Object Recognition with the Higher-Order Singular Value Decomposition of the Multi-Dimensional Prototype Tensors

Bogusław Cyganek

AGH University of Science and Technology,
Al. Mickiewicza 30, 30-059 Kraków, Poland
cyganek@agh.edu.pl

Abstract. In the paper an extension of object recognition based on the Higher-Order Singular Value Decomposition (HOSVD) to the 4th dimension is discussed. HOSVD based object recognition expands the concept of object recognition in the pattern spaces spanned by the PCA decomposition of vector patterns into the higher dimensions. However, contrary to the PCA, in the HOSVD the bases of the pattern space are tensors rather than 1D vectors. Nevertheless, the already presented works on HOSVD recognition were limited to the images with only scalar valued pixels. In the proposed framework images are allowed to contain multi-dimensional pixels, which adds an additional dimension to the pattern tensor. The proposed method opens new possibility of the HOSVD based recognition to color or other multi-valued images. Experimental results show improved accuracy as compared to the scalar valued data, as well as fast execution time.

1 Introduction

Recently, tensor based methods found great interest in pattern recognition domain. In computer vision these were also shown to provide excellent results in object recognition [1][16][18][5]. Their success lie in the fact that tensor based methods explicitly account for multidimensional nature of processed data.

In this paper an extension of object recognition based on the Higher-Order Singular Value Decomposition (HOSVD) to the 4th dimension is presented. HOSVD method of object recognition exploits the concept of object recognition in the pattern spaces. The best known method in this category is the PCA decomposition. However, PCA operates with vector-like data. Also, the bases of the spaces spanned by PCA are vectors. However, contrary to the PCA in the HOSVD the bases of the pattern spaces are tensors rather than 1D vectors.

The HOSVD classifier shows good results when applied to multi dimensional data, such as images [16][5]. This is due to tensor processing which allows separate control of all intrinsic dimensions of data. Let us recall that in the classical PCA-based classification method, images are first vectorized and, in the result, the obtained subspaces are spanned by vector bases [17]. However, this also leads to the lost of information on spatial relations among pixels. Contrary to this, in the HOSVD method the bases of the orthogonal pattern subspace are spanned by the higher-dimensional tensors. Nevertheless, to the best of our knowledge, the reported works on HOSVD recognition were always limited to the images with only scalar valued pixels [16][3]. In the case of scalar valued images, the HOSVD bases are two-dimensional. Contrary to these, in the proposed framework images are allowed to contain multi-dimensional pixels, which adds an additional dimension to the pattern tensor. However, thanks to this, the proposed method opens new possibility for the HOSVD based recognition of color or any-length pixel images. In other words, due to the proposed extension to multi-valued pixels, the base tensors are three-dimensional, as will be discussed. When classifying an unknown pattern, the tested patters are projected onto the subspaces of each of the trained classes and the best fitting projection is returned. However, in the tensor case the bases are multidimensional, as already mentioned. The proposed method can be also extended to higher dimensions, leading to the 5th, 6th, and higher dimensional bases, depending on a type of the input signals.

The method was tested on the problem of face recognition in the difficult set of color face images. Experimental results show improved accuracy as compared to only scalar valued, i.e. gray-valued, images. The proposed method can be compared to the methods reported by other researchers [8][14][10], although it was not optimized particularly for the face recognition problem.

Apart from the above, in the proposed system a parallel version of the HOSVD algorithm is applied. Concurrency is obtained through the functional and data decompositions on different levels of computations. Parallel operation is also possible at the response time of the system, since each subspace projection can be computed independently.

The paper is organized as follows. In the next section the tensor based pattern recognition framework is presented. Experimental results are presented and discussed in Section 3. In this section also implementation details are provided. Conclusions are presented in Section 4 of this paper.

