A structured dictionary learning framework for sparse representation

Yin Wei

Information Science and Engineering, Wuhan University of Science and Technology, Wuhan 430000, Hubei, China

Keywords: sparse coding, face recognition, fisher criterion, spatial pooling, svm classifier.

Abstract. With the development of the computer, BOW model and SC model are applied to a large number of image classifications, and exhibit excellent performance, which become a hot topic in the field of computer vision. In the paper, we proposed a new framework of dictionary learning. the objective function based on sparse representation just consider the sparsity, while ignoring the spatial information of image and the correlation information of features, we apply spatial pyramid matching and add the discrimination fisher regularized penalty, by performing iterative optimization can get an excellent dictionary for representing image features, finally, we use max pooling and svm classifier for image classification. Experimental results in ORL and YALE face database show that the method has a high resolution.

1. Introduction

Face recognition has been a hot topic in the field of computer vision. Many image representations have been proposed, for example, BOW (Bags of Words) model, Part-based model and so on. Among these models, BOW model which shows excellent performance is widely used in practice, Initially Bow model is applied to text classification by representing the document as feature vectors. In 2006, S. Lazebnik, C. Schmid etc. proposed SPM model based on BOW model, they combine both model to make up the ignore of spatial information in BOW model, building the classic BOW and SPM (Spatial Pyramid Matching) model, and applying it to the classification of natural images.

BOW and SPM model includes three modules: 1) image blocks and feature extraction; 2) generate the codebook and vector quantization; 3) spatial pooling.

In the three modules, the establishment of codebook and spatial pooling impact on the quality of the image representation directly. Codebook is a collection of the potential pattern of all samples. In the traditional BOW and SPM model, the establishment of codebook applies K-means algorithmto cluster all sample features, and ultimately get the K cluster centers, which are called codebook. However, this cluster method is the computational complexity, and the cluster centers can't represent all sample features well. In 2009, Yang J proposed a method of using the sparse coding instead of the traditional K-means algorithm, which obtained a codebook by solving a convex optimization, then applied SPM and Max pooling to Summarize all the SIFT features of an image to a vector, finally, using the svm classifier for image classification. Experimental shows that the established codebook has an excellent image representation, and result in a high-resolution. However, in the establishment of codebook, the optimization strategy just considers the single sample feature, but ignores the spatial information and correlation information of sample features. To make up this drawback, the paper proposes a dictionary learning framework based on fisher regularied penalty combined the ability of reserved spatial information of ScSPM, building a new framework for face recognition.

2. Sparse coding model

In 2009, Wright J proposed a face recognition method based on sparse representation. Suppose c class training sample$B = \{B_1, B_2, \dots, B_c\} \in R^{m\times k}$, B_i is the i-th class sample, now there is a test sample$x \in R^{m\times 1}$, then x can be formed of B approximated by a linear fit. The process of sparse coding is as follow:

Sparsely code x on B via L1-regularied minimization:

$\hat{s} = \mathrm{argmin}_s \|s\|_1$ (1)

$$s.t.Bs = x$$

Or

$\hat{s} = arg\min_{s}\|x - Bs\|^2 + \lambda\|s\|_1$ (2)

Where λ is an adjustment factor to balance reconstruction error and sparsity.

Do classification via minimizing the reconstruction error:

$\text{identity(x)} = \text{argmin}_i(e_i)$ (3)

Here $e_i = \|x - B_i\hat{s}_i\|$, i is the class label.

3. Dictionary learning based on fisher regularied penalty

Dictionary learning, as a kind of unique signal sparse representation model, has caused the attention of many scholars in recent years. Its essence is to learn a very complete dictionary with a lot of training sample learning to represent the input signal. The quality of the dictionary determines the quality of the image sparse representation directly. There are many scholars who put forward several of improvement measures, and proposed a consensus: the sparse coding model with the classifier penalty can improve the classification performance effectively. Gangeh M J proposed a dictionary learning method which combined maximum the correlation of signal and response class; Ramirez, Sprechmann make the sparse coefficient stronger discrimination by adding the items of categories in the sparse coding model; Zhang etc. proposed a K-SVD dictionary learning method based on category. In this article, we put forward an alternative optimization model based on the category of sparse coefficient.

