Tensor Discriminant Analysis for View-based Object Recognition

Yong Wang and Shaogang Gong
Department of Computer Science
Queen Mary, University of London, E1 4NS, United Kingdom
{ywang, sgg}@dcs.qmul.ac.uk

Abstract

In this paper, we use a general M^{th} order tensor discriminant analysis approach [11] for view based object recognition. This method is an extension of the 2D image coding technique [10] to general M^{th} order tensors for discriminant analysis, and has good convergence property. We demonstrate the performance advantages of this approach over existing techniques using experiments on the COIL-100 and the ETH-80 datasets. Specifically, our experimental results on ETH-80 show the particular strength of this tensor discriminant analysis method when only a small number of training samples with big intra-class variation are available.

1. Introduction

View based 3D object recognition remains to be a hard problem for computer vision. Several methods have been proposed in the literature. Poggio and Edelman [6] proposed an exemplar-based approach with a network of generalized radio basis functions for recognising a 3D object from its 2D images. Another approach, known as parametric appearance eigenspace, was proposed by Murase and Nayar [5] and has been further extended effectively to modelling face images by Gong et al. [2]. Generic machine learning techniques have also been used for view based object recognition. Particularly, support vector machines (SVMs) have been extensively evaluated for object recognition. Both linear and non-linear kernels have been used and achieved good results on benchmark data sets [9, 7]. Another machine learning technique, called Sparse Network of Winnows (SNoW), has also been shown to be effective for view based object recognition [13]. SNoW is able to learn explicitly a representation of an object, unlike SVMs which only define discriminating boundaries. However, all of the above has one common characteristic: representing 2D images by 1D vectors. This vectorization is rather adhoc and not optimal because it does not preserve any nonlinear structure and shape information of the data. It can also result in a very large image representational space with poor numerical properties and computational tractability.

In this work, we use a tensor discriminant analysis approach by Tao et al. [11, 1] for view based object recognition. This method represents a color image as a 3rd order tensor, resulting in a much smaller dimension. Take a 64×64 color image as an example, the dimension will be $64 \times 64 \times 3$ if we vectorize the image but only 64 in a tensor structure because the tensor framework will deal with each dimension separately. A similar idea of 2D tensor discriminant analysis has also been proposed by Ye et al. [14]. This was further extended to a general M^{th} order tensor by Yan et al. [12]. Alternatively, the method we use here has extended the tensor rank one decomposition for image coding proposed by Shashua and Levin [10] in the following two ways: First, the rank one decomposition is optimized for classification instead of representation. Second, it is a generalization from 2D images to M^{th} order tensors. A significant advantage of this method over that of Yan et al. [12] is its superior convergence property while the latter is difficult to converge if not impossible. For classification, we map the original tensor objects into a low dimensional feature space and use a nearest neighbor classifier and AdaBoost.

The remaining parts of this paper are organized as follows. We outline the algorithm of tensor discriminant analysis in section 2. In section 3 we introduce the dataset and experimental setup we used. We report the experimental results in section 4 and conclude this paper in section 5.

2. Tensor Discriminant Analysis

2.1 Tensor Rank-one Decomposition

To model 2-D data without rasterization, Shashua and Levin [10] demonstrated the advantage of tensor rank-one decomposition (TROD) over PCA for image coding. Given a set of M^{th} order tensors $\{\mathbf{X}_i \in \mathbb{R}^{I_1 \times I_2 \times \ldots I_M} | i = 1, 2, \ldots, N\}$, and define a rank one tensor τ as the outer product of M unit vectors $\{u^d \in \mathbb{R}^{I_d} | d = 1, 2, \ldots, M\}$: $\tau = \prod_{d=1 \times_d}^{M} u^d$. TROD looks for the optimal rank one tensors $\{\tau_r |\ r = 1, 2, \ldots, R\}$ to minimize the overall re-

The 18th International Conference on Pattern Recognition (ICPR'06)
0-7695-2521-0/06 $20.00

construction error E with minimal possible R:

$$E = \sum_{i=1}^{n} \left\|\varepsilon_i^R\right\|_F^2 = \sum_{i=1}^{n} \left\|\mathbf{X}_i - \sum_{s=1}^{R} \lambda_{i,s}\tau_s\right\|_F^2 \quad (1)$$

$$\varepsilon_i^r = \mathbf{X}_i - \sum_{s=1}^{r} \lambda_{i,s}\tau_s \quad (2)$$

$$\lambda_{i,r} = \left(\mathbf{X}_i - \sum_{s=1}^{r-1} \lambda_{i,s}\tau_s\right) \prod_{d=1}^{M} \times_d u_r^d \quad (3)$$

where ε_i^r is the r^{th} reconstruction error of i^{th} tensor. $\lambda_{i,r}$ is the projection coefficient of the $(r-1)^{th}$ reconstruction error of the i^{th} tensor on τ_r. Given the rank one tensors $\{\tau_r\}$, a tensor object $\mathbf{X}_i$ can be approximately reconstructed by its corresponding coefficients $\{\lambda_{i,r}\}$.