2 Multi-Valued Image Recognition in the Tensor Subspaces

Tensors in data mining can be interpreted as multidimensional data-cubes. Processing and analysis of multi-dimensional data, such as images, fits well into this framework. An example of image representation in a tensor form is presented in Fig. 1. However, an analysis of data content requires proper decomposition of pattern tensors. In this respect, HOSVD is one of the most powerful tensor decomposition methods [1][11][12][13][5]. As shown, HOSVD can be used to build orthogonal spaces which can be then used for pattern recognition in a way similar to the subspace projection methods [6][17]. This procedure is briefly outlined in this section. More information on tensors in signal processing can be found in literature, e.g. [1][11][12][13][5].

Let us briefly present the underlying theory behind multi-dimensional data representation and analysis by means of tensors and their decompositions. In this respect, the first concept is the *k-mode vector* of a *P-th* order tensor $\mathcal{T} \in \Re^{N_1 \times N_2 \times \ldots N_P}$. It is a vector obtained from the elements of $\mathcal{T}$ by changing only one index n_k, and keeping all other fixed. The second important concept is the operation of the k-mode flattening of a tensor. For a tensor $\mathcal{T}$, a result of its k-mode flattening is the following matrix [12][11].

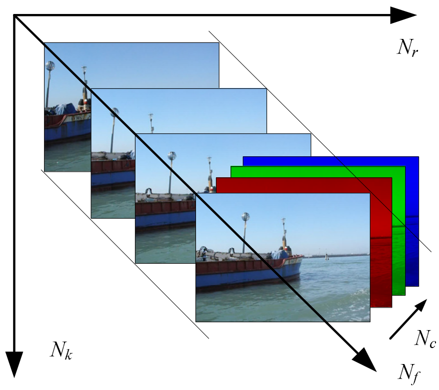

Fig. 1. A series of color images represented as a four-dimensional data cube. This can be seen as a 4-dimensional tensor.

$$\mathbf{T}_{(k)} \in \Re^{N_k \times \left(N_1 N_2 \ldots N_{k-1} N_{k+1} \ldots N_P\right)}. \tag{1}$$

Now we can define the HOSVD decomposition for pattern tensors constructed of a series of 3D images, that is, each having two spatial and one pixel-value coordinates. Therefore our pattern tensors will be four-dimensional (4D). Thus, the further discussion is confined to the 4D tensors.

As already mentioned, important information about the pattern space are revealed after its HOSVD decomposition. That is, any 4D tensor can be represented as the following tensor product [12][11]

$$\mathcal{T} = \mathcal{Z} \times_1 \mathbf{S}_1 \times_2 \mathbf{S}_2 \times_3 \mathbf{S}_3 \times_4 \mathbf{S}_4 . \tag{2}$$

In the above formula $\mathbf{S}_k$ are *unitary* matrices of dimensions $N_k \times N_k$, (called mode matrices), and $\times_j$ denotes the so called j-mode product of a tensor and a matrix. The mode matrices $\mathbf{S}_k$ are responsible for representation of column spaces related to each different index (dimension) of a tensor. On the other hand, the tensor $\mathcal{Z} \in \Re^{N_1 \times N_2 \times N_3 \times N_4}$ is called a core tensor, and fulfills properties of the sub-tensor orthogonality and decreasing energy value [12][11].

De Lathauwer proposed a method of computation of the HOSVD which is based on successive application of the matrix SVD decompositions to the flattened matrices of a given tensor [12]. The HOSVD decomposition algorithm for a 4-dimensional tensor $\mathcal{T}$ is outlined in Fig. 3. It can be easily observed that computation of the HOSVD requires a series of computations of the SVD decompositions on the flattened tensor representations (i.e. matrices). These are independent versions (different modality) of the input tensor. Therefore it is possible to run all these SVD decompositions concurrently, which must be synchronized on a barrier just before computation of the core tensor in (8), however. Figure 3 shows the algorithm for computation of the HOSVD. Its grayed area can be run concurrently, as discussed.

