Elsevier Editorial System(tm) for Engineering Applications of Artificial Intelligence
Manuscript Draft

Manuscript Number:

Title: Multidimensional Data Classification with Chordal Distance Based Kernel and Support Vector Machines

Article Type: Contributed Paper

Keywords: machine learning; pattern recognition; support vector machine; chordal distance; tensor decomposition

Corresponding Author: Dr. Boguslaw Cyganek, PhD, DSc

Corresponding Author's Institution: AGH University of Science and Technology

First Author: Boguslaw Cyganek, PhD, DSc

Order of Authors: Boguslaw Cyganek, PhD, DSc; Bartosz Krawczyk; Michal Wozniak, PhD,DSc

Abstract: Proposing novel methods for tackling complex and multidimensional data is a focus of current machine learning research. The problem of representation of such data, in order to find a trade-off between easy processing and maintaining discriminative power is one of the crucial issue. In this paper, we propose a new method for efficient classification of multidimensional data based on tensor-based kernel applied in Support Vector Machines. We represent data as tensors, in order to preserve the spatial dependencies among the data and allow to process complex structures (such as color images or video sequences) as single objects. To allow for an effective classification, we augment a Support Vector Machine trained with Sequential Minimal Optimization procedure with a chordal distance-based kernel for efficient computation of tensor-based objects. We present full implementation details, required to use such a kernel in practice. The proposed method is evaluated in both binary and multi-class classification problems. Comprehensive experimental analysis carried on face recognition problem, shows the high usefulness of the proposed approach.

LaTeX Source Files

Click here to download LaTeX Source Files: final.zip

Multidimensional Data Classification with Chordal Distance Based Kernel and Support Vector Machines

Bogusław Cyganek[a,*], Bartosz Krawczyk[b], Michał Woźniak[b]

[a] *AGH University of Science and Technology, Al. Mickiewicza 30, 30-059 Krakw, Poland*
[b] *Department of Systems and Computer Networks, Wrocław University of Technology, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland*

Abstract

Proposing novel methods for tackling complex and multidimensional data is a focus of current machine learning research. The problem of representation of such data, in order to find a trade-off between easy processing and maintaining discriminative power is one of the crucial issue. In this paper, we propose a new method for efficient classification of multidimensional data based on tensor-based kernel applied in Support Vector Machines. We represent data as tensors, in order to preserve the spatial dependencies among the data and allow to process complex structures (such as color images or video sequences) as single objects. To allow for an effective classification, we augment a Support Vector Machine trained with Sequential Minimal Optimization procedure with a chordal distance-based kernel for efficient computation of tensor-based objects. We present full implementation details, required to

*Corresponding author

Email addresses: cyganek@agh.edu.pl (Bogusław Cyganek), bartosz.krawczyk@pwr.wroc.pl (Bartosz Krawczyk), michal.wozniak@pwr.wroc.pl (Michał Woźniak)

use such a kernel in practice. The proposed method is evaluated in both binary and multi-class classification problems. Comprehensive experimental analysis carried on face recognition problem, shows the high usefulness of the proposed approach.
Keywords: machine learning, pattern recognition, support vector machine, chordal distance, tensor decomposition.

---

1. Introduction

Majority of the classical pattern recognition methods relies on vector spaces [13]. This reflects basic properties of simple measurements which stack different feature values of measurements into one-dimensional (1D) vectors, which are then assigned to predefined classes. However, many phenomena lead to measurements, which change specifically depending on a chosen dimension or a coordinate. Well known examples are video sequences, which are composed of two-dimensional frames containing three-valued pixels, displayed a number of times per second. Naturally, they are four-dimensional data which become even five-dimensional considering sound. Such examples arise in many domains when measuring signals under different settings of an experiment. Such data is called multidimensional or tensor like signals [10, 25, 26]. However, they do not fit well into the classical 1D vector based framework. Although there are many ways to vectorize multidimensional data, it has been observed that such operation usually leads to significant loss of important information, since some values which were close (in terms of a chosen coordinate) become differently separated if data are arbitrarily linearized into a vector. Therefore in recent years much attention gained development of pattern recognition methods which inherently consider multidimensionality of the classified data [8, 37, 40, 41]. On the other hand, one of the newest and highly appreciated achievements in pattern recognition of recent two decades are Support Vector Machines (SVMs) which, operating on vector spaces, successively classify different types of data with support of the kernel functions [7, 13].

