KERNEL SPARSE REPRESENTATION WITH LOCAL PATTERNS FOR FACE RECOGNITION

Cuicui Kang[1], Shengcai Liao[2], Shiming Xiang[1], Chunhong Pan[1]

[1] National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences
[2] Department of Computer Science and Engineering, Michigan State University

cuicui.kang@ia.ac.cn scliao@msu.edu {smxiang, chpan}@nlpr.ia.ac.cn

ABSTRACT

In this paper we propose a novel kernel sparse representation classification (SRC) framework and utilize the local binary pattern (LBP) descriptor in this framework for robust face recognition. First we develop a kernel coordinate descent (KCD) algorithm for $l1$ minimization in the kernel space, which is based on the covariance update technique. Then we extract LBP descriptors from each image and apply two types of kernels (χ^2 distance based and Hamming distance based) with the proposed KCD algorithm under the SRC framework for face recognition. Experiments on both the Extended Yale B and the PIE face databases show that the proposed method is more robust against noise, occlusion, and illumination variations, even with small number of training samples.

Index Terms— sparse representation, face recognition, kernel, local binary pattern

1. INTRODUCTION

Face recognition is an important research area in computer vision. It has many useful applications in real life, such as face attendance, access control, security surveillance, etc. Face recognition is also a challenging problem, which suffers from aging, occlusion, pose, illumination, and expression variations. Many researchers have been attracted to solve these problems, making a great development in face recognition techniques during the past two decades.

Recently Wright et al. proposed the SRC framework for robust face recognition [1], which makes use of the well-known $l1$-norm constrained least-square reconstruction minimization technique [2]. They showed that SRC obtained impressive results against illumination variations, occlusions, and random noise. In their work image pixel values were used to represent faces. The training set needs to be carefully constructed, i.e. each subject in the training set is constituted with many images representing various lighting conditions, so that a probe image of certain illumination condition can be represented by a sparse linear combination of the training samples. However, in realistic applications it is hard for every enrolled user to have such varying lighting images. In terms of this

drawback, Yuan et al. [3] and Chan et al. [4] proposed to utilize the LBP descriptor in the SRC framework, so that the system would be more robust against illumination variations, making it potentially possible to apply SRC for small-sample-size face recognition problem.

LBP was originally proposed by Ojala et al. for texture classification [5]. It is a binary string resulted from local neighboring pixel comparisons. Ahonen et al. have applied LBP to face recognition and proved that it is robust for illumination variations [6]. In [3], LBP is also employed to handle illumination variations in face recognition, showing that it is more robust than original pixel values under the SRC framework. In their work the LBP encoded images were directly used in the linear sparse representation system. However, since LBP codes are converted from binary comparisons, i.e. they are not regular numerical values, thus it is not very reasonable to linearly combine LBP encodings directly. Actually LBP is mostly used in the form of histogram features counted in local regions, and the χ^2 distance is preferred to calculate distance between two LBP histogram features [5, 6]. Different from [3], Chan et al. [4] proposed to extract LBP histogram features instead for SRC based face recognition.

In this work, we propose a novel kernel SRC framework, and apply it together with LBP descriptor to face recognition. The contribution lies in two-folds. First, we propose a novel kernel coordinate descent (KCD) algorithm with covariance update technique for the $l1$ minimization problem in the kernel space. Second, we further apply it in SRC framework to face recognition, in which we are able to take advantage of the powerful LBP descriptors in two types of kernels: the χ^2 distance based and the Hamming distance based. We conduct several experiments on the Yale-B and the PIE face databases to illustrate the effectiveness of the proposed approach.

2. KERNEL SPARSE REPRESENTATION WITH LOCAL PATTERNS

2.1. Sparse Representation Classification

Suppose $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n] \in \mathbb{R}^{m \times n}$ is a training dictionary with each sample $\mathbf{x}_i$ having zero mean and unit length,

then given a test sample $\mathbf{y} \in \mathbb{R}^m$, the l_1 minimization solves a linear representation of $\mathbf{y}$ in $\mathbf{X}$ with the Lasso constraint [2] as follows

$$\min_{\beta} \frac{1}{2}\|\mathbf{X}\beta - \mathbf{y}\|_2^2 + \lambda\|\beta\|_1, \quad (1)$$

where $\beta \in \mathbb{R}^n$ is a sparse vector. Given the solution from Eq. (1), the SRC algorithm [1] classifies $\mathbf{y}$ based on

$$\min_{c} r_c(\mathbf{y}) \doteq \|\mathbf{X}\delta_c(\beta) - \mathbf{y}\|_2, \quad (2)$$

where $\delta_c(\cdot)$ is the characteristic function [1] that selects coefficients related with the cth class and makes the rest to zero.

