Benefits of Separable, Multilinear Discriminant Classification

Christian Bauckhage
Deutsche Telekom Laboratories
10587 Berlin, Germany
christian.bauckhage@telekom.de

Thomas Käster
Bielefeld University
33615 Bielefeld, Germany
tkaester@techfak.uni-bielefeld.de

Abstract

This paper presents an empirical investigation of the merits of tensor-based discriminant classification for visual object detection. First, we briefly discuss 2D separable discriminant analysis for grey value image analysis. Then, we contrast this tensorial approach with classical linear discriminant analysis. Our findings on a standard data set for object detection in natural environments show that, for the task of image analysis, tensor-based discriminant classifiers perform very robust. They learn and run faster and also generalize better than conventional techniques based on vectorial representations of the data.

1 Introduction

Traditionally, appearance based object recognition requires transforming image patches $\boldsymbol{\mathcal{X}}$ of spatial extension $m \times n$ into vectors $\mathbf{x} \in \mathbb{R}^{mn}$. Recent findings, however, suggest, that treating images for what they really are, namely *multiindexed objects* or *higher order tensors*, yields better results in tasks such as image coding or data reduction [7, 8, 9]. Moreover, several recent contributions indicate that multilinear algebra also offers substantial advantages for the classification of image data [2, 10]. In this paper, we quantitatively evaluate the benefits of tensor classifiers for view-based object detection.

We briefly describe an extension of linear discriminant analysis (LDA) that applies to multilinear objects of arbitrary orders. In contrast to other recent contributions, it does not make use of the technique of n-mode SVD. Instead, we repeatedly apply tensor contractions to a set of training samples and consider an alternating least squares procedure to obtain an R-term approximation of a rank deficient projection tensor for higher order discriminant analysis. As the rank deficiency constraint considerably reduces the number of free parameters, the resulting multilinear classifiers train rapidly. Furthermore, they are separable and thus also provide fast runtime. In addition to fast training- and runtime, the method generalizes well and provides good accuracy. Empirical results in multilinear object detection show that the tensor-based discriminant analysis performs reliably on complex, unconstrained natural scenes.

2 Theoretical Background

This section briefly summarizes an algorithm for fast training of tensor discriminant classifiers.

2.1 Mathematical Preliminaries

If the elements of a real-valued, nth-order tensor $\boldsymbol{\mathcal{U}}$ are denoted by $\mathcal{U}_{i_1 i_2 \ldots i_n}$, the *inner product* of two tensors $\boldsymbol{\mathcal{U}}$ and $\boldsymbol{\mathcal{V}}$ is defined as

$$\boldsymbol{\mathcal{U}} \cdot \boldsymbol{\mathcal{V}} = \sum_{i_1=1}^{m_1} \sum_{i_2=1}^{m_2} \ldots \sum_{i_n=1}^{m_n} \mathcal{U}_{i_1 i_2 \ldots i_n} \mathcal{V}_{i_1 i_2 \ldots i_n} \tag{1}$$

$$\equiv \mathcal{U}_{i_1 i_2 \ldots i_n} \mathcal{V}_{i_1 i_2 \ldots i_n}, \tag{2}$$

where, in (2), we made use of Einstein's summation convention. An nth-order tensor $\boldsymbol{\mathcal{U}}$ is a *decomposed* or a *rank 1 tensor*, if it can be written as

$$\boldsymbol{\mathcal{U}} = \mathbf{u}_1 \otimes \mathbf{u}_2 \otimes \ldots \otimes \mathbf{u}_n. \tag{3}$$

Here, $\otimes$ denotes the vector outer product and the $\mathbf{u}_j$ are vectors in $\mathbb{R}^{m_j}$. Note that for decomposed second-order tensors we may simply write

$$\boldsymbol{\mathcal{U}} = \mathbf{u}_1 \otimes \mathbf{u}_2 = \mathbf{u}_1 \mathbf{u}_2^{\mathrm{T}}. \tag{4}$$

Two decomposed tensors $\boldsymbol{\mathcal{U}}, \boldsymbol{\mathcal{V}}$ are *completely orthogonal*, if $\mathbf{u}_j \perp \mathbf{v}_j$ for all *modes* $j = 1, \ldots n$.