In this paper we analyze properties of the SVMs employed to a multidimensional data classification task. More specifically, we consider classification of the monochrome and color images, directly treated as 2D and 3D tensor respectively, and with help of the recently proposed chordal kernel for tensor data [37]. The kernel is employed to different versions of SVMs trained with the Sequential Minimal Optimization (SMO) algorithm [35]. The purpose of this work is to verify usefulness of such approach to the common image classification problem, as well as to scrutinize its basic properties and implementation issues. To the best of our knowledge this is a first research that directly shows properties of the chordal tensor and SMO trained SVMs. Also, the obtained results lead us to the important conclusions on favor of such approach which we believe can be successively employed to many systems requiring image classification with significant improvement to the classification accuracy. We also analyze computational costs of this new approach.

The main contributions of this work are as follows:

- Proposal of efficient tool for handling complex and multidimensional data (e.g., color images, hyperspectral images or video sequences) as tensors, based on SVM with chordal distance-based kernel and trained with SMO algorithm.

- Detailed information and solutions to the implementation issues considered with efficient running of the described tensor kernel.

- Comprehensive experimental evaluation and statistical assessment of the proposed system, carried out on face recognition datasets.

The rest of this paper is organized as follows. The next Section describes state-of-the-art connected with SVMs and tensors for image processing and classification. Then, the used chordal distance-based kernel, together with all of the required details for a highly efficient implementation are described. In Section 4 the experimental study is presented, while the last section concludes the paper.

2. Related works

Let's briefly review recent works which were influential to our work. Kernel design is still an active research topic. However, there is a gap between the kernel methods and multidimensional data which stems from the fact that kernel functions in their basic definition accept two vector arguments, whereas tensor data frequently are not identical with vectors. Thus, a common practice when classifying objects with SVM is to vectorize image patterns before applying them to a classifier. However, such strategy leads to worse classification results compared to methods which directly account for multi-dimensionality of the input data [33, 37, 40, 41].

In [26] de Lathauwer introduced tensors to the signal processing domain.

He proposed the higher order singular value (HOSVD) decomposition of tensors and analyzed the chosen tensor's properties.

A tensor approach for recognition of human faces from a series of images with exemplary faces under different view, pose and illumination, was originally proposed by Vasilescu and Terzopoulos [40]. Their method, called tensorfaces, relies on HOSVD decomposition of a data tensor containing all the prototype face patterns. This decomposition allows classification of an unknown face image to the best fit person class, as well as a synthesis of a view of any person under new illumination or pose conditions [41].

The problem of tensor factorization for face recognition was addressed in many works. Tenenbaum and Freeman proposed a bilinear model which allows face analysis in the context of two unknown factors, such as people and views, peoples and expressions, or peoples and illumination conditions [38]. Similarly, Lin et al. [29] proposed an alternating update and consists in changing one factor while the other is kept fixed. Peng and Qian proposed a method for online gesture spotting from visual hull data in which pose features are extracted using the HOSVD decomposition and the alternating least-squares algorithm [34]. This method was also applied to an individual subspace modeling of pattern faces belonging to a different person is proposed [33]. Such approach leads to reduction of the number of factors which need to be determined in the prototype tensor factorization to three factors, i.e., to the face, expression and illumination. On the other hand, the factorization problem is stated as the least squares problem for the Kronecker product of all of the unknown parameters with the quadratic equality constraint. Park and Savvides proposed also kernel version of their factorization method. It was shown that mapping of the patterns into a higher dimensional feature space by a kernel function leads to higher accuracy at an acceptable computational cost, mostly due to the well known kernel trick [33].

An introduction of the kernel-based approach to the multi-factor analysis for image synthesis and recognition was proposed in [27]. This method relies on kernel-based factorization with the HOSVD method and it is suitable for new image synthesis and underlying factor estimation. Further overview of tensors in pattern recognition can be found in the book by Cyganek [8].