2.2. Kernel Sparse Representation

Here we consider the Lasso problem of Eq. (1) in the kernel space, *i.e.*

$$\min_{\beta} J(\beta) \doteq \frac{1}{2}\|\sum_{i=1}^{n}\beta_i\varphi(\mathbf{x}_i) - \varphi(\mathbf{y})\|_2^2 + \lambda\|\beta\|_1, \quad (3)$$

where $\varphi(\cdot)$ is an implicit mapping which maps a feature vector to a kernel space. We assume that $\varphi(\cdot)$ satisfies $\varphi(\mathbf{x})^T\varphi(\mathbf{x}) = 1$ when $\|\mathbf{x}\|_2 = 1$. To solve Eq. (3), in the following we develop a Kernel Coordinate Descent (KCD) algorithm which employs the Coordinate Descent approach [7] due to its simplicity and efficiency. First, taking partial derivative of $J(\beta)$ with respect to $\beta_j (\neq 0)$, we have

$$\frac{\partial J(\beta)}{\partial \beta_j} = \varphi(\mathbf{x}_j)^T[\sum_{i=1}^{n}\beta_i\varphi(\mathbf{x}_i) - \varphi(\mathbf{y})] + \lambda\text{sign}(\beta_j). \quad (4)$$

Then similar with [7], by setting Eq. (4) to zero, we get the update of β_j as

$$\beta_j \leftarrow \text{sign}(\alpha)(|\alpha| - \lambda)_+, \quad (5)$$

where

$$\alpha = \varphi(\mathbf{x}_j)^T[\varphi(\mathbf{y}) - \sum_{i=1,i\neq j}^{n}\beta_i\varphi(\mathbf{x}_i)], \quad (6)$$

and $(s)_+$ equals s if $s > 0$, or 0 otherwise.

Further considering the covariance update idea suggested in [7], Eq. (6) can be rewritten as

$$\alpha = K(\mathbf{x}_j, \mathbf{y}) - \sum_{i=1,i\neq j}^{n}\beta_i K(\mathbf{x}_j, \mathbf{x}_i), \quad (7)$$

where $K(\mathbf{x}, \mathbf{y}) \doteq \varphi(\mathbf{x})^T\varphi(\mathbf{y})$ is the kernel function. Therefore, providing a kernel function K, the KCD algorithm is able to update β iteratively in kernel space by Eq. (5) and (7).

For classification, we also develop the corresponding KCD-SRC criterion as follows

$$\begin{aligned} c &= arg\min_{c} \|\sum_{i=1,l(i)=c}^{n}\beta_i\varphi(\mathbf{x}_i) - \varphi(\mathbf{y})\|_2^2 \\ &= arg\min_{c} \delta_c(\beta)^T\mathbf{R}\delta_c(\beta) - 2\mathbf{z}^T\delta_c(\beta), \end{aligned} \quad (8)$$

where $\mathbf{R} \doteq (K(\mathbf{x}_i, \mathbf{x}_j))_{n\times n}$ is the training kernel matrix, $\mathbf{z} \doteq (K(\mathbf{x}_i, \mathbf{y}))_{n\times 1}$, and $l(i)$ is the class label of the ith sample.

The kernelization of the Lasso problem (Eq. (1)) has also been suggested by Yuan et al. [3] and Gao et al. [8]. In [3] the formulation is addressed for a multi-task learning setting, and the resulted optimization problem is solved via the accelerated proximal gradient method [9]. Differently, our formulation does not need to combine multiple features in face recognition. For problem (3), based on feature-sign search [10], Gao et al. developed a gradient descent algorithm by finding the sparse codes and codebook alternatively [8]. In their work, Gaussian kernel is employed for face recognition. A drawback could be that simply using Gaussian kernel may not achieve high efficiency under small dictionary. In contrast, here we use coordinate descent approach to solve the optimization problem. The computation in each iteration is simple and the convergence can be guaranteed [7]. Furthermore, other than Gaussian kernel, we will develop two types of efficient kernels with LBP features below, resulting in high performance under only a few training samples.