Fisher criterion is a good rule of the linear discriminant analysis, can be used as a linear classifier. Assuming donate all the training sample by random sampling SIFT features as $X = \{X_1, X_2, \ldots, X_c\} \in R^{m\times n}$, X_i is the collection of i-th class samples, the codebook that we intend to learn is defined as $B = \{b_1, b_2, \ldots, b_k\} \in R^{m\times k}$, and $S = \{S_1, S_2, \ldots, S_c\} \in R^{k\times n}$ is the sparse representation matrix that X represents on B, then $X \approx BS$. Here S_i is the sparse representation matrix that i-th class samples represent on B. So the Divergence of between classes KL_b and with-in classes KL_w can be expressed as:

$KL_b = \sum_{i=1}^{c} n_i\,(m_i - m)(m_i - m)^T$ (4)

$KL_w = \sum_{i=1}^{c}\sum_{S_i\in S}(S_i - m_i)\,(S_i - m_i)^T$ (5)

Where m_i and m are the mean of the i-th sparse coefficient and the entire sparse sample coefficient,n_i is the number of i-th class samples.

Here we define $f(S)$as $Tr(KL_w - KL_b)$, $Tr(°)$ express the trace of a matrix. However, $f(S)$ is a non-convex function, in order to ensure that the final model is a convex optimization problem, we amend $f(S)$ as follows:

$f(S) = Tr(KL_w - KL_b) + \eta\|S\|_F^2$ (6)

In the above formula, η is the adjustment factor.

3.1 fisher criterion penalty model

By the formula (2) and (6), we obtain the following model:

$J(B,S) = argmin_{B,S}\|X - BS\|_F^2 + \lambda_1\|S\|_1 + \lambda_2(Tr(KL_w(S) - KL_b(S)) + \eta\|S\|_F^2)$ (7)

As can be seen from the above equation, $J\,(B,S)$ is a convex function, so this is a typical convex optimization problem, which includes two variables, for this convex optimization problem, the general method is iterative optimization algorithm:

fixed$B$, updated S

fixed$S$, updated B

Currently, many algorithms of solving this optimization have been proposed, the next section describes the optimization process of our method.

3.2 Dictionary learning

We can see from (7) that this optimization problem is high-dimensional and complex, here we use the common multi-variable optimization method to optimize *B* and *S* separately (detailed procedure in Algorithm 1).

Firstly, random initialize dictionaries B, and get the training samples $X = \{X_1, X_2, \ldots, X_c\} \in R^{m\times n}$byrandom sampling the SIFT features.

Keep dictionary $\boldsymbol{B}$ fixed, updating sparse coefficient $\boldsymbol{S}$. The object function reduces to the following optimization functions:

$$J(S_i) = min\|X_i - BS_i\|_F^2 + \lambda_1 \|S_i\| + \lambda_2 f(S_i) \quad (8)$$

Here X_i is the collection of i-th class samples, S_i is the sparse coefficient matrix that X_i represents on B.λ_1,λ_2 is adjustment factor. $f(S_i) = \|S_i - M_i\|_F^2 - \sum_{k=1}^{c}\|M_i - M\|_F^2 + \eta\|S_i\|_F^2$,define $E_i^i = [1]_{n_i\times n_i}$, then $f(S_i)$ is changed as follow:

$$f(S_i) = \|S_i N_i\|_F^2 - \|S_i P_i - G\|_F^2 - \sum_{\substack{k=1\\k\neq i}}^{c}\left\|Z - S_i C_i^k\right\|_F^2 + \eta\|S_i\|_F^2 (9)$$

Here $N_i = I_{n_i\times n_i} - E_i^i/n_i$, $P_i = \frac{E_i^i}{n_i} - \frac{E_i^i}{n}$, $Z = \frac{S_k E_k^k}{n_k} - \sum_{j=1,j\neq i}^{c} S_j C_j^j$, $G = \sum_{k=1,k\neq i}^{c} S_k C_k^i$, $C_i^j = E_i^j/n$. Seen from formula (8) and (9), we make sense that the object function is differentiable and strictly convex. Thus we can use gradient descent to solve this kind of optimal problem. Lorenzo R published a report in 2009. He proposed an iterative projection method for structural optimization of sparse regularization, and described the procedure of iterative projection optimization method based on several of models in details. This method is used in this article.