SVM belong to the newest and most effective classifiers, originally published by Cortes and Vapnik [7]. The most original properties of the SVM is a construction of a maximal separable classifier between two classes, as well as transformation of data into the higher dimensional feature space in which data separation can be done with a linear hyper-surface. Since then SVMs gain much attention due to their superior accuracy and relatively easy parameter tuning. Due to these properties, SVMs found broad applications in image classification which are the simplest examples of multidimensional signals. There are plethora of examples of their successful applications in this domain. Ko and Byun proposed combination of SVM classifiers for multi-class problem with application to face recognition [24]. The authors analyzed the one-per-class, as well as pairwise coupling output combiners method. However, the obtained accuracies do not exceed 93% tested on the AT&T Database of Faces (formerly ORL database) [1].

Wang et al. [42] analyzed application of the SVM endowed with the

[1] `http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html`

error correction mechanism to face recognition task. In their approach both, the PCA and Fisherface features were used as input vectors. The authors advocate for good abilities of the the error-correcting output codes (ECOC) [11], which improved overall accuracy of the system. The obtained results on the ORL face database reach up to 97%. Shavers et al. [36] presented a face detection system based on SVM with the polynomial kernels. A method for face image gender recognition based on the Gabor transform and SVM was proposed by Yan [43]. In this work the 2D Gabor features are fed to the SVM classifier with the linear, polynomial, as well as RBF and sigmoid kernels. The reported accuracy of the final gender response is in the range of up to 83%. SVM classifiers are also employed in the face recognition system based on the curvelet features, proposed by Mandal et al. [30]. In their paper the curvelet transform is reported to produce better results than the well known wavelet transforms. The experiments were performed on the Georgia Tech Face Database[2], as well as on the AT&T Database of Faces. Performance of the multi-class SVM for face recognition is also reported in the work by Lihan et al. [28]. In their system, PCA is first used for dimensionality reduction and feature extraction, then, the one-versus-all SVMs were trained. The obtained accuracies are 93.5% for the ORL, and 97.3% for the Yale face databases[3], respectively.

Valuvanathorn et al. proposed a multi-feature face recognition based on PSO-SVM [39]. In this system the global and local features are investigated for the face recognition. The features are obtained based on histograms,

[2] `http://www.anefian.com/research/face_reco.htm`

[3] `http://vision.ucsd.edu/content/yale-face-database`

PCA, as well as two-dimensional PCA (2D-PCA) techniques. For classification the SVM classifier was joined with the particle swarm optimizatio (PSO) to automatically determine classification parameters. As reported, the 2D-PCA provides the best results, for which the obtained accuracies of face recognition reached 95.6%.

The face recognition problem, discussed in this paper, belongs to the difficult classification tasks for which many methods have been proposed, as Jiang et al. [22] proposed a subspace approach which regularizes and extracts co called eigenfeatures. Face recognition with help of the random projections was described by Goel et al. [17].In the work by Nefian [32] the embedded Bayesian network (EBN), which is a generalization of the embedded hidden Markov model, was proposed to the face recognition task. Nevertheless, as we will show later the method presented in this paper outperforms majority of the results reported in the aforementioned experiments.

A method of decomposing the tensor kernel support vector machine for neuroscience data with structured labels is discussed in the paper by Hardoon and Shawe-Taylor [20]. In their approach the tensor product kernels is analyzed towards decomposing the tensor kernel SVM weight vector without accessing the feature spaces. It was shown that the decomposed weights can also be used as single source classifiers as well as to the task content based information retrieval.

An important kernel - called a chordal tensor - that allows direct application to any dimensional tensors, was introduced by Signoretto et al. [37]. The proposed non-parametric tensor-based model showed very high discriminative power, compared to the previously proposed methods. It overcomes the commonly applied naive tensor comparisons, by taking into consideration each of the spaces spanned by each of the dimension spaces spanned by the flattened versions of the input tensors. In the presented experiments the least-squares version of the SVM classifiers was employed, however. Influenced by good results obtained with the chordal tensor, in this paper we investigate its properties to the common problem of image recognition with the SMO versions of the SVM classifiers.

3. Chordal Distance Between Pattern Tensors

In this Section, we will discuss the basis of tensors applied in pattern recognition and machine learning, as well as methodology for computing the chordal distance for tensor-based kernels.