Let us now observe that, thanks to the commutative properties of the k-mode multiplication, for each mode matrix $\mathbf{S}_i$ in (2) the following sum can be constructed

$$\mathcal{T} = \sum_{h=1}^{N_P} \mathcal{T}_h \times_4 \mathbf{s}_4^h \ . \tag{3}$$

Further, it can be shown that tensors

$$\mathcal{T}_h = \mathcal{Z} \times_1 \mathbf{S}_1 \times_2 \mathbf{S}_2 \times_3 \mathbf{S}_3 \tag{4}$$

in (3) constitute the basis tensors and $\mathbf{s}^h{}_P$ are columns of the unitary matrix $\mathbf{S}_P$ [12][11]. Thus, they form an orthogonal basis which spans a subspace. This property is used to construct a HOSVD based classifier [16][3]. However, in this case they are *3D tensors*, as shown in Fig. 2. This constitutes *a novelty* of the proposed method.

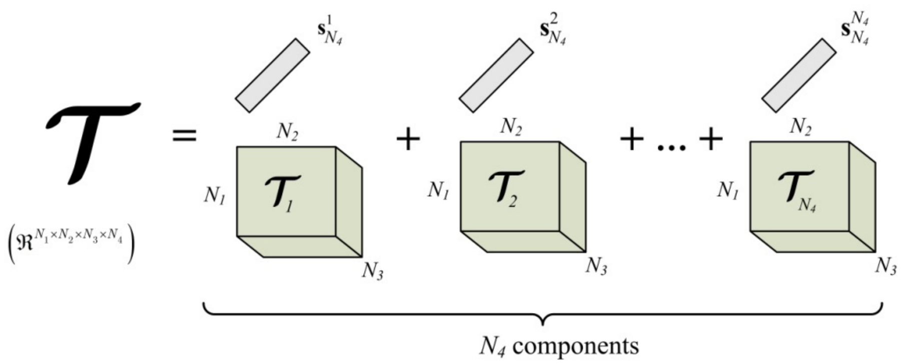

Fig. 2. Decomposition of the pattern tensor into a sum of products of the 3D base tensors and mode vectors. The base tensors form an orthonormal subspace used for pattern recognition.

In each subspace spanned by tensors $\mathcal{T}_h$, object recognition can be formulated as a testing of a distance of a given test pattern $\mathbf{P}_x$ to its projections in each of the spaces spanned by the set of the bases $\mathcal{T}_h$ in (4). That is, the following optimization process needs to be solved [16]:

$$\min_{i,c_h^i} \underbrace{\left\| \mathbf{P}_x - \sum_{h=1}^{K} c_h^i \boldsymbol{\mathcal{T}}_h^i \right\|^2}_{Q_i}, \qquad \textbf{(5)}$$

where the scalars c^i_h denote unknown coordinates of the pattern $\mathbf{P}_x$ in the space spanned by $\mathcal{T}_h^i$, and $K \leq N_P$ denotes a number of chosen dominating components. It can be further shown that to minimize (5) we need to maximize the following value [16][5]

$$\hat{\rho}_i = \sum_{h=1}^{K} \left\langle \hat{\boldsymbol{\mathcal{T}}}_h^i, \hat{P}_x \right\rangle^2, \qquad (6)$$

where $\left\langle \hat{\boldsymbol{\mathcal{T}}}_h^i, \hat{P}_x \right\rangle$ denotes the inner product operation. In other words, the (single) HOSVD based classifier returns a class i for which its ρ_i from (6) is the largest. It is worth recalling that in our framework the base tensors $\mathcal{T}_\mathrm{h}$ are 3D. However, in the response time, computation of the inner product in accordance with (6) is very fast.

