to its discriminative power and computational efficiency. Practically, it assigns a binary string to each pixel $I_c = (x_c, y_c)$ of an image by thresholding P sampling pixels on a circle of radius R with the center pixel I_c. The decimal form of the binary string for pixel I_c (LBP code) is calculated as:

$$LBP_R^P(I_c) = \sum_{n=0}^{P-1} s(I_n - I_c)2^n \tag{4}$$

where I_c and I_n are the value of center and neighboring pixels, respectively. The thresholding function $s(x)$ is defined as:

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{5}$$

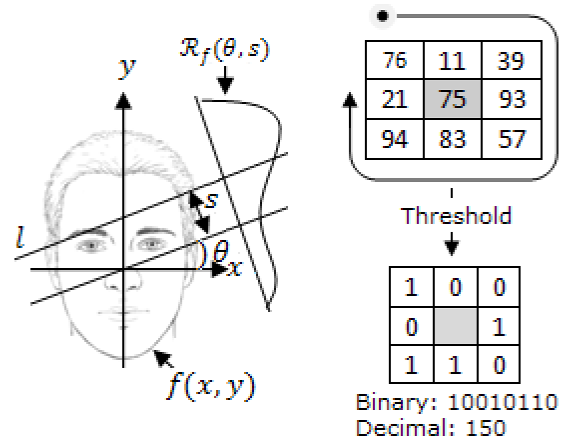
An example of LBP_1^8 is illustrated in Figure 1 (b).

2.3. Local Radon Binary Pattern

The LBP has been originally applied to encode face texture based on pixel's values [14]. Later, some approaches argued that the accuracy of LBP can be improved by using more discriminative features in encoding stage [15,16]. In LRBP, we apply LBP on transformed image in Radon space. By applying the LBP operator on Radon transform, LRBP encodes the shape of face which is invariant to different modalities. Given an image $f:(x,y) \rightarrow g$, the 2D Radon transform $\mathcal{R}_f(\theta,s)$ can be calculated by Equation 3. Then, the local radon binary pattern (LRBP) is computed as:

$$LRBP_R^P\left(I_c^{\mathcal{R}_f}\right) = \sum_{n=0}^{P-1} s(I_n^{\mathcal{R}_f} - I_c^{\mathcal{R}_f})2^n \tag{6}$$

where $I_i^{\mathcal{R}_f}$ denotes Radon transform $\mathcal{R}_f(\theta,s)$ and s(.) is thresholding function defined by Equation 5.



(a) (b)

Fig. 1. (a) Radon transform, (b) LBP operator: LBP_1^8

2.4. Pyramid Representation of LRBP

In order to model both local and global characteristics of face shape, LRBP is calculated in a coarse to fine fashion using spatial pyramid [13]. In spatial pyramid with L levels, an image at level *l=0,..,L-1* is represented by dividing the original image into 2^l spatial grids along each axis direction.

Given a face image *f*, a spatial pyramid with L levels is represented by $SP_f = \{sp_f^l, 0 \leq l \leq L-1\}$. where each sp_f^l contains 2^{2l} spatial grids $sg_f^{l,i}$, $sp_f^l = \{sg_f^{l,i}, 1 \leq i \leq 2^{2l}\}$. For each spatial grid $sg_f^{l,i}$, $LRBP_R^P(sg_f^{l,i})$ is computed using Equation 6 on its Radon transform followed by computing a histogram $h_{LRBP_R^P(sg_f^{l,i})}$ as:

$$h_{LRBP_R^P(sg_f^{l,i}),j} = \sum_{\mathbb{R}^2} I\{LRBP_R^P(sg_f^{l,i}) = j\} \tag{7}$$

where $0 \leq j < 2^P - 1$ and

$$I(A) = \begin{cases} 1, & A\ is\ true \\ 0, & A\ is\ false \end{cases} \tag{8}$$

then, all histograms computed from spatial grids $sg_f^{l,i}, 1 \leq i \leq 2^{2l}$ at level l are concatenated to a histogram sequence $H_{f,l}$:

$$H_{f,l} = \{h_{LRBP_R^P(sg_f^{l,i})}, 1 \leq i \leq 2^{2l}, 0 \leq l \leq L-1\} \tag{9}$$