3.1. Tensors for Pattern Recognition

Tensors play an important role in physics, especially in mechanics and relativistic physics [12][23]. They are used to describe relations among physical values, which follow changes of the coordinate systems change in accordance with strict rules called tensor transformation laws. The other definition of tensors can be constructed with help of the multi-linear functions operating on a vector field and its dual [3]. However, in data analysis tensors are limited to represent multidimensional cubes of data. In other words, data that depends on multiple factors, or degree of freedom, can be grouped into such a multidimensional array. An example can be measurements of groups of clients buying specific groups of merchandize at certain days, prices, etc. Similarly, a color video signal can be seen as a four dimensional cube of values changing in accordance with the x-y spatial, c color, and t time dimensions.

Thus, frequently patterns which are expressed as multidimensional cubes of data need to be analyzed with mathematical tools relevant to tensor analysis. In this paper we focus on tensor representation of different image types. Theretofore, a brief introduction to tensor representation for data analysis with special stress on image processing is presented. A more detailed description with further explanations can be found in literature [5][8]. A *tensor*

$$\mathcal{A} \in \Re^{N_1 \times N_2 \times \ldots N_L} \tag{1}$$

is a L-dimensional "cube" of real valued data, in which each dimension corresponds to a different factor of the input data space. For further discussion, scalars are denoted with small letters, such as a, column vectors with bold $\mathbf{a}$, matrices with the bold capitals, such as $\mathbf{A}$, and higher order structures - tensors - with bold calligraphic letters, such as $\mathcal{A}$.

With the above definition of a tensor, the j-mode vector of the K-th order tensor is a vector obtained from elements of $\mathcal{A}$ by varying only one its index N_j while keeping all other indices fixed. Further, if from the tensor $\mathcal{A}$ the matrix

$$\mathbf{A}_{(j)} \in \Re^{N_j \times (N_1 N_2 \ldots N_{j-1} N_{j+1} \ldots N_L)} \tag{2}$$

is created, then the columns of $\mathbf{A}_{(j)}$ are j-mode vectors of $\mathcal{A}$. Also $A_{(j)}$ is a matrix representation of the tensor $\mathcal{A}$, called a j-mode tensor flattening (known also as tensor matricization). The j-th index becomes a row index of $\mathbf{A}_{(j)}$, while its column index is a product of all the rest L-1 indices. An analysis of sufficient computer representations of (2) are discussed in many publications [26]. Figure 1 shows three flattenings of a 3D tensor.

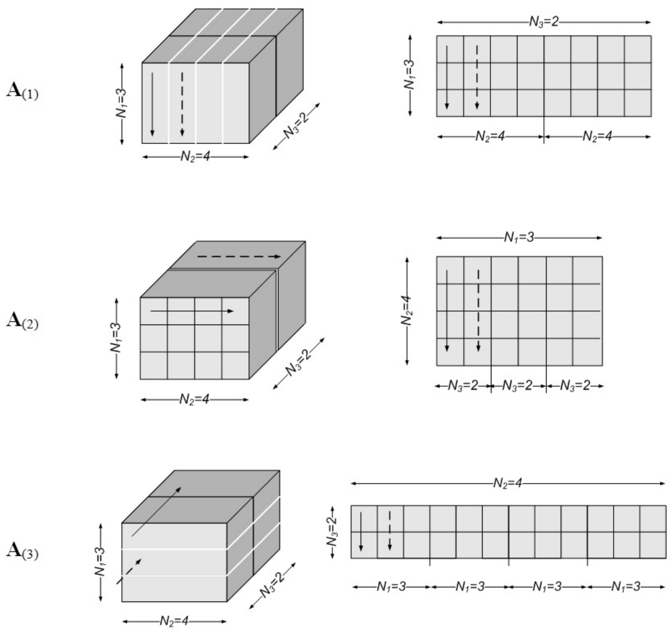

Figure 1: Examples of 3D tensor flattening in the forward index permutation mode.

The three distinct flattenings of a $N_1{\times}N_2{\times}N_3$ (3×4×2) tensor shown in Figure 1 assume a forward mode of index permutations which is more suitable for video processing [9]. The matrix $\mathbf{A}_{(1)}$, which directly reflects video location in memory, has dimensions $N_1{\times}N_2N_3$, $\mathbf{A}_{(2)}$ – $N_2{\times}N_3N_1$, and $\mathbf{A}_{(3)}$ – $N_3{\times}N_1N_2$, respectively.