2.3. LBP with Kernels

LBP is a powerful descriptor [5] computed from local neighboring pixel comparisons (see Fig. 1). It is a binary string of length N or a discrete label in $\{0, 1, \cdots, 2^N - 1\}$, where N is the number of neighboring pixels. LBP histogram with these discrete bins is often computed from local image region, and the χ^2 distance is employed for image classification [5, 6]. Inspired by this, we first define our χ^2 kernel as follows

$$K_{\chi^2}(\mathbf{a}, \mathbf{b}) = 2\sum_{i=1}^{L}\frac{a_i b_i}{a_i + b_i}, \quad (9)$$

where $\mathbf{a}$ and $\mathbf{b}$ are two normalized LBP histograms with $L = 2^N$ bins. It can be easily verified that $0 \leq K_{\chi^2} \leq 1$ and $K_{\chi^2}(\mathbf{a}, \mathbf{a}) = 1$. Alternatively, we propose another kernel with hamming distance, which can be applied directly on LBP images. The definition is

$$K_H(\mathbf{x}, \mathbf{y}) = \frac{1}{mN}\sum_{i=1}^{m}h(x_i, y_i), \quad (10)$$

where $\mathbf{x}$ and $\mathbf{y}$ are two LBP encoded images, and $h(\cdot, \cdot)$ is a function that counts the number of consistent bits between two local binary patterns. Note that K_H also ranges in [0,1], and $K_H(\mathbf{x}, \mathbf{x}) = 1$. With these two kernels, we are able to apply the proposed KCD-SRC algorithm with LBP features for face recognition.

3. EXPERIMENTS

To verify the effectiveness of the proposed method, several experiments were conducted on the Extended Yale B [11] and the CMU-PIE [12] face database. We tested the KCD-SRC algorithm with LBP features and two kernels proposed in Section 2.3. For convenience, we denote these two proposed methods by "LBP-K_H" and "LBP-Hist-K_{χ^2}" respectively.

Note that the proposed KCD-SRC algorithm with linear kernel is actually equivalent as the original SRC algorithm with coordinate descent solver for the l_1 minimization problem. Thus we also implemented two other methods in this setting. One is the SRC algorithm with LBP histogram features developed by [4] (denoted by "LBP-Hist"), and the other is SRC with raw pixels proposed in [1] (denoted by "Pixel"). For all experiments, the parameter λ was consistently set to be 0.01 and the maximum iteration number was 100.

For the LBP-K_H algorithm, the $\text{LBP}_{8,1}$ operator was performed on each image for feature extraction. Then the algorithm can be applied directly on the LBP encoded images. In contrast, histograms of the $\text{LBP}_{8,1}^{u2}$ operator [6, 4] was computed for the other two LBP based methods. For this the image was first divided into non-overlapping subregions of 8×8 pixels. Then a 59-bin histogram was counted within each subregion and all sub-histograms were concatenated to form the final representation. The final histogram was normalized to unit length with l_2-norm for LBP-Hist, and l_1-norm for LBP-Hist-K_{χ^2} due to their corresponding requirements.

For each of the two databases we designed experiments in three scenarios: with original images, with noise, and with occlusion. For all experiments we randomly separated training samples and testing samples from the database, and repeated each evaluation for 10 times to obtain mean recognition rates and the corresponding standard deviations.

3.1. On Extended Yale B Database

The Extended Yale B Database [11] consists of 16,128 facial images of 38 subjects under 9 poses and 64 illumination conditions. We selected 2,414 frontal images of all the 38 subjects under 64 illumination conditions for experiments. All faces are cropped to 64×56 pixels. Figure 1 shows some example images from this database, with the corresponding LBP encoded images. As we can see, under illumination variations the LBP features could reserve more local image structures, which benefits face recognition.
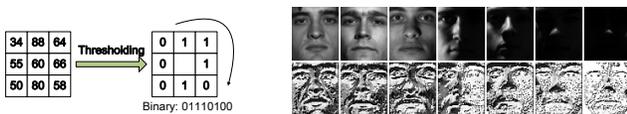

Fig. 1. LBP operator (left) and samples of LBP encodings (right).