2.2. Discriminant Tensor Rank-one Decomposition

While TROD looks for the optimal $\{\tau_r\}$ to minimize the overall reconstruction error, discriminant tensor rank-one decomposition (DTROD) [11] looks for the optimal $\{\tau_r\}$ so that the projection coefficients $\{\lambda_r\}$ are optimal for discrimination. Similar to the relationship between PCA and LDA, TROD is optimal for representation and DTROD is optimal for classification. Suppose we have tensor objects from C classes. The c^{th} class has n_c tensor objects and the total number of tensor objects is n. Let $\mathbf{X}_i^c$ be the i^{th} object in the c^{th} class. The mean tensor of the c^{th} class $\mathbf{M}_c$ and the overall mean tensor $\mathbf{M}$ are defined as:

$$\mathbf{M}_c = \frac{1}{n_c}\sum_{i=1}^{n_c} \mathbf{X}_i^c, \qquad \mathbf{M} = \sum_{c=1}^{C} \frac{n_c}{n}\mathbf{M}_c \quad (4)$$

The between-class scatter $S_b(\tau)$ and within-class scatter $S_w(\tau)$ with regard to the rank one tensor τ is defined as

$$S_b(\tau) = \frac{1}{n}\sum_{c=1}^{C} n_c < (\mathbf{M}_c - \mathbf{M}), \tau >^2 \quad (5)$$

$$S_w(\tau) = \frac{1}{n}\sum_{c=1}^{C}\sum_{i=1}^{n_c} < (\mathbf{X}_i^c - \mathbf{M}_c), \tau >^2 \quad (6)$$

$$< \mathbf{A}, \tau > = \mathbf{A}\prod_{d=1}^{M} \times_d u_r^d \quad (7)$$

The inner product between a general tensor $\mathbf{A}$ and a rank one tensor τ is defined by Eq.(7). Compared to classic LDA, we can find that τ functions in the similar way to that of a single projection vector (a column vector in the expected transformation matrix in LDA). We optimize τ according to:

$$\tau = \arg\max_{\tau}(S_b(\tau) - \zeta S_w(\tau)) \quad (8)$$

with an optimal ζ [8], note that τ consists of M unit vectors.

The final objective of DTROD is to find a set of rank one tensors $\{\tau_r | r = 1, 2, \ldots, R\}$. In details, after obtaining the first $(r-1)$ rank one tensors, we replace $\mathbf{X}_i^c$ with its $(r-1)^{th}$ reconstruction error as defined in Eq.(2) and look for the next rank one tensor. For the optimization problem in Eq.(8), we adopt an alternative least square (ALS) approach. In ALS, we can obtain the optimal base vectors on one mode by remaining fixed the base vectors on the other modes and cycle for the remaining variables. The details of DTROD are listed in Algorithm 1.

Algorithm 1 Discriminant Tensor Rank-one Decomposition

Input: Tensor objects $\{\mathbf{X}_i^c\}$ from C classes, $\mathbf{X}_i^c$ denotes the i^{th} tensor object in the c^{th} class.
Initialize: $\mathbf{X}_{i,1}^c = \mathbf{X}_i^c$
for $r = 1$ to R **do**
  Randomly initialize $u_r^d \in \mathbb{R}^{I_d}$ for $d = 1, 2, \ldots, M$
  for $l = 1$ to T or *until converge* **do**
    for $d = 1$ to M **do**
      • $y_i^c = \mathbf{X}_{i,r}^c \prod_{s=1, s\neq d}^{M} \times_s u_r^s$.
      • Calculate S_b and S_w as that in traditional linear discriminant analysis, by taking y_i^c as the individual sample.
      • u_r^d = **SVD**$(S_b - \zeta S_w)$ with an optimal ζ.
    end for
  end for
  $\lambda_{i,r}^c = \mathbf{X}_{i,r-1}^c \prod_{d=1}^{M} \times_d u_r^d$
  $\mathbf{X}_{i,r+1}^c = \mathbf{X}_{i,r}^c - \lambda_{i,r}^c \prod_{d=1}^{M} \times_d u_r^d$
end for
Output: R rank one tensors $\tau_r = \prod_{d=1}^{M} \times_d u_r^d$, for $r = 1, 2, \ldots, R$.