A useful concept of tensor algebra is a p-mode product of a tensor $\mathcal{A} \in \Re^{N_1 \times N_2 \times \ldots N_L}$ with a matrix $\mathbf{M} \in \Re^{Q \times N_p}$. A result of this operation is the tensor $\mathcal{B} \in \Re^{N_1 \times N_2 \times \ldots N_{p-1} \times Q \times N_{p+1} \times \ldots N_L}$ whose elements are as follows

$$\mathcal{B}_{n_1 n_2 \ldots n_{p-1} q n_{p+1} \ldots n_L} = (\mathcal{A} \times_p \mathbf{M})_{n_1 n_2 \ldots n_{p-1} q n_{p+1} \ldots n_L} = \sum_{n_p=1}^{N_p} a_{n_1 n_2 \ldots n_{p-1} n_p n_{p+1} \ldots n_L} m_{q n_p}. \tag{3}$$

As was shown, the *p-mode* product can be equivalently represented in terms of the flattened versions of the tensors $\mathbf{A}_{(p)}$ and $\mathbf{B}_{(p)}$ [26][25]. That is, if the following holds

$$\mathcal{B} = \mathcal{A} \times_p \mathbf{M} \tag{4}$$

then

$$\mathbf{B}_{(p)} = \mathbf{M}\,\mathbf{A}_{(p)} \tag{5}$$

An important property of the tensor flattening is that each gives rise to a different matrix with specific properties. Thus, an analysis of the space properties spanned by each flattening matrix $\mathbf{A}_{(j)}$, gives unique information of data cube seen from the j-th dimension. This property is used to build the higher-order singular value decomposition (HOSVD), as well as will be used to construct a suitable kernel for data analysis, as will be discussed in the next section.

To analyze properties of a space spanned by each matrix $\mathbf{A}_{(j)}$, it is decomposed with the Singular Value Decomposition (SVD) decomposition, as follows [31]

$$\mathbf{A}^{(j)} = \mathbf{S}^{(j)}\mathbf{V}^{(j)}\mathbf{D}^{T(j)} = \sum_{i=1}^{R_{A_{(j)}}} v_i^{(j)}\mathbf{s}_i^{(j)}\mathbf{d}_i^{T(j)} = \begin{bmatrix} \mathbf{S}_{\mathbf{A},1}^{(j)} & \mathbf{S}_{\mathbf{A},2}^{(j)} \end{bmatrix} \begin{bmatrix} \mathbf{V}_{\mathbf{A},1}^{(j)} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{D}_{\mathbf{A},1}^{T(j)} \\ \mathbf{D}_{\mathbf{A},2}^{T(j)} \end{bmatrix}. \tag{6}$$

In the above, $_{\mathbf{A},1}$ and $_{\mathbf{A},2}$ denote indices of block matrices related to the kernel and null spaces of $\mathbf{A}_{(j)}$, respectively. It holds also that $\mathbf{S}_{\mathbf{A},1}^{(j)}$ and $\mathbf{D}_{\mathbf{A},1}^{T(j)}$

are unitary matrices of the kernel of $\mathbf{A}_{(j)}$. Finally,$\mathbf{V}^{(j)}_{\mathbf{A},1}$ is a diagonal matrix with R_A non-zero elements, whose size determines rank of the matrix $\mathbf{A}_{(j)}$. It immediately follows also that

$$\mathbf{D}^{T(j)}_{\mathbf{A},1}\mathbf{D}^{(j)}_{\mathbf{A},1} = \mathbf{I}_{R_{\mathbf{A}} \times R_{\mathbf{A}}} \tag{7}$$

The analogous conditions hold for the j-th mode flattening of the tensor $\mathcal{B}$. However, in this case, its rank can be different which is further denoted by R_B.

In pattern recognition one of the most important concepts is a distance between patterns. In the well know vector space, frequently used is the Euclidean distance. It can be also directly applied to the patterns represented as tensors. However, such simplistic approach disregards important information hidden behind the spatial composition and interrelations among data. In this case a useful concept is to consider distances of the subspaces spanned by the flattening matrices of pattern tensors. In this formulation a more appropriate is a distance among principal angles, called *projection Frobenius norm* [6] or *a chordal distance* [19]. For two tensors in their *j-th* flattened mode matrices $\mathbf{A}_{(j)}$ and $\mathbf{B}_{(j)}$, their chordal distance is defined as follows [37]