With Original Images: The performance of four algorithms with original images from the Extended Yale B database is shown in Figure 2. For each individual we randomly chose 5,10,15,20,25,30 images for training and the rest for testing. It can be seen from Figure 2 that, under different amount of training samples, the proposed LBP-K_H algorithm consistently performs the best, followed by LBP-Hist-K_{χ^2}, LBP-Hist, and Pixel. With the decreasing amount of training samples, the performances of all algorithms generally drop.

However, LBP-K_H is the most robust one, with small performance degradation. This is because the LBP operator encodes intrinsic structure of individual face images with even small training sample size, and the KCD-SRC framework successfully makes use of this advantage in the Hamming kernel. It is impressive that under only five training samples, the recognition rate of LBP-K_H can still reach over 97% while that of Pixel is only 70.18%, with over 20% improvement.
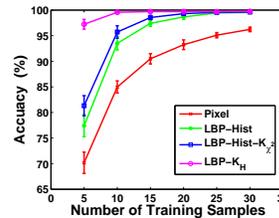

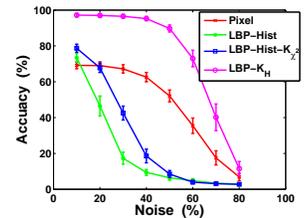**Fig. 2**. Performance with original Extended Yale B images.

Fig. 3. Performance with noise on Extended Yale B images.

With Noise: Next, we tested the robustness of the KCD-SRC algorithm with noise corrupted images. In this experiment we fixed the training sample size to be 5. Then for each test image we randomly corrupted a proportion of pixels from 10% to 80% (with a step of 10%) with noise. These chosen pixels were replaced by uniform distributed independent samples in [0, 255]. The recognition results are shown in Figure 3. As expected, the performance drops with the increasing amount of noise. Besides, it can be observed that LBP-K_H outperforms all other algorithms notably.

Surprisingly, LBP-Hist and LBP-Hist-K_{χ^2} drops drastically. They perform worse than Pixel after 30% of noise corruption. This may mainly due to the fact that local noise can change the neighboring comparison results in some directions (see Fig. 1), and so the affected LBP codes could disturb the histogram distribution over LBP bins seriously because of the discrete nature of LBP. To illustrate this, we draw an example in Figure 4 showing local LBP histograms with and without noise. Clearly, the distribution of LBP codes in the same region is severely affected by noise. In contrast, since Hamming kernel is calculated bitwisely, the ruined local comparisons might affect only a fraction (e.g. 30%) of Hamming counts. Therefore LBP-K_H is more robust against noise.
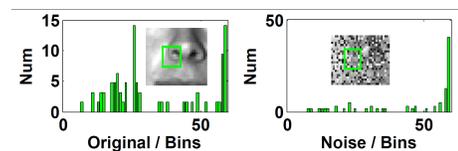

Fig. 4. An example of LBP histograms in a 8×8 local region without noise (left) and with 30% of noise (right).

With Occlusion: Finally we conducted experiments with random contiguous occlusions. Like in [1], we chose an ir-

related picture, resized it and attached it on each test image in an independent random position. 7 levels of occlusions were tested, with the size of the attached picture corresponding to 8×7, 16×14, $\cdots$, 56×49 respectively (see Figure 5). We also evaluated the four algorithms in this scenario, with randomly selected 5 samples per subject for training. Figure 5 shows corresponding results under different levels of occlusions. As illustrated, LBP-K_H performs the best, followed by LBP-Hist-K_{χ^2}, LBP-Hist, and Pixel. Obviously, LBP based methods are more robust against occlusion, compared to raw pixel based one. With the same LBP histograms, the proposed KCD-SRC with χ^2 kernel slightly outperforms that with linear kernel. It is also observed that the performances of all algorithms drop with the increasing amount of occlusion. It should be noticed that Hamming kernel with LBP maintains over 90% recognition rate even with up to 56% of occlusion. It outperforms the second best algorithm (LBP-Hist-K_{χ^2}) by over 15% for all the 7 levels of occlusions.
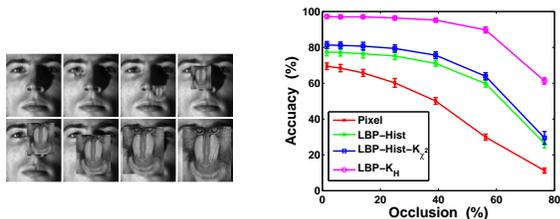

Fig. 5. Performance (right) under 7-levels occlusion (left) on Yale B

3.2. On PIE Database

Similar experiments were also done on the CMU-PIE database [12]. CMU-PIE database is built up with 41,368 images of 68 subjects under 13 different poses, 43 different illumination conditions, and with 4 different expressions. We only selected frontal face images for experiments, and cropped and resized them to 64×64 [13]. Due to the limited space, we only show experiment results with five training samples per subject in Table 1. In the results the noise level is 50%, and the occlusion level is 56.25% (48×48).

