Journal of Signal and Information Processing, **2013, 4, 86-90**
doi:10.4236/jsip.2013.43B015 Published Online August 2013 (http://www.scirp.org/journal/jsip)

Scientific Research

Combined Dictionary Learning in Facial Expression Recognition

Ziyang Zhang, Kaamran Raahemifar

Department of Electrical and Computer Engineering, Ryerson University, Toronto, Canada.
Email: zhangzyster@gmail.com

Received April, 2013.

ABSTRACT

Dictionary learning has been applied to face recognition and gets good results. However few works applied dictionary learning in facial expression recognition. This paper investigates the application of K-SVD in facial expression recognition. Since K-SVD focuses on reconstruction and lacks discriminant capability. It has similar classification performance with image pixel values. To address this problem, this paper proposes a Combined Dictionary Scheme, which uses combination of separate dictionaries. This yields better performance than the original single dictionary scheme in terms of both recognition rate and computation complexity.

Keywords: Facial expression Recognition; Dictionary Learning; K-SVD

1. Introduction

Emotion Recognition is an area that witnessed great amount of research efforts. Driven by the prospect that computer can fully understand human emotion and behaving like real people in the future, researchers are trying to find more and better ways to recognize human emotion.

Recognizing facial expressions from facial images is an important part of these efforts. Gabor features have been used in many works because of its insensitivity to face registration and good performance in facial expression recognition. [1] Used fisher discriminant criteria to select a subset of Gabor filters to reduce feature dimension and computation complexity, and then feed the reduced feature to a PCA+FLD classification scheme to recognize facial expressions. Though there are ways to somehow reduce the redundancy of Gabor features, it still suffers from high dimension and large amount of computation required. Further dimension reduction such as PCA is needed before Gabor feature is ready for classification.

One way to overcome this disadvantage is to use sparse representation, instead of Gabor features. Dictionary Learning and Sparse Representation are areas under extensive research in recent years, which have seen promising applications in signal compression, reconstruction, and pattern recognition, especially face recognition. However, few works were reported applying dictionary learning to facial expression recognition.

Given an over-complete dictionary, various sparse representation methods such as OMP (orthogonal matching pursuit) find a sparse signal that can best represent the original one. Though traditional DCT or various wavelets matrix can be used as dictionary, it is reported that better performance will result from learning dictionaries for specific signals to be represented. K-SVD is an iterative algorithm to build a dictionary for sparse representation that yields good reconstruction performance [2].

However, one drawback of K-SVD is that while doing well in reconstruction, it lacks discrimination capability to separate different classes. To solve this problem, [3] built a dictionary for each class, and use the reconstruction error on different dictionaries to classify a new sample. [4] Proposed a modification that introduces a discriminative part into the original K-SVD objective function, and by solving the new optimization problem, their method get better results in face recognition. These proposed methods all require more computation and are more time consuming than original K-SVD.

In this work, K-SVD is investigated to learn dictionary for facial expressions. Then OMP is applied to find sparse representations for facial images. Simple classification method, nearest neighbor, is used to test performance of K-SVD. To address the problem of lacking discriminant information in the learned dictionary, we propose a new scheme of combined dictionary, which achieves better recognition performance, while needs even less computation, than original single dictionary

scheme.

The rest of this paper is organized as follows: Section 2 illustrates the preprocessing of images. Section 3 introduces the K-SVD algorithms and how it can be applied to facial expression recognition. Section 4 discusses the experiment method and shows results of the original single dictionary scheme. The proposed combined dictionary scheme is illustrated in section 5, and the results are given and analyzed. Section 6 draws conclusion and talks about possible future work.

2. Normalization of Faces

A practical facial expression recognition system requires automatic detection and extraction of human faces. There are many efficient ways to accomplish this task, such as the famous Viola Jones Detector [5], which has been implemented in the Open CV library.

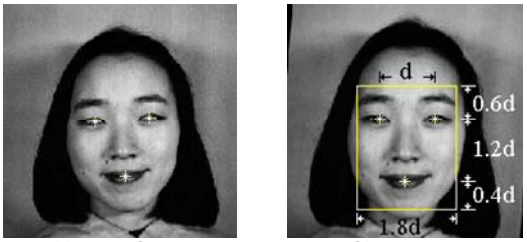
In this work, in order to focus on feature extraction using dictionary learning, only manually extraction of face is used. The faces then need to be normalized to minimize the difference of lighting conditions. Similar with the method used in [1] and [6], there are 3 steps to obtain a normalized face:

- Manually locate center of eyes and mouth, as shown in **Figure 1(a)**. Build an affine transform to adjust the 3 points to fixed position. This transform may consists of translation, scaling and rotation.
- Apply a geometric face model to crop out the face area, as shown in **Figure 1(b)**.
- Perform histogram equalization on the rectangular faces obtained in step 2.