$$D^2_{ch}\left(\mathbf{A}_{(j)}, \mathbf{B}_{(j)}\right) = D^2_F\left(\Pi_{\mathbf{A}_{(j)}}, \Pi_{\mathbf{B}_{(j)}}\right) = \left\|\Pi_{\mathbf{A}_{(j)}} - \Pi_{\mathbf{B}_{(j)}}\right\|^2_F \tag{8}$$

where $\Pi_{\mathbf{A}_{(j)}}$denotes a projector matrix of $\mathbf{A}_{(j)}$, defined as follows [31]

$$\Pi_{\mathbf{A}_{(j)}} = \mathbf{D}^{(j)}_{\mathbf{A},1}\mathbf{D}^{T(j)}_{\mathbf{A},1} \tag{9}$$

Inserting (9) into (8) yields

$$D^2_{ch}\left(\mathbf{A}_{(j)},\mathbf{B}_{(j)}\right) = \left\|\mathbf{D}^{(j)}_{\mathbf{A},1}\mathbf{D}^{T(j)}_{\mathbf{A},1} - \mathbf{D}^{(j)}_{\mathbf{B},1}\mathbf{D}^{T(j)}_{\mathbf{B},1}\right\|^2_F \tag{10}$$

Based on D^2_{ch} a tensor kernel can be defined as follows [37]

$$\begin{aligned} K_j\left(\mathcal{A},\mathcal{B}\right) &= \exp\left(-\tfrac{1}{2\sigma^2}D^2_{ch}\left(\mathbf{A}_{(j)},\mathbf{B}_{(j)}\right)\right) \\ &= \exp\left(-\tfrac{1}{2\sigma^2}\left\|\mathbf{D}^{(j)}_{\mathbf{A},1}\mathbf{D}^{T(j)}_{\mathbf{A},1} - \mathbf{D}^{(j)}_{\mathbf{B},1}\mathbf{D}^{T(j)}_{\mathbf{B},1}\right\|^2_F\right). \end{aligned} \tag{11}$$

Thus, for a L-dimensional tensor a product kernel can be defined as follows

$$K\left(\mathcal{A},\mathcal{B}\right) = \prod_{j=1}^{L} K_j\left(\mathcal{A},\mathcal{B}\right) = \prod_{j=1}^{L}\exp\left(-\frac{1}{2\sigma^2}\left\|\mathbf{D}^{(j)}_{\mathbf{A},1}\mathbf{D}^{T(j)}_{\mathbf{A},1} - \mathbf{D}^{(j)}_{\mathbf{B},1}\mathbf{D}^{T(j)}_{\mathbf{B},1}\right\|^2_F\right) \tag{12}$$

Evidently, computation of (12) requires prior computations of $2{\cdot}L$ SVD decompositions, after which the Frobenius norm needs to be computed out of the kernel space matrices $\mathbf{D}$. However, in the case of large tensors this might require a prohibitive time of computations and the expression can be simplified, as will be discussed.

3.2. Computation of the Chordal Distance

Let us denote the squared norm in (10) as follows

$$\left\|\mathbf{P}-\mathbf{Q}\right\|^2 = Tr\left(\mathbf{P}^T\mathbf{P}\right) - 2Tr\left(\mathbf{P}^T\mathbf{Q}\right) + Tr\left(\mathbf{Q}^T\mathbf{Q}\right) \tag{13}$$

where the two matrices $\mathbf{P}$ and $\mathbf{Q}$ are defined as follows

$$\mathbf{P} = \mathbf{D}^{(j)}_{\mathbf{A},1}\mathbf{D}^{T(j)}_{\mathbf{A},1}\;,\quad \mathbf{Q} = \mathbf{D}^{(j)}_{\mathbf{B},1}\mathbf{D}^{T(j)}_{\mathbf{B},1} \tag{14}$$