From Table 1 we can see that with original images all LBP based method perform consistently well. They have little difference because of the increased dimensions. Similar with the Extended Yale B database, LBP-Hist and LBP-Hist-K_{χ^2} have very low recognition rates with random noise. Yet LBP-K_H and Pixel are affected less by noise. As for the occlusion scenarios, LBP based methods perform much better than Pixel. Clearly, LBP-K_H is again the best algorithm among all.

4. CONCLUSION

In this paper, we have proposed a novel kernel coordinate descent (KCD) algorithm based on the covariance update technique for l_1 minimization problem in the kernel space. We

Table 1. Recognition Rates on CMU-PIE (%).

	Pixel	LBP-Hist	LBP-Hist-K_{χ^2}	LBP-K_H
Original	90.55±1.13	95.98±0.56	96.08±0.66	94.66±1.15
Noise	68.12±4.32	2.12±0.75	6.15±1.65	85.40±1.38
Occlusion	48.81±1.39	77.03±1.22	80.88±1.51	88.99±1.10

have applied the new algorithm in the sparse representation classification framework (which we call KCD-SRC) for face recognition, and have shown that the powerful LBP descriptor can be utilized in the proposed framework with two kernels based on the χ^2 distance and the Hamming distance. We have evaluated the proposed approach on both the Extended Yale B and the CMU-PIE face database. Experimental results show that KCD-SRC with LBP encodings and Hamming kernel performs impressively well against illumination variations, random noise, and continuous occlusions, even when there are only 5 training samples per subject. In most cases this combination can achieve 90% recognition rate, outperforming the original SRC with raw pixel values by 20% in the same experimental settings.

5. REFERENCES

[1] John Wright, Allen Y. Yang, Arvind Ganesh, S. Shankar Sastry, and Yi Ma, "Robust face recognition via sparse representation," *IEEE Trans. on PAMI*, vol. 31, pp. 210–227, 2009.

[2] Robert Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society, Series B*, vol. 58, pp. 267–288, 1994.

[3] Xiao-Tong Yuan and Shuicheng Yan, "Visual classification with multi-task joint sparse representation," in *IEEE Conference on CVPR*, 2010, pp. 3493 –3500.

[4] Chi Ho Chan and Josef Kittler, "Sparse representation of (multiscale) histograms for face recognition robust to registration and illumination problems," in *IEEE International Conference on Image Processing*, 2010, pp. 2441 –2444.

[5] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on feature distributions," *PR*, vol. 29, no. 1, pp. 51–59, January 1996.

[6] T. Ahonen, A. Hadid, and M.Pietikäinen, "Face recognition with local binary patterns," in *ECCV*, 2004, pp. 469–481.

[7] Jerome Friedman, Trevor Hastie, and Rob Tibshirani, "Regularization paths for generalized linear models via coordinate descent," *Journal of Statistical Software*, 2009.

[8] Shenghua Gao, Ivor Wai-Hung Tsang, and Liang-Tien Chia, "Kernel sparse representation for image classification and face recognition," in *ECCV:Part IV*, 2010, pp. 1–14.

[9] P. Tseng, "On accelerated proximal gradient methods for convex-concave optimization," *SIAM Journal on Optimization*, 2008.

[10] Honglak Lee, Alexis Battle, Rajat Raina, and Andrew Y. Ng, "Efficient sparse coding algorithms," in *NIPS*, 2007, pp. 801–808.

[11] K.C. Lee, J. Ho, and D. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," *IEEE Trans. on PAMI*, vol. 27, no. 5, pp. 684–698, 2005.

[12] Terence Sim, Simon Baker, and Maan Bsat, "The cmu pose, illumination, and expression (pie) database," in *IEEE International Conference on FG*, May 2002.

[13] "http://www.zjucadcg.cn/dengcai/data/data.html," .