2.2 Tensor Discriminant Analysis

Traditional, two-class linear discriminant analysis as introduced by Fisher [4] deals with vectorial data. Given a training set of pairs $\{(\mathbf{x}^\alpha, y^\alpha)\}_{\alpha=1,\ldots,N}$, where $\mathbf{x}^\alpha \in \mathbb{R}^m$

0-7695-2521-0/06/$20.00 (c) 2006 IEEE

and $y^\alpha \in \{-1,+1\}$, LDA seeks a projection $\mathbf{w} \cdot \mathbf{x}^\alpha$ which maximizes the inter-class distance of the resulting scalars. Following Fisher's proposal, $\mathbf{w}$ results from solving the generalized eigenvalue problem $\mathbf{S}_b\mathbf{w} = \beta\mathbf{S}_w\mathbf{w}$, where $\mathbf{S}_b$ and $\mathbf{S}_w$ are the between-class and within-class scatter matrices of the data, respectively. Fisher himself pointed out that this may also be cast as a least squares regression problem

$$\mathbf{w} = \underset{\tilde{\mathbf{w}}}{\operatorname{argmin}} \|\mathbf{y} - \mathbf{X}\tilde{\mathbf{w}}\|^2 \tag{5}$$

where the $N \times m$ sample matrix $\mathbf{X}$ contains all sample vectors and the class labels are gathered in $\mathbf{y} \in \mathbb{R}^N$. This provides a closed form solution for the optimal projection direction:

$$\mathbf{w} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}. \tag{6}$$

Although the least squares approach to LDA generalizes to discriminant analysis of multilinear objects of arbitrary order, in this paper we focus on second-order tensors $\mathcal{U} \in \mathbb{R}^{m\times n}$ which can be used to represent grey-value images of size $m \times n$. Given a training set $\{(\mathcal{X}^\alpha, y^\alpha)\}_{\alpha=1,\dots,N}$, where the $\mathcal{X}^\alpha \in \mathbb{R}^{m\times n}$ are tensors sampled from two classes, tensor discriminant analysis requires a projection tensor $\mathcal{W}$ which solves the regression problem

$$\mathcal{W} = \underset{\tilde{\mathcal{W}}}{\operatorname{argmin}} \sum_\alpha \left(y^\alpha - \tilde{\mathcal{W}} \cdot \mathcal{X}^\alpha\right)^2. \tag{7}$$

If $\mathcal{W}$ is constrained to be a sum of R rank 1 tensors, i.e.

$$\mathcal{W} = \sum_{r=1}^{R} \mathbf{u}_r \mathbf{v}_r^{\mathrm{T}}, \tag{8}$$

solving (7) reduces to a series of simpler optimization problems within an alternating least squares scheme. The basic idea becomes apparent from the case $R = 1$ where we have

$$\mathbf{u}\mathbf{v}^{\mathrm{T}} \cdot \mathcal{X}^\alpha = (\mathbf{u}\mathbf{v}^{\mathrm{T}})_{kl}\mathcal{X}^\alpha_{kl} = u_k v_l \mathcal{X}^\alpha_{kl} = \mathbf{u}^{\mathrm{T}}\mathcal{X}^\alpha\mathbf{v}. \tag{9}$$

Therefore, given a random guess for $\mathbf{u}$, we may compute a set of vectors $\mathbf{x}^\alpha$ resulting from the tensor contractions $\mathcal{X}^\alpha_{kl}\ u_k$, stack them into a sample matrix $\mathbf{X}$ and use (6) to solve for $\mathbf{v}$. Given $\mathbf{v}$, the solution for $\mathbf{u}$ can be refined similarly and both steps are iterated until convergence.

Figure 1 displays the algorithm for deriving a projection tensor $\mathcal{W}$ with $R > 1$. The orthogonalization steps in this algorithm avoid redundancy and ensure that the resulting projection $\mathcal{W}$ favors directions of maximum variance in the data tensor space. Further details on alternating least squares for tensor classifiers can be found in [2].

Finally, note that discriminant analysis of a whole image is equivalent to a filter operation. In the case of classical LDA, the projection vector $\mathbf{w} \in \mathbb{R}^{mn}$ has to be treated as an $m \times n$ matrix (or second-order tensor) $\mathcal{W}$. Convolving an image $\mathcal{I}$ with $\mathcal{W}$ will result in a response map $\mathcal{Y} = \mathcal{I} * \mathcal{W}$, where entry $\mathcal{Y}_{ij}$ is the projection of image patch $\mathcal{X}_{ij}$ centered at image coordinate (i,j), i.e. $\mathcal{Y}_{ij} = \mathcal{W} \cdot \mathcal{X}_{ij}$.