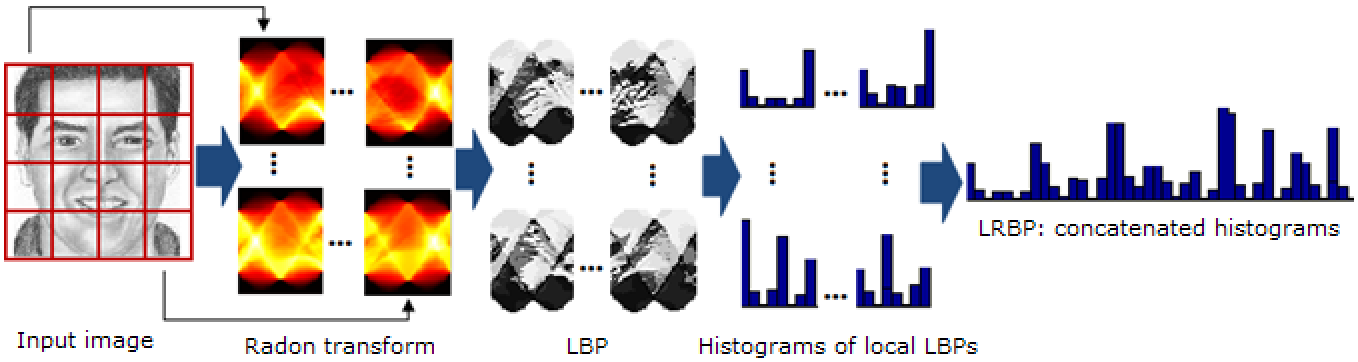


Fig. 2. Framework of LRBP at level 2

Finally, the pyramid representation of LRBP is represented by $H_{f,L} = \{H_{f,l}, 0 \leq l \leq L-1\}$. The overall framework of LRBP of level 2 is illustrated in Figure 2.

A distance measure based on Pyramid Match Kernel (PMK) [13] is used to compare two histograms (LRBPs) $H_{p,L}$ and $H_{s,L}$ as:

$$d(H_{p,L}, H_{s,L}) = \frac{\chi^2(H_{p,0}, H_{s,0})}{2^L} + \sum_{l=1}^{L} \frac{\chi^2(H_{p,l}, H_{s,l})}{2^{L-l+1}} \quad (10)$$

This distance assigns higher weights to matching histograms at finer levels. $\chi^2(H_{p,l}, H_{s,l})$ calculates the *chi-square* distance between two histograms $H_{p,l}$ and $H_{s,l}$. $\chi^2(x,y)$ is defined as:

$$\chi^2(x,y) = \sum_k \frac{(x_k - y_k)^2}{(x_k + y_k)} \quad (11)$$

Let $\mathrm{I} = \{\iota_j, 1 \leq j \leq M\}$ denotes a set of face photos with known identities, $\iota_j: \mathbb{R}^2 \to \mathbb{R}$. Given a query face sketch $\varsigma: \mathbb{R}^2 \to \mathbb{R}$, face sketch recognition is performed by:

$$\hat{\iota}_j \leftarrow \underset{\mathrm{I}=\{\iota_j, 1 \leq j \leq M\}}{\operatorname{argmin}} d(H_{I,L}, H_{\varsigma,L}) \quad (12)$$

where L indicates the level of spatial pyramid.

3. EXPERIMENTAL RESULTS

Two experiments are conducted to evaluate the effectiveness of LRBP for face photo-sketch matching. The CUFSF [1] and CUFS [5] datasets are used for the experiments. All 606 photos in CUFS are taken under a normal illumination condition. For each photo, there is a sketch without shape exaggeration drawn by an artist when viewing its corresponding photo. The CUFSF consists of 1194 photo/sketch pairs with lighting variations (for photos) and shape exaggeration (for sketches). All face images are translated, rotated and scaled such that the centers of eyes and mouth are at fixed positions. All images are then cropped to a same size of 224×192 pixels. Some examples of face photo/sketch pairs are shown in Figure 3.