Keep the sparse coefficient $\boldsymbol{S}$ fixed, updating dictionary. The objective function reduces to the following optimization functions:

$$\mathrm{argmin}_B \|X - BS\|_\mathrm{F}^2$$
$$s.t. \sum_{i=1}^{k} b_{i,j}^2 \leq \varphi, \forall j = 1,2, \ldots, n \ (10)$$

This is a typical problem of least squares optimization. Generally, this optimization problem can be solved by gradient descent method. And we also can solve it with the optimization tools, for example, CVX etc. However, these methods are generally time-consuming. As can be seen from the above equation, we can use the Lagrangian dual optimization methods to solve it. It's simple and efficient. Its Lagrange form is as follow:

$$L(B, \vec{\lambda}) = \|X - BS\|_F^2 + \sum_{j=1}^{n} \lambda_j \sum_{i=1}^{k}(b_{i,j}^2 - \varphi) \quad (11)$$

Here λ_j is a dual variable. Minimizing over B analytically, we obtain the Lagrange dual:

$$D(\vec{\lambda}) = min_B\, Tr(X^T X - XS^T(SS^T + \Lambda)^{-1})\,(XS^T)^T - \varphi\Lambda)(12)$$

Seeking the maximum of $D(\vec{\lambda})$, $\Lambda = diag(\vec{\lambda})$, Compute the gradient and hessian matrix of $D(\vec{\lambda})$:

$$\frac{\partial D(\vec{\lambda})}{\partial \lambda_i} = \|XS^T(SS^T + \Lambda)^{-1} e_i\| - \varphi \qquad (13)$$

$$\frac{\partial^2 D(\vec{\lambda})}{\partial \lambda_i \partial \lambda_j} = -2((SS^T + \Lambda)^{-1}(XS^T)^T(SS^T + \Lambda)^{-1})_{i,j}((SS^T + \Lambda)^{-1})_{i,j} (14)$$

Here e_i is unit vector. Conjugate gradient descent method is used to optimize (12) for getting a suitable resolution (or optimize by build-in functions of Matlab). Compute the partial derivative of B in (11) and set it to zero. Then get the equation (15):

$$B^T = (SS^T + \Lambda)^{-1}(XS^T)^T \qquad (15)$$

After the multiple iterations above, finally we get an excellent dictionary B which can represent the samples sparsity. Then we do sparse representation for SIFT features of every image with this dictionary. In this paper, we do BOW representation for each sub-region by maximum spatial pool. Because studies of many scholars show that maximum space pool and sparse coding is efficient in image classification. Finally, we cascade the BOW representation vectors of all region to represent the entire image's features. And classify the face images with SVM Classifier.

4. Experimental research

In order to evaluate the feasibility and high resolution of this method, we select the classic face database ORL and YALE for experimental research. Figure 1 list the parts of two image databases, as can be seen from the image, two face databases are gray images, different image has the obvious difference because of light causes and the display brightness, some face images have obvious deflection and glasses. ORL face image database has a total of 400 pieces of 128×128 image, 40 individuals, 10 images per person; YALE face image database is 165 pieces of 100×100 image, 15 individuals, 11 images per person. In the experimental formula, we set λ_1 to 0.015, λ_2 to 0.05, sampling 5000 SIFT features from all samples for dictionary learning.

Fig 1 ORL face database at left, YALE face database at right

In this paper, we mainly evaluate the algorithm from three aspects: 1. Effect on image recognition rate of dictionary size; 2. the influence on results of the number of training images; 3. experimental results were compared with the results of other algorithms.

Since each image has a large number of SIFT features, we are to sample a certain amount from SIFT features of each image in dictionary training, and at the end of the classification, we divide the image into two parts, one part is used for the training set, the other part is the test set.

Figure 2 shows the effect on face recognition rate of dictionary size, as can be seen from the figure, the face recognition rate increases while the number of the dictionary atoms increases, and can reach about 99% in the YALE image database. In Table 1 we studied the effect on image resolution of the number of training images, took respectively 3, 5, 7 pairs of images from two image database for SVM classification, finally, experimental results show that the influence on image resolution of the number of training images is large.

We apply this method and other methods to carry on the contrast research, select the YALE image database, as can be seen from table 2, this method has good classification ability in face recognition, and improves the recognition rate of about 2% in contrast to the ScSPM method.