Some of the normalized faces are shown in **Figure 1(c)**.

3. Learning Dictionary of Expressions Using K-SVD

The goal of dictionary learning is to find a dictionary that can achieve the best sparse representation of a given signal, which can be modeled as (1).



(a) Locate key points. (b) Crop Image

(c) Normalized faces.

Figure 1. Examples of face normalization.

$$\min_{\mathbf{D},\mathbf{X}} \left\|\mathbf{Y}-\mathbf{D}\mathbf{X}\right\|_F^2 \text{ subject to } \forall i, \left\|\mathbf{x}_i\right\|_0 \le T \tag{1}$$

where $\mathbf{Y}=[\mathbf{y}_1,\mathbf{y}_2,...,\mathbf{y}_M]$ is a $N\times M$ matrix containing M original signals with N dimensions; $\mathbf{X}=[\mathbf{x}_1,\mathbf{x}_2,...,\mathbf{x}_N]$ is a $K\times M$ sparse representation matrix, with less than T non-zero elements in each column, while $\mathbf{D}=[\mathbf{d}_1,\mathbf{d}_2,...,\mathbf{d}_K]$ is the dictionary, which has K atoms.

K-SVD is an iterative algorithm to learn a dictionary that minimizes the representation error in (1). Each iteration consists of two steps: sparse coding and dictionary updating [2].

In the sparse coding step, a sparse representation using the current dictionary is calculated, using methods like OMP, BP, or FOCUSS. The objective of this step is described in (2).

$$\begin{aligned}\mathbf{x}_i = \arg\min_{\mathbf{x}_i} &\left\|\mathbf{y}_i-\mathbf{D}\mathbf{x}_i\right\|_2^2 \\ \text{subject to } &\left\|\mathbf{x}_i\right\|_0 \le T, \\ \text{for } &i=1,2,...,\mathrm{M}\end{aligned} \tag{2}$$

In the dictionary updating step, for each dictionary atom $\mathbf{d}_k$, first find the error of current reconstruction without that atom.

$$\mathbf{E}_k = \mathbf{Y}-\sum_{j\ne k}\mathbf{d}_j\cdot\mathbf{X}(j,:) \tag{3}$$

where $\mathbf{X}(j,:)$ uses MATLAB terminology, which represents the j^{th} row of $\mathbf{X}$.

Then select the those columns in the error matrix that corresponding to which original signal use the atom $\mathbf{d}_k$.

$$\mathbf{E}_k^R = \mathbf{E}_k\left(:,\text{find}\left(\mathbf{X}(k,:)\ne 0\right)\right) \tag{4}$$

Perform SVD (singular value decomposition) on $\mathbf{E}_k^R$, and use the resulting column of $\mathbf{U}$ and $\mathbf{V}$ to update $\mathbf{d}_k$ and its corresponding non-zero coefficients.

Repeat the two steps until certain error is satisfied or certain number of iteration has finished. Then, use the resulting dictionary D in sparse coding to find the final sparse representation.

In the previous section, normalized face images have been achieved. Arrange the pixel values of each image into a column vector and perform K-SVD on training samples to learn a dictionary. Some of the dictionary atoms are shown in **Figure 2**. Note that the dictionary atoms are very similar to the Eigen-faces derived from PCA [7], so it can be regarded as a method of dimension reduction.

Since each image will have a sparse representation based on the learned dictionary, the sparse representation then can be used for classification or reconstruction. From this point of view, the sparse representation can also be regarded as feature of the input image.

Figure 2. First 16 atoms in the learned dictionary.

4. Experiment Results on Single Dictionary Scheme

The learning of dictionary focuses on minimizing the reconstruction error, and does not require class labels. Thus, a convenient way of constructing the dictionary is to use all the training samples to perform K-SVD. We call this a Single Dictionary Scheme, which is the case in most traditional works performing classification after dictionary learning.

In our work, a simple nearest neighbor classifier is used to test the classification performance of learned dictionary. For each class, the mean is calculated from the sparse representation of training samples. A test sample is assigned to the class, whose mean is closest to the sample, in terms of Euclidian distance.

The test is performed on JAFFE (Japanese Female Facial Expression) database. It consists of 10 persons with 7 different expressions and has been widely used to test various methods of facial expression recognition.

In order to find an optimal parameter set, we change the number of atoms and sparsely constraints, and test on randomly selected images. **Figure 3** shows the results of recognition rate and time of learning dictionaries. As the figure shows, an increase in the number non-zero coefficients will increase the recognition rate to some certain level, and then the recognition rate stays insensitive to the sparsity constraint. Besides, **Figure 3(b)** indicates the time consumption is mainly determined by the sparsely constraint, with an exponential relation. Size of the dictionary, which is the number of atoms, have a very small impact on the recognition rate. Overall, the best result comes when there are 70 atoms in the dictionary and 20 non-zero coefficients. Following experiments will use this parameter set.