In experiment 1, we compare the new descriptor with original LBP [14], LGBPHS [15] and Sobel LBP [16] descriptors. For each of these four descriptors, there are four free parameters which have to be tuned, including size of local patches *(S×S)*, LBP parameters *(P,R)* and histogram bins for LBP quantizing (*B*) which are represented by 4-tuple *(S,P,R,B)*. The parameters tuning is performed by exploring parameters space *S={8×8, 16×16, 32×32}*, *P={4, 8,16}*, *R={1,2}* and *B={8, 16, 32}* to find the best combination of parameters with highest performance. Eventually, we choose *(8,8,2,8)* for each of three existing approaches and *(8,8,2,32)* for LRBP with a spatial pyramid of five levels, *L=5*. When tuning the parameters, we realized that all four descriptors in this experiment are robust against *P*, *R* and *B* parameters. In addition, the accuracy of these descriptors except LRBP changes by the size of local patches (*S*). Lower accuracy is obtained by increasing the size of local patches.

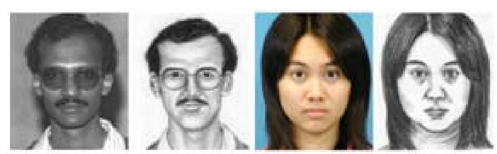


Fig. 3. Face photo and sketch examples of (left) CUFSF, (right) CUFS

The results of experiment 1 are illustrated in Table 1. As shown in Table 1, our descriptor significantly outperforms the other descriptors for both databases. Moreover, the original LBP obtains the lowest accuracy of recognition due to the texture information which is lost in face sketches (modality gap). Increasing the recognition rate of the order of original LBP, Sobel LBP, LGBPHS and LRBP indicates that the discrimination power of LBP based descriptors is improved by extracting features which are less dependent on image texture. Furthermore, due to the large shape degradation of CUFSF compared with CUFS, the accuracy obtained by all descriptors on the former is obviously much less than those on the latter dataset.

The experiment 2 compares LRBP descriptor with three state-of-the-art approaches on CUFS and CUFSF datasets:

MRF+RS-LDA [5]: First, a MRF-based photo synthesis is trained to synthesize pseudo photos from input sketches [5]. Then, a face recognition based on RS-LDA [17] is performed to match pseudo photos against a gallery of face photos with known identities.

LFDA [8]: A discriminate projection is learnt by fusing LBP and SIFT descriptors in a single feature vector. LBP is computed with four radii $r = \{1,3,5,7\}$ and 8 sampling points. SIFT is computed from patches with size 32 [8].

CITP [1]: CITP is characterized by five trees in CITP forest and 256 nodes for each tree. The pattern sampling is performed by a single ring of radius 2. PCA-LDA classifier [18] with dimension of 600 is used for feature reduction.

The above approaches are tuned to the best setting according to their reference papers. The results of experiment 2 are shown in Table 2. According to Table 2, the CITP achieved the highest recognition performance on both databases followed by the LRBP. The poor accuracy obtained by MRF+RS-LDA on CUFSF denotes that MRF-based photo synthesis is affected by the shape distortion and the variations of lighting in training set. Furthermore, the high performance of LRBP and LFDA on CUFSF suggests that fusing features of different spatial partitioning can compensate the influence of shape degradation of sketches.

Table. 1. Recognition rates of Experiment 1

	Descriptors			
	Original LBP [14]	Sobel LBP [16]	LGBPHS [15]	LRBP
CUFS	39.10%	57.26%	76.73%	99.51%
CUFSF	24.03%	45.22%	62.98%	91.12%