Table 1 Influence on image resolution of the number of training image

The number of training image		3	5	7
This method	ORL	91.2%	97.4%	98.6%
	YALE	92.9%	98.3%	99.2%

Table 2 the experimental results of different algorithms on YALE image database

algorithm	SRC	SVM	ScSPM[4]	This method
recognition	90.0%	88.8%	96.7%	98.8%

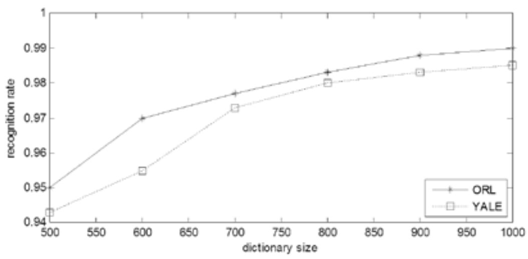

Fig 2 Influence on image resolution of dictionary size

5. Conclusion

In this article, we propose a new sparse coding framework. Since the model of ScSPM and the standard sparse coding take the sparsity and the reconstruction error into consider, but ignore the spatial information and correlation between samples, we add the Fisher regularization penalty term in the model combined with the ability ofScSPM reserving the image spatial information, finally, apply SVM classifier for classification in image database. Experimental results show that the method has high recognition rate. But I have to say that this method has a fatal weakness, which is that dictionary learning is too time consuming. Therefore, how to improve the efficiency of dictionary learning becomes a future research.

Reference

[1] Chen J, Towfic Z J, Sayed A H. Online dictionary learning over distributed models[C]. //Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on. IEEE, 2014:3874 - 3878.

[2] Engan K, Hå K S A J, Husø K. Family of iterative LS-based dictionary learning algorithms, ILS-DLA, for sparse signal representation[J]. Digital Signal Processing, 2007,3 (1):32-49.

[3] Galvez-López D, Tardos J D. Bags of Binary Words for Fast Place Recognition in Image Sequences[J]. Robotics, IEEE Transactions on, 2012, 28(5):1188 - 1197.

[4] Gangeh M J, Ghodsi A, Kamel M S. Kernelized Supervised Dictionary Learning[J]. GnalRongRanaon on, 2013, 7(19):4753-4767.

[5] G, u, i, et al. Classification and clustering via dictionary learning with structured incoherence and shared features[J]. Proceedings / CVPR, IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2010, 23(3):3501 - 3508.

[6] Gang L, Xiaochi W. Improved Bags-of-Words Algorithm for Scene Recognition[J]. Physics Procedia, 2010,7(24):1255–1261.

[7] He B, Xu D, Nian R. Fast Face Recognition Via Sparse Coding and Extreme Learning Machine[J]. Cognitive Computation, 2014, 6(2):264-277.

[8] Lorenzo R,S F Mosci. Iterative Projection Methods for Structured Sparsity Regularization[R].Computer Science and Artificial Intelligence Laboratory Technical Report,2009:1-28.

[9] S G, IW T, LT. C. Laplacian Sparse Coding, HypergraphLaplacian Sparse Coding, and Applications[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(1):92-104.

[10] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: spatial pyramid matching for recognizing natural scene categories[C].In CVPR, 2006,11: 2169-2178.

[11] Skretting K, Engan K. Recursive Least Squares Dictionary Learning Algorithm[J]. Signal Processing, IEEE Transactions on, 2010, 58(4):2121 - 2130.

[12] Wright J, Yang A Y, Ganesh A, et al. Robust Face Recognition via Sparse Representation[J]. ArnAnaly and MahnNllgnRanaon on, 2009, 31(2):210 - 227.

[13] Wei D, Chen F, Zhang T. Least Square-Support Vector Regression based car-following model with sparse sample selection[C]. //Intelligent Control and Automation (WCICA), 2010 8th World Congress on. IEEE, 2010:1701 - 1707.

[14] Xu J, Man H. Dictionary Learning Based on Laplacian Score in Sparse Coding[J]. Machine Learning and Data Mining in Pattern Recognition, 2011,9(6871):253-264.

[15] Yang J, Yu K, Gong Y, et al. Linear spatial pyramid matching using sparse coding for image classification[C]. //Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009:1794 - 1801.

[16] Zhang Q, Li B. Discriminative K-Svdfor Dictionary Learning In Face Recognition[J]. Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, 2010, 119(5):2691 - 2698.