Input: a training set $\{(\mathcal{X}^\alpha, y^\alpha)\}_{\alpha=1,\dots,N}$ of image patches $\mathcal{X}^\alpha \in \mathbb{R}^{m\times n}$ with class labels $y^\alpha \in \{-1,+1\}$
Output: a R-term solution of a second-order projection tensor $\mathcal{W} = \sum_r \mathbf{u}_r \otimes \mathbf{v}_r$

---

for $r = 1,\dots,R$
  $t = 0$
  randomly initialize $\mathbf{u}_r(t)$
  orthogonalize $\mathbf{u}_r(t)$ w.r.t. $\{\mathbf{u}_1,\dots,\mathbf{u}_{r-1}\}$
  repeat
    $t \leftarrow t+1$
    contract $x^\alpha_k = \mathcal{X}^\alpha_{kl}\ u_{r_k}(t)$
    compute $\mathbf{v}_r(t) = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}$
    orthogonalize $\mathbf{v}_r(t)$ w.r.t. $\{\mathbf{v}_1,\dots,\mathbf{v}_{r-1}\}$
    contract $x^\alpha_l = \mathcal{X}^\alpha_{kl}\ v_{r_l}(t)$
    compute $\mathbf{u}_r(t) = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}$
    orthogonalize $\mathbf{u}_r(t)$ w.r.t. $\{\mathbf{u}_1,\dots,\mathbf{u}_{r-1}\}$
  until $\|\mathbf{u}_r(t) - \mathbf{u}_r(t-1)\| \le \epsilon \ \vee\ t > t_{\max}$
endfor

Figure 1. Algorithm for computing a second-order tensor discriminant classifier $\mathcal{W}$.

3 Empirical Evaluation

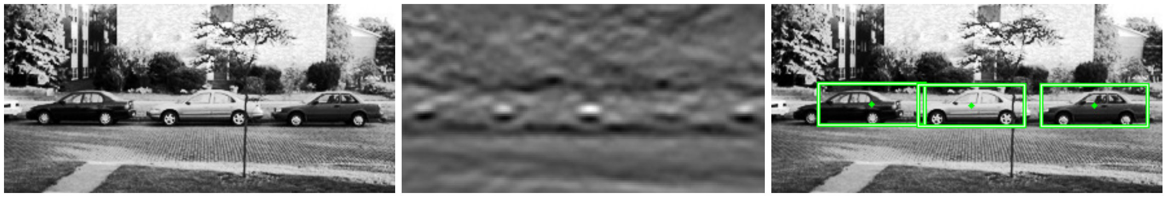
This section compares the performance of tensor-based discriminant analysis in visual object detection with that of classical linear discriminant analysis. For the latter, we consider both: the generalized eigenvalue approach and the least squares regression technique.

3.1 Experimental Setting

All experiments were carried out on a 3GHz Xeon PC and considered the UIUC database [1] of annotated side views of cars in natural environments. For each classifier, the set ω_p of positive training examples consisted of 124 images of cars of size 81×31. The set ω_n of negative examples consisted of 2442 patches randomly cut from the background of half of the images in the database; testing was done on 99 independent images of an average size of 120×116 pixels.

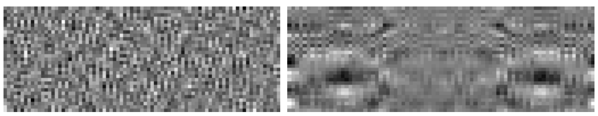
For training and in the application phase, the data was normalized to zero mean. This barely affects runtime. With $\hat{\mathcal{X}}$ denoting the mean of the training samples, we have

$$\left(\mathcal{X} - \hat{\mathcal{X}}\right) \cdot \mathcal{W} = \mathcal{X} \cdot \mathcal{W} - \hat{\mathcal{X}} \cdot \mathcal{W}. \tag{10}$$



(a) input image (b) smoothed response map (c) results after thresholding

Figure 2. Example of object detection using tensor-based discriminant analysis.

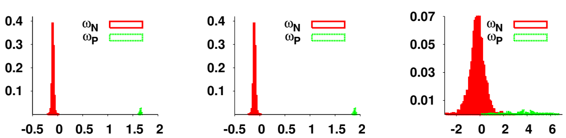


(a) $\mathbf{w} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y} \in \mathbb{R}^{mn}$ (b) $\boldsymbol{\mathcal{W}} = \sum_{r=1}^{R=6} \mathbf{u}_r \mathbf{v}_r^{\mathrm{T}} \in \mathbb{R}^{m \times n}$

Figure 3. 2D visualization of projection vector $\mathbf{w}$ and rank deficient projection tensor $\boldsymbol{\mathcal{W}}$.

method		t_{train}	t_{test}	EER
vector LDA	$\mathbf{S}_b\mathbf{w} = \beta\mathbf{S}_w\mathbf{w}$	260s	11s	60%
	$\mathbf{w} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}$	35s	11s	60%
tensor LDA	$\boldsymbol{\mathcal{W}} = \sum_{r=1}^{R=6} \mathbf{u}_r \mathbf{v}_r^{\mathrm{T}}$	5s	4s	88%

Table 1. Summary of quantitative results.