The main difference of the tensor based approach to building the spanning pattern subspaces thus lies in 4-times computed column space, whereas in the PCA method this is computed once on a vectorized data, no matter what dimensionality they had originally.

begin

> **for** each k=1, …, 4 **do**
>
> **1.** From Eq.(1) compute k-mode flattened matrix $\mathbf{T}_k$ of tensor $\boldsymbol{\mathcal{T}}$
>
> **2.** Compute $\mathbf{S}_k$ from the SVD decomposition of $\mathbf{T}_k$
>
> $$\mathbf{T}_k = \mathbf{S}_k \mathbf{V}_k \mathbf{D}_k^T \qquad (7)$$
>
> **end**

Compute the core tensor from all matrices $\mathbf{S}_k$

$$\boldsymbol{\mathcal{Z}} = \boldsymbol{\mathcal{T}} \times_1 \mathbf{S}_1^T \times_2 \mathbf{S}_2^T \times_3 \mathbf{S}_3^T \times_4 \mathbf{S}_4^T \qquad (8)$$

end

Fig. 3. Algorithm for computation of the Higher-Order Singular Value Decomposition of tensors of 4th dimensions. The shaded steps can be executed concurrently.

More details on implementation of the HOSVD decomposition can be found in [5]. Fig. 3 contains pseudo-code of the four-dimensional HOSVD decomposition. The grayed area represents the part of the algorithm which can be run concurrently. This can lead to the computation speed-up.

3 Experimental Results

The presented method was implemented in C++, supported by the *DeRecLib* software from [5] and the *OpenMP* library for the multicore processing [2][15]. The experiments were carried out on the computer with 8 GB RAM and the Pentium® Quad Core Q 820 microprocessor (eight cores due to the hyper-threading technology [9]).

In order to evaluate the method a database with multi-valued features is required. For this purpose the Georgia Tech Face Database (GTFD) [7] was employed which contains color images. Images of persons in the GTFD are acquired in different sessions, various poses and illuminations. Some of the photographed persons in some sessions wear glasses, as well as many persons were photographed from different viewpoints. Therefore, this database is known as highly demanding for the face recognition algorithms [8][14]. It contains images of 50 persons taken at multiple sessions. There are 750 images with 15 images per person. The images for each person contain the frontal pose, as well as different facial expressions, various illuminations, and scale. Exemplary faces from this database are shown in Fig. 4. However, contrary to other works in our experiments the images are not preprocessed.

Thus, for each 15 available exemplars, the experiment was carried out always randomly taking 12 images of a person for training, and then testing on the remaining 3 images. Such tests were run ten times and the average results are reported in Tab. 1.

Fig. 4. Examples of the test images from the Georgia Tech Face Database [7]. There are 50 subjects, for each there are 15 images from which 10 were randomly selected for training and the remaining 5 for testing in different runs of the system.

Fig. 5 and Fig 6 depict slices of the 3D base tensors $\mathcal{T}_h$ computed in accordance with the formula (4) for two subjects shown in Fig. 4.

Fig. 5. Slices of the three base tensors of the second subject shown in Fig. 4.

In our experiments parallelism on different levels of computations were measured. Also, each parallel realization was analyzed in the context of memory requirements. The parallel implementation allows up to two times speed-up in computations, as compared to a serial version.

Tab. 1. Average accuracy of face recognition with the multi-dimensional 3D and 4D HOSVD based classifier (first column). Accuracy measured with a condition on best match separation of at least 1% (second column).

Experiment conditions	Accuracy (a) [%]	Accuracy (b) [%]
3D tensors (scalar)	83.4	87.2
4D tensors (multi-valued)	87.3	91.9

To verify our assumptions the experiments were performed the same number of times for the monochrome, as well as color versions of the same images. Results show that utilization of color information, in the form of a 4th dimension of the input pattern tensor, leads to better accuracy. The number of components used in (6) was 7 in all experiments. Lower values led to slightly smaller accuracy, although even for 3 first components the differences do not exceed 1% in overall accuracy. On the other hand, higher values resulted in no higher accuracy, requiring more computations at the same time.