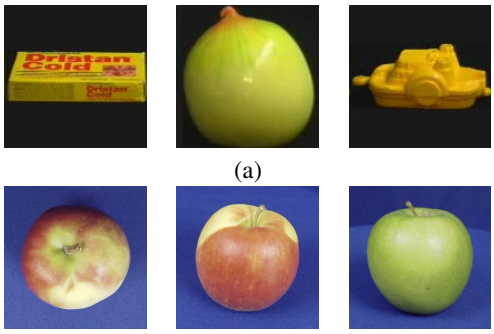
After obtaining the rank one tensors $\{\tau_r | r = 1, \ldots, R\}$ by DTROD, we can project each tensor object into these rank one tensors. For classification, the projection coefficients $\{\lambda_r | r = 1, \ldots, R\}$ defined in Eq.(3) can represent the extracted feature vectors and can be inputted into any other classification algorithm.

3. Experiment

3.1. Datasets

We use two datasets to validate DTROD for view-based object categorisation and recognition. The first is the Columbia COIL-100 image library [5]. It consists of color images of 72 different views of 100 objects. The images

were obtained by placing the objects on a turntable and taking a view every 5^o. The second dataset is the ETH Zurich CogVis ETH-80 dataset [3]. This dataset was setup by Leibe and Schiele [3] to explore the capabilities of different features for object class recognition. In this dataset, eight object categories including apple, pear, tomato, cow, dog, horse, cup and car have been collected. There are 10 different objects spanned large intra-class variance in each category. Each object has 41 images from viewpoints spaced equally over the upper viewing hemisphere. On the whole we have 3280 images, 41 images for each object and 10 object for each category. Fig.1 shows some sample images from these two datasets.



(a)

(b)

Figure 1. Some sample images. (a) Sample images from three different objects in COIL-100. (b) Sample images from the same category (apple) in ETH-80.

3.2. Experimental Setup

We take different experimental setup in these two datasets. For COIL-100, our objective is to discriminate between the 100 individual objects. For ETH-80, we aim to discriminate between the 8 object categories. In the previous experiments on object recognition using COIL-100, the number of views used as training set for each object varied from 36 to 4. When 36 views are used for training, the recognition rate using SVM was reported approaching 100% [7]. In practice, however, only very few views of an object are available. In our experiment, we used only 4 views of each object for training and the rest 68 views for testing. In total it is equivalent to 400 images for training and 6800 images for testing. The error rate is the overall error rate over 100 objects. To enable the tensor framework to capture enough variance on the change of viewpoint, we sampled 4 viewpoints evenly from the 72 viewpoints.

Previous experiments using ETH-80 dataset all adopted leave-one-object-out cross-validation. The training set consists of all views from 9 objects from each category. The testing set consists of all views from the remaining object from each category. In this setting, objects in the testing set have not appeared in the training set, but those belonging to the same category have. Classification of a test image is a process of labeling the image by one of the categories. Reported results are based on average error rate over all 80 possible test objects [3]. Similar to the above, instead of taking all possible views of each object in a training set, we take only 5 views of each object as training data. By doing so we have decreased the number of the training data to $1/8$ of that used in [3, 4]. The testing set consists of all the views of an object. In all our experiments, each training image is resized to 64×64 and represented as a $3rd$ tensor object using raw RGB data. The dimension of each tensor object is $64 \times 64 \times 3$.

3.3. Results

On each dataset, we carried out two kinds of experiments. First, we compare DTROD with traditional PCA+LDA by vectorizing the 3^{rd} tensor objects and using a nearest neighbor classifier. Second, we combine DTROD and AdaBoost. On COIL-100, the best results are obtained using non-linear SVM [9]. The results obtained using SNoW [13] is comparable to that of non-linear SVM. Our results are slightly better than both (see Table 1). On ETH-80, we compare our results with two recent results. Leibe and Schiele [3] compared the performance of different features for object categorization, they reported an error rate of 0.35 using only color feature. Marée et al. [4] reported an error rate of 0.26 using their random subwindows method. We achieved 0.24 error rate although we used only $1/8$ of the training data. These results are shown in Table 2.

Table 1. Overall recognition error rates on COIL-100. The results of DTROD and DTROD+AdaBoost are compared with that of PCA+LDA, SVM(linear&non-linear) [7, 9], Columbia 3D object recognition system [5] and SNoW [13].

Method	#training	#testing	error rate
PCA+LDA	400	6800	0.320
SVM(linear)	400	6800	0.215
SVM(non-linear)	400	6800	0.177
Columbia	400	6800	0.225
SNoW	400	6800	0.185
DTROD	400	6800	0.203
DTROD+AdaBoost	400	6800	0.155

In the experiment using nearest neighbor classifier, we used only the first 30 projection coefficients as features be-

Table 2. Overall recognition error rates on ETH-80, RSW denotes random subwindow method [4] and LS denotes the results from Leibe and Schiele [3].