Table. 2. Recognition rates of Experiment 2

	Methods			
	MRF+RS-LDA [5]	LFDA [8]	CITP [1]	LRBP
CUFS	96.30%	99.47%	99.87%	99.51%
CUFSF	29.54%	90.78%	98.70%	91.12%

4. CONCLUSIONS

We proposed a new face descriptor, called LRBP, to address the problem of face sketch recognition in presence of modality gap between face photos and sketches. The LRBP descriptor encodes the shape of face image in Radon space using LBP operator. LRBP benefits low computational complexity and also there is no critical parameter to be tuned. Experiments on CUFS and CUFSF datasets demonstrate that the LRBP significantly outperforms previous LBP based face descriptors in face sketch recognition. In addition, LRBP obtains promising results compared with the state-of-the-art approaches in face sketch recognition. For future works, we will focus on face sketch recognition in presence of extreme shape degradation for real-world situations in which the eye-witness cannot properly recall the detail of suspect's face. Moreover, a face sketch dataset drawn based on recollection of eye-witness is a necessity for future researches.

5. AKNOWLEDGMENT

We acknowledge the generous support of NUS grant # R -252 -000-451-112 for this work.

6. REFERENCES

[1] W. Zhang, X. Wang and X. Tang, “Coupled information-theoretic encoding for face photo-sketch recognition”, in *Proc. of CVPR*, pp. 513-520, 2011.

[2] X. Tang and X. Wang, “Face sketch synthesis and recognition”, in *Proc. of ECCV*, pp. 687-694, 2003.

[3] X. Tang and X. Wang, “Face sketch recognition”, *IEEE Trans. Circuits and Systems for Video Technology*, vol. 14, no. 1, pp. 50-57, 2004.

[4] Q. Liu, X. Tang, H. Jin, H. Lu and S. Ma, “A nonlinear approach for face sketch synthesis and recognition”, in *Proc. of CVPR'05*, pp. 1005-1010, 2005.

[5] X. Wang and X. Tang“, Face photo-sketch synthesis and recognition”, *IEEE TPAMI*, vol. 31, no. 11, pp. 1955-1967, 2009.

[6] B. Klare and A. Jain, “Sketch to photo matching: a feature-based approach”, *SPIE Conference on Biometric Technology for Human Identification*, 2010.

[7] D. Lowe, “Distinctive image features from scale-invariant keypoints”, *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91-110, 2004.

[8] B. Klare, Z. Li and A. Jain, “Matching forensic sketches to mug shot photos”, *IEEE TPAMI*, vol. 33, no. 3, pp. 639-646, 2011.

[9] H. Kiani and T. Sim, “Inter-modality face sketch recognition”, in *Proc. of ICME'12*, 2012.

[10] D. V. Jadhav, R. S. Holambe, "Feature extraction using Radon and wavelet transforms with application to face recognition," *Neurocomputing*, vol. 72, no. 7-9, pp. 1951-1959, 2009.

[11] S. Srisuk, M. Petrou, W. Kurutach, and A. Kadyrov, “Face authentication using the trace transform”, in *Proc. of CVPR'03*, pp. 305-312, 2003.

[12] T. Ojala, M. Pietikäinen, and T. Mäenpää, “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns”, *IEEE TPAMI*, vol. 24, pp. 971-987, 2002.

[13] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: spatial pyramid matching for recognizing natural scene categories, in *Proc. of CVPR'06*, pp. 2169-2178, 2006.

[14] T. Ahonen, A. Hadid, and M. Pietikäinen, “Face description with local binary patterns: Application to face recognition”, *IEEE TPAMI*, vol. 28, pp. 2037-2041, 2006.

[15] W. Zhang, S. Shan, W. Gao, X. Chen, and H. Zhang, "Local Gabor Binary Pattern Histogram Sequence (LGBPHS): A novel non-statistical model for face representation and recognition," in *Proc. of ICCV*, pp. 786-791, 2005.

[16] S. Zhao, Y. Gao, B, Zhang, “Sobel LBP”, in *Proc. of ICIP'08*, pp. 2144-2147, 2008.

[17] X. Wang and X. Tang, “Random sampling for subspace face recognition”, *International Journal of Computer Vision*, vol. 70, no. 1, pp. 91-104, 2006.

[18] X. Wang and X. Tang, “A unified framework for subspace face recognition”, *IEEE TPAMI*, vol. 26, no. 9, pp. 1222-1228, 2004.