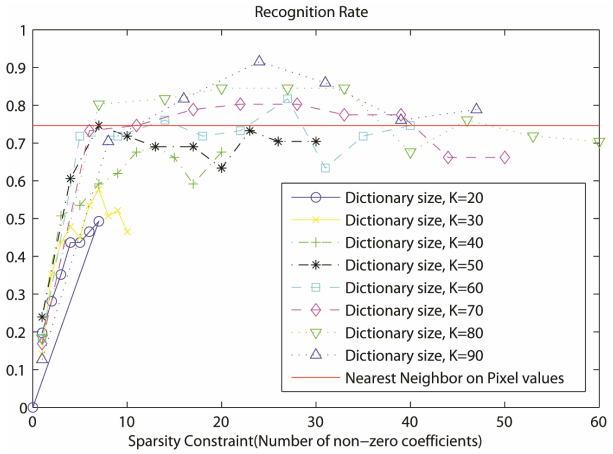
The recognition rates of the Single Dictionary Scheme are listed in **Table 1**. Two test methods [8] are used here: Cross validation randomly chose one image per person per class as test samples, while the rest as training samples, repeat the training and testing 10 times to acquire a mean recognition rate; Leave-one-out method use one image as test sample and all the rest images as training samples, repeat this procedure to test all the images in the database and find the mean recognition rate.

5. Combined Dictionary Scheme

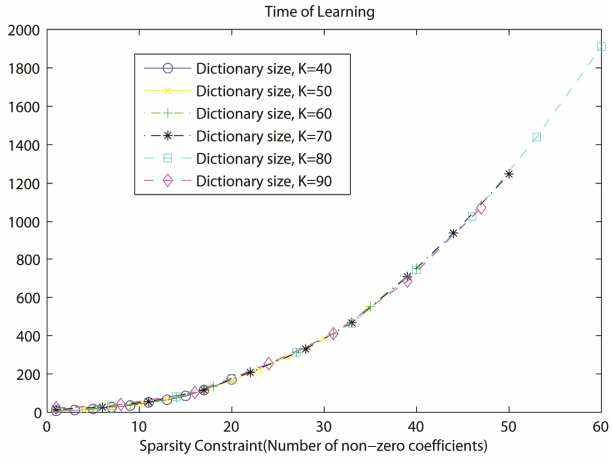
Results from **Table 1** show that sparse representations from the learned dictionary get similar recognition rate with original pixel values. This again indicates that K-SVD leads to good representation, but can barely increase the discriminant capability.

Table 1. Recognition Rate of Single Dictionary Scheme

	Test Method	Recognition Rate
Single Dictionary Scheme	Cross validation	84.51%
	Leave one out	63.85%
Directly from Pixel values	Cross validation	76.06%
	Leave one out	61.50%



(a)



(b)

Figure 3. Results on different parameter sets.

Though there are ways to include a discriminant factor into the objective function solve the new optimization problem and thus improve the classification performance, they are often more time consuming to make additional computation. One easier approach is to keep the dictionary learning process unchanged, and finds better ways to reconstruct the dictionary using class information.

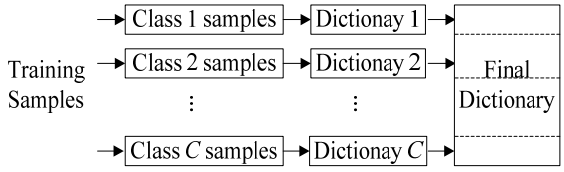


Figure 4. Combined Dictionary Scheme.

As a simple way of reconstructing dictionary, the proposed Combined Dictionary Scheme combine multiple dictionaries into one. More specifically, learn dictionary from the training samples of each single class, and combine these dictionaries into a C times larger diction- ary, as shown in **Figure 4**, where C is the number of classes.

Therefore, in order to keep the same dictionary size and sparsity constraint with Single Dictionary Scheme, the dictionaries need to be learned can have C times smaller size and sparsity constraint. It has been discussed from **Figure 3** that the sparsity constraint (number of non-zero coefficients) mainly determines the computation complexity in an exponential way. It can be expected that the Combined Dictionary Scheme will dramatically reduces the time of learning dictionaries. As in the previous section, dictionary size of 70 and sparsity constraint 20 are used, here for each of the 7 dictionaries, the sizes are set to 10 and sparsity constraints are set to 3.