(a) $\mathbf{S}_b\mathbf{w} = \beta\mathbf{S}_w\mathbf{w}$ (b) $\mathbf{w} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}$ (c) $\boldsymbol{\mathcal{W}} = \sum_r \mathbf{u}_r \mathbf{v}_r^{\mathrm{T}}$

Figure 4. Projections of training samples onto corresponding discriminant directions.

Since the scalar constant $\hat{\boldsymbol{\mathcal{X}}} \cdot \boldsymbol{\mathcal{W}}$ can be computed beforehand, shifting the data to zero mean requires only a single operation per pixel.

The classifier response maps were smoothed (see Fig. 2) and subjected to a non-maximum suppression in order to reduce the number of false positives. A car was said to have been detected where the response exceeded a suitable threshold θ.

3.2 Results

Figure 3(a) depicts a matrix visualization of the projection vector $\mathbf{w}$ which we obtained from least squares regression training; Fig. 3(b) shows an $R = 6$ term projection tensor $\boldsymbol{\mathcal{W}}$. While there is no perceivable semantic structure in the visualization of $\mathbf{w}$, one dimly recognizes a car-like shape in the visualization of $\boldsymbol{\mathcal{W}}$.

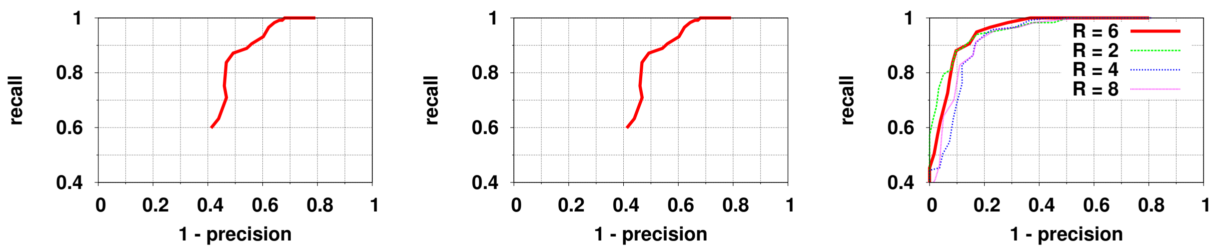
Figure 4 shows how the tested discriminant classifiers map the training samples onto the discriminant direction. While the vector-based approaches perfectly separate the positive and negative training samples, in the projection produced by the tensor-based predictor, there is an overlap. Figure 5 characterizes classifier performances by means of precision/recall curves resulting from varying θ between the centers of the modes of the projections in Fig. 4. Obviously, independent of the choice of R, the tensor-based approach outperforms the vectorial discriminant classifiers. Equal error rates as well as average training- and runtimes of the three tested methods are summarized in Tab. 1. Not only does the tensor-based technique yield the best performance, it also has the shortest training- time and best runtime.

Finally, it is interesting to see that multilinear object detection performs comparably to recent part-based methods. Table 2 lists equal error rates obtained on the UIUC database. Except for the technique treated here, all figures result from approaches that learn extensive lexicons of salient object parts and statistical part relations. Note that runtimes or training times of these approaches have not been reported.

3.3 Discussion

The major advantage of tensor discriminant classifiers appears to be the small number of free parameters. Because of this, multilinear classifiers train faster than usual LDA predictors. If multivariate data of size $m \times n$ are vectorized, conventional linear discriminant analysis requires inverting matrices of sizes $mn \times mn$. However, since the matrix inverses that appear in the tensor-based algorithm are of much smaller sizes $n \times n$ and $m \times m$, the multilinear technique significantly shortens training. In practice, we found that it reduces training times by up to two orders of magnitude.