Method	#training	#testing	error rate
PCA+LDA	360	328	0.37
DTROD	360	328	0.30
DTROD+AdaBoost	360	328	0.24
RSW	2952	328	0.26
LS	2952	328	0.35

cause we found there was no significant improvement in recognition rate when more features were used. Our results show that DTROD outperforms PCA+LDA. When we used DTROD for feature extraction and combined it with AdaBoost (DTROD+AdaBoost), the performance is comparable to other methods such as SVM and SNoW. It demonstrated particularly good performance on ETH-80 when taking a much small sample size (see a comparison in Table 2). The convergence property of DTROD is also demonstrated by Fig. 2. This figure shows the convergence curve of the difference between two consecutive iterations when calculate a single rank one tensor. Although the difference fluctuates in the beginning, it converges to a local minimum with the decreasing of the magnitude of fluctuation.

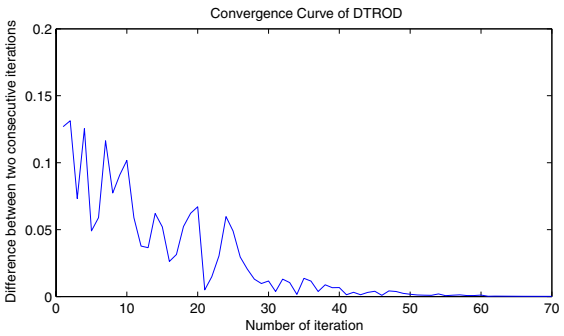


Figure 2. The convergence curve of calculation of a rank one tensor.

4. Conclusion

In this paper, we used a tensor discriminant analysis approach for view based object categorisation and recognition. The tensor approach avoids the need to vectorize 3D color images or 2D grayscale images into high dimensional feature vectors and has good property in convergence. Our experiment shows that discriminant analysis in multi-order tensor space outperforms LDA in a vectorised/flattened feature space for view-based object categorisation and recognition but the computation is more expensive. Using tensor discriminant analysis project coefficients as features for a AdaBoost based classification, we demonstrated comparable recognition results to the state of the art on the COIL-100 dataset and much better categorisation results on the ETH-80 dataset using only a much smaller number of training samples. Although the ETH-80 dataset contains big intra-class variance, the tensor discriminant analysis approach still achieved good results. However, we should also acknowledge that the tensor discriminant analysis approach for object categorisation and recognition shares the common drawbacks with other subspace approaches such LDA and PCA. They are not robust to illumination change, background clutter, occlusion and severe 3D pose change. These challenges still remain unsolved.

References

[1] D. Tao, X. Li, et al. Elapsed time in human gait recognition: A new approach. In *ICASSP*, 2006.

[2] S. Gong, S. McKenna, and J. Collins. An investigation into face pose distributions. In *AFGR*, pages 265–270, 1996.

[3] B. Leibe and B. Schiele. Analyzing appearance and contour based methods for object categorization. In *CVPR*, June 2003.

[4] R. Marée, P. Geurts, J. Piater, and L. Wehenkel. Random subwindows for robust image classification. In *CVPR 2005*, volume 1, pages 34–40, June 2005.

[5] H. Murase and S. K. Nayar. Visual learning and recognition of 3D objects from appearance. *International Journal of Computer Vision*, 14:5–24, 1995.

[6] T. Poggio and S. Edelman. A network that learns to recognize 3D objects. *Nature*, 343:263–266, 1990.

[7] M. Pontil and A. Verri. Support vector machines for 3D object recognition. *PAMI*, pages 637–646, 1998.

[8] Qingshan Liu, Xiaoou Tang, et al. Kernel scatter-difference based discriminant analysis for face recognition. In *ICPR*, Aug 2004.

[9] D. Roobaert and M. V. Hulle. View-based 3D object recognition with support vector machines. In *IEEE International Workshop on Neural Networks for Signal Processing*, 1999.

[10] A. Shashua and A. Levin. Linear image coding for regression and classification using the tensor-rank principle. In *CVPR (1)*, pages 42–49, 2001.

[11] D. Tao, X. Li, W. Hu, S. Maybank, and X. Wu. Supervised tensor learning. In *IEEE International Conference on Data Mining (ICDM)*, 2005.

[12] S. Yan, D. Xu, Q. Yang, L. Zhang, X. Tang, and H.-J. Zhang. Discriminant analysis with tensor representation. In *CVPR,2005*, volume 1, pages 526–532, 2005.

[13] M.-H. Yang, D. Roth, and N. Ahuja. Learning to recognize 3D objects with SNoW. In *Proceedings of the Sixth European Conference on Computer Vision*, pages 439–454, 2000.

[14] J. Ye, R. Janardan, and Q. Li. Two-dimensional linear discriminant analysis. In *Advances in Neural Information Processing Systems 17*, pages 1569–1576, Cambridge, MA, 2004.

0-7695-2521-0/06 $20.00