Once again, nearest neighbor is used to test performance. **Table 2** and **Table 3** compare the recognition rate and time of computation with Single Dictionary Scheme and an-other Gabor based algorithm in literature.

Table 2. Comparison of Recognition Rate.

	Test Method	Recognition Rate
Single Dictionary Scheme	Cross validation	84.51%
	Leave one out	63.85%
Combined Dictionary Scheme (*Proposed*)	Cross validation	91.55%
	Leave one out	84.04%
Gabor + Nearest Neighbor	Cross validation	83.09%
	Leave one out	-
Gabor + PCA + FLD [1]	Cross validation	-
	Leave one out	93.90%

As shown in **Table 2** and **Table 3**, the proposed Combined Dictionary Scheme achieves better results than the original Single Dictionary scheme. The recognition rate is increased by 7 and 20 percentage, respectively in the two test methods, while reducing the computation complexity by approximately 3 times.

Table 3. Comparison of Computation Complexity.

	Training Time*	Testing Time*
Single Dictionary Scheme	61.14	0.0382
Combined Dictionary Scheme (*Proposed*)	16.99	0.0101
Gabor + PCA + FLD [1]	99.06	1.26

*Training time refers to the average time of training 140 samples; Testing time refers to average time of testing one sample; All times are in seconds, acquired from Matlab running on a desktop computer.

Though the recognition performance of dictionary learning is not as good as Gabor+PCA+FLD [1], the average time of testing one image is more than 100 times less. We should also note that the recognition rate here is based on nearest neighbor, it could be expected that the performance would be better if more powerful classification algorithms are used. If the same nearest neighbor classifier is applied to Gabor feature, the recognition rate is worse than that of dictionary learning. This indicates that dictionary learning and sparse representation might get more promising results in the future.

6. Conclusions

This paper investigates the application of K-SVD dictionary learning in facial expression recognition. The sparse representation of a facial image is regarded mainly as a way of extracting features and at the same time with low dimensions Sparse representation based on learned dictionary directly from all training samples (Single Dictionary Scheme) yields similar recognition performance with original image pixel values. This demonstrates that K-SVD focuses on minimizing reconstruction error, and does not provide good discriminate capability.

In order to improve the classification performance, a Combined Dictionary Scheme is proposed, so that the class information can contribute to the dictionary construction. First learn a separate dictionary for each class using the corresponding samples in the training set. Then combine them into a larger dictionary for the final training and classification. The proposed method gets better classification performance than the traditional Single Dictionary Scheme, in terms of both recognition rate and computation complexity.

Though the current recognition rate is not as good as classification systems using other features, it might be due to the simple nearest neighbor classifier used for testing. Also dictionary learning needs much less time to compute than Gabor features. Since currently there are few works applying dictionary learning to facial expression recognition, the performance would be further improved in the future, if more powerful classification algorithms are used.

REFERENCES

[1] Z. Y. Zhang, X. M. Mu and L. Gao, "Recognizing Facial Expressions Based on Gabor Filter Selection," *Proceedings of 2011 4th International Congress on Image and Signal Processing (CISP)*. IEEE, 2011, Vol. 3, pp. 1544–1548.

[2] Michal Aharon, Michael Elad, and Alfred Bruckstein, "K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation," *IEEE Transactions on Signal Processing*, Vol. 54, No. 4, 2006, pp. 4311–4321.doi:10.1109/TSP.2006.881199

[3] Julien Mairal and Francis Batch et al., "Discriminative Learned Dictionaries for Local Image Analysis," *Proceedings of 2008 IEEE Conference on Computer Visionand Pattern Recognition (CVPR)*, IEEE, 2008, pp. 1–8.

[4] Z. L. Jiang, L. Zhe and L. S. Davis, "Learning a Discriminative Dictionary for Sparse Coding via Label Consistent K-SVD," *Proceedings of 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 2011, pp. 1697-1704.

[5] Paul Viola and Michael Jones, "Robust Real-time Face Detection," *International Journal of Computer Vision*, vol. 57, No. 2, 2004, pp. 137–154. doi:10.1023/B:VISI.0000013087.49260.fb

[6] S. Dubuisson, F. Davoine and M. Masson, "A Solution for Facial Expression Representation and Recognition," *Signal Processing: Image Communication*, Vol. 17, No. 9, 2002, pp. 657–673. doi:10.1016/S0923-5965(02)00076-0

[7] N. Peter Belhumeur, P. Joao Hespanha and David J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, 1997, pp. 711-720. doi:10.1109/34.598228

[8] F. Y. Shih, C. F. Chuang and Patrick, S. P. Wang, "Performance Comparisons of Facial Expression Recognition in JAFFE Database," *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 22, Vol. 3, 2008, pp. 445-459.doi:10.1142/S0218001408006284