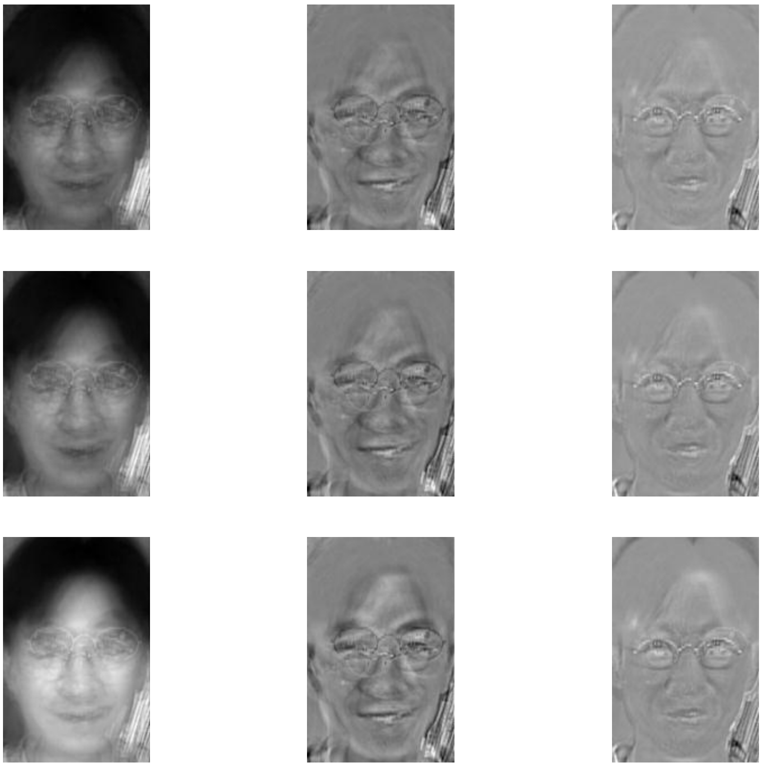

Fig 6. Slices of the three base tensors of the fourth subject shown in Fig. 4.

It is worth noticing the difference of the proposed HOSVD based method compared with the PCA approach. The proposed methods works better since multi-dimensional data (color faces in our case) are decomposed independently in each dimension (four in our experiments), whereas PCA does decomposition only in one dimension regardless of the internal dimensionality of data. Thus, with the HOSVD a more in-depth information of the contained patterns is extracted which leads to higher accuracy. However, we would like to emphasize that the proposed method is not the best face recognition algorithm. Especially problematic is recognition of multi class patterns, containing dozens of classes. In this case the variability between classes can be even lower than within a single class which leads to lowered accuracy. To remedy the situation we added an additional constraint on the best match value, as well as the second best match (i.e. the value in formula (6)). More precisely, the following condition is checked

$$1-\frac{\hat{\rho}_{2nd}}{\hat{\rho}_{1st}} > \tau \ , \tag{9}$$

where $\hat{\rho}_{1st}, \hat{\rho}_{2nd}$ denote the 1st largest and the 2nd largest value of the residuum computed in accordance with (6), respectively, and τ denotes a threshold value. In our experiments the latter was set to 1%. Application of (9) allowed an increase of accuracy at a cost of some missing recognitions (false negatives), as shown in Tab. 1.

However, our purpose was to show the difference between the HOSVD operating in different dimensions. That is, a difference between the 3D and 4D pattern tensors. Our experimental results show that operations in the higher dimensional space lead to better results, at a negligible additional computations in the response stage allowing real-time operations.

4 Conclusions

In this paper a new version of the HOSVD based tensor classifier is presented. This is an extension of the highly successful HOSVD classifier to the 4th dimension, representing multi-valued pixels of the input images. Thanks to this, the input prototypes can contain other than scalar values. Thus, the proposed method allows recognition of color or other multi-valued signals. Experimental results in color face recognition show improved accuracy as compared to the scalar valued representations. Summarizing, the key features of the presented method are as follows:

- The method achieves high accuracy.
- The proposed pattern recognition method can be used to any patterns (not only images).
- The method can be easily extended to higher dimensional "cubes" of data (such as video, hyperspectral, etc.).
- The parallel algorithms for training and testing were outlined.
- The method allows real-time operation even in software implementation (simple inner product computation).