Moreover, in the tensor-based approach, the convolution $\boldsymbol{\mathcal{I}} * \boldsymbol{\mathcal{W}}$ can be computed as a sequence of one-dimensional convolutions $\sum_r \big(\boldsymbol{\mathcal{I}} * \mathbf{u}_r\big) * \mathbf{v}_r^{\mathrm{T}}$. If $\boldsymbol{\mathcal{W}}$ is rank deficient, i.e. $R < \min\{m, n\}$, the effort reduces to $O(R(m+n))$ per pixel. In our experiments, this was an order of magnitude smaller than the per pixel effort of $O(mn)$ for the classical



(a) $\mathbf{S}_b\mathbf{w} = \beta\mathbf{S}_w\mathbf{w}$ (b) $\mathbf{w} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}$ (c) $\boldsymbol{\mathcal{W}} = \sum_{r=1}^{R} \mathbf{u}_r\mathbf{v}_r^{\mathrm{T}}$

Figure 5. Precision/recall curves obtained from varying the classification threshold θ.

method	LDA	Agarwal et al.[1]	Fergus et al. [3]	Garg et al. [5]	**tensor LDA**	Leibe et al. [6]
EER	60	77%	88%	88%	88%	97%

Table 2. Equal error rates obtained on the UIUC database of cars.

approaches. Also, the algorithm does not suffer from *small sample sizes*. In conventional LDA, the within-class scatter matrix may be singular because the number of training samples is much smaller than the dimension of the embedding space. This is why we considered fairly large numbers of training samples in our experiments. As the matrices required for training tensor-based classifiers are rather small, small sample sizes do not hamper tensor discriminant analysis. Due to lack of space, however, a detailed quantitative analysis will have to be reported elsewhere.

Finally, multilinear predictors with few parameters due to rank 1 constraints generalize better than their unconstrained vectorial counterparts. Although both training methods for classical linear discriminant classifiers yield a perfect separation of the training set, the mediocre performance on the test set (EER 60%) reveals that vector-based classifiers suffer from overfitting. The tensor-based classifier, in contrast, separates the training samples less well but yields a much higher equal error rate (88%) on the independent test set. From these figures and from the visualizations in Fig. 3, it appears that multiliner discriminant classifiers capture important visual structures contained in the training samples more faithfully than vector-based classifiers.

4 Summary

This paper presented a quantitative analysis of the benefits tensor-based discriminant classifiers offer for visual object detection. Given a standard database of objects in natural environments and a set of positive and negative examples, we found that R-term rank 1 tensor discriminant classifiers, which result from an alternating least squares algorithm for training, are characterized by very short training times. Also, compared to classical discriminant predictors, tensor discriminant classifiers are faster in runtime and yield considerably increased robustness and performance. With respect to future work, it appears that the short training times of the tensorial approach open up interesting perspectives for online learning in interactive scenarios.

References

[1] S. Agarwal, A. Awan, and D. Roth. Learning to Detect Objects in Images via a Sparse, Part-based Representation. *IEEE T. Pattern Anal. Machine Intell.*, 26(11):1475–1490, 2004.

[2] C. Bauckhage and J. Tsotsos. Separable Linear Discriminant Classification. In *Proc. DAGM*, volume 3663 of *LNCS*, pages 318–325. Springer, 2005.

[3] R. Fergus, P. Perona, and A. Zisserman. Object Class Recognition by Unsupervised Scale-Invariant Learning. In *Proc. CVPR*, volume II, pages 264–272, 2003.

[4] R. Fisher. The Use of Multiple Measurements in Taxonomic Problems. *Ann. Eugenics*, 7:179–188, 1936.

[5] A. Garg, S. .Agarwal, and T. Huang. Fusion of Global and Local Information for Object Detection. In *Proc. ICPR*, volume III, pages 723–727, 2002.

[6] B. Leibe and B. Schiele. Scale-Invariant Object Categorization using a Scale-Adaptive Mean-Shift Search. In *Proc. DAGM*, volume 3175 of *LNCS*, pages 145–153. Springer, 2004.

[7] A. Shashua and A. Levin. Linear Image Cosing for Regression and Classification using the Tensor-rank Principle. In *Proc. CVPR*, volume I, pages 42–40, 2001.

[8] M. Vasilescu and D. Terzopolos. Multilinear Analsysis of Image Ensembles: Tensorfaces. In *Proc. ECCV*, volume 2350 of *LNCS*, pages 447–460. Springer, 2002.

[9] H. Wang and N. Ahuja. Compact representation of multidimensional data using tensor rank-one decomposition. In *Proc. ICPR*, volume I, pages 44–47, 2004.

[10] S. Yan, D. Xu, L. Zhang, X. Tang, and H.-J. Zhang. Discriminant Analysis with Tensor Representation. In *Proc. CVPR*, volume I, pages 526–532, 2005.