Nevertheless it is in order to mention some problems associated with the proposed method. First, size of the input tensor very frequently is too high to fit into the memory. This also concerns time necessary for the HOSVD decomposition. Therefore, future research is to develop methods which allow partial computation of the HOSVD. The second problem is threshold necessary to distinguish the in-class from the ext-class patterns. In the presented experiments this is achieved by comparison with the external face classes. However, in some practical situations such ext-class examples are not available. Further research will focus also on testing the proposed method with different datasets, application of image transformations, such as computation of the extended structural tensor, as well as further extension to classification of the video patterns, i.e. processing of the 5th order pattern tensors. Also interesting is application of the presented method to data other than images.

Acknowledgement

The work was supported in the years 2013-2014 from the funds of the Polish National Science Centre NCN, contract no. DEC-2011/01/B/ST6/01994.

References

1. Cichocki, A., Zdunek, R., Amari, S.. Nonnegative Matrix and Tensor Factorization. IEEE Signal Processing Magazine, Vol. 25, No. 1, 142-145 (2009)
2. Chapman, B., Jost, G., Van Der Pas, A.R. Using OpenMP. Portable Shared Memory Parallel Programming. MIT Press (2008)
3. Cyganek, B. An Analysis of the Road Signs Classification Based on the Higher-Order Singular Value Decomposition of the Deformable Pattern Tensors, Lecture Notes in Computer Science, Vol. 6475, Springer, 191–202 (2010)
4. Cyganek, B. Adding Parallelism to The Hybrid Image Processing Library in Multi-Threading and Multi-Core Systems. IEEE International Conference on Networked Embedded Systems for Enterprise Applications, 103–110 (2011)
5. Cyganek, B. Object Detection in Digital Images. Theory and Practice, Wiley (2013)
6. Duda, R.O., Hart, P.E., Stork, D.G. Pattern Classification. Wiley (2000)
7. Georgia Tech Face Database. *http://www.anefian.com/research/face_reco.htm* (2013)
8. Goel, N., George Bebis and Ara V. Nefian. Face recognition experiments with random projections, SPIE Conference on Biometric Technology for Human Identification (2005)
9. Intel. *www.intel.com* (2013)
10. Jiang, X., Mandal, B., Kot, A. C-C.: Eigenfeature Regularization and Extraction in Face Recognition. IEEE Trans. Pattern Anal. Mach. Intell. 30(3): 383-394 (2008)
11. Kolda, T.G., Bader, B.W. Tensor Decompositions and Applications. SIAM Review, Vol. 51, No. 3, 455-500 (2008)
12. Lathauwer de, L. Signal Processing Based on Multilinear Algebra. PhD dissertation, Katholieke Universiteit Leuve (1997)
13. Lathauwer de, L., Moor de, B., Vandewalle, J. A Multilinear Singular Value Decomposition. SIAM Journal of Matrix Analysis and Applications, Vol. 21, No. 4, 1253-1278 (2000)
14. Nefian, A. V.: Embedded Bayesian networks for face recognition, IEEE International Conference on Multimedia and Expo, August (2002)
15. OpenMP. *www.openmp.org* (2013)
16. Savas, B., Eldén, L. Handwritten Digit Classification Using Higher Order Singular Value. Pattern Recognition, Vol. 40, No. 3, 993-1003 (2007)
17. Turk, M., Pentland, A. Eigenfaces for Recognition. Journal of Cognitive Neuroscience Vol. 3, No. 1, 71–86 (1991)
18. Vasilescu, M.A.O., Terzopoulos, D. Multilinear Analysis of Image Ensembles: TensorFaces, Lecture Notes in Computer Science, Vol. 2350, Springer, 447-460 (2002)