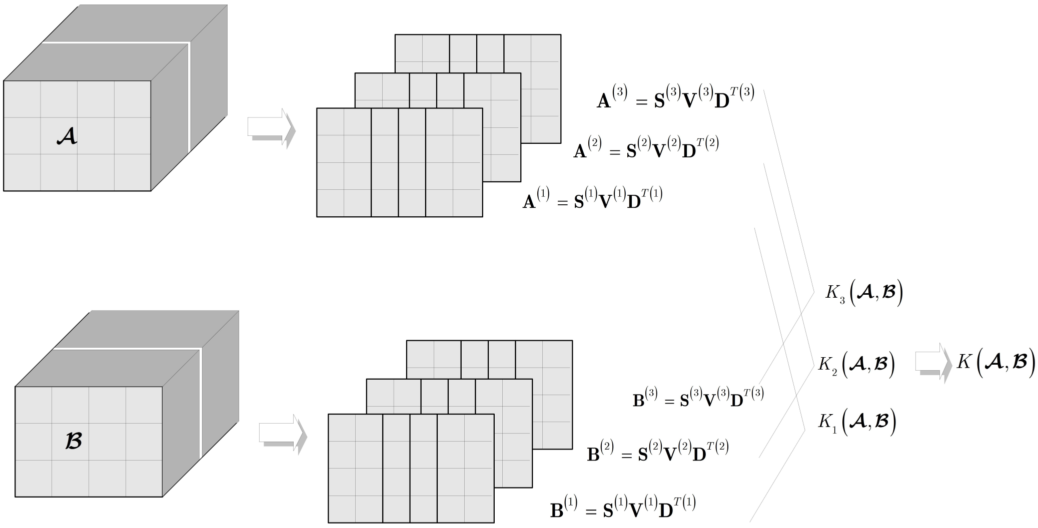

Figure 2: Visualization of the computation of the chordal distance between two tensors $\mathcal{A}$ and $\mathcal{B}$.

The two matrices $\mathbf{P}$ and $\mathbf{Q}$ are of the same size. Now, for the consecutive terms in (13), the following is obtained (we skip the indices 1 and 2 for simplicity)

$$Tr\left(\mathbf{P}^T\mathbf{P}\right) = Tr\left(\left(\mathbf{D_A}\mathbf{D_A}^T\right)^T\left(\mathbf{D_A}\mathbf{D_A}^T\right)\right) = Tr\left(\mathbf{D_A}\underbrace{\mathbf{D_A}^T\mathbf{D_A}}_{\mathbf{I}}\mathbf{D_A}^T\right) = \tag{15}$$
$$Tr\left(\mathbf{D_A}\mathbf{D_A}^T\right) = Tr\left(\mathbf{D_A}^T\mathbf{D_A}\right) = R_\mathbf{A}.$$

Similarly, it holds that

$$Tr\left(\mathbf{Q}^T\mathbf{Q}\right) = R_\mathbf{B} \tag{16}$$

On the other hand, the middle term in (13) can be expressed as follows

$$Tr\left(\mathbf{P}^T\mathbf{Q}\right) = Tr\left(\left(\mathbf{D_A}\mathbf{D_A}^T\right)^T\mathbf{D_B}\mathbf{D_B}^T\right) = Tr\left(\mathbf{D_A}\mathbf{D_A}^T\mathbf{D_B}\mathbf{D_B}^T\right) =$$
$$Tr\left(\mathbf{D_A}^T\mathbf{D_B}\mathbf{D_B}^T\mathbf{D_A}\right) = Tr\left(\left(\underbrace{\mathbf{D_B}^T\mathbf{D_A}}_{\mathbf{G}}\right)^T\underbrace{\mathbf{D_B}^T\mathbf{D_A}}_{\mathbf{G}}\right) = Tr\left(\mathbf{G}^T\mathbf{G}\right). \tag{17}$$

Thus, (13) can be written as follows

$$D_{ch}^2\left(\mathbf{A}_{(j)}, \mathbf{B}_{(j)}\right) = R_{\mathbf{A}} + R_{\mathbf{B}} - 2\,Tr\left(\mathbf{G}_{(j)}^T\mathbf{G}_{(j)}\right) \tag{18}$$

where

$$\mathbf{G}_{(j)} = \mathbf{D}_{\mathbf{B},1}^{T(j)}\mathbf{D}_{\mathbf{A},1}^{(j)} \tag{19}$$

Expressions (18) and (19) are easier for computation than (10) since only the matrix $\mathbf{G}_{(j)}$ needs to be computed after computation of the SVD decompositions of j-th mode flattened versions $\mathbf{A}_{(j)}$ and $\mathbf{B}_{(j)}$ of the tensors $\mathcal{A}$ and $\mathcal{B}$, respectively. For the computation of the chordal kernel distance, such computations need to be repeated L times, which is a dimensionality of the two tensors.

In practice, choice of the non-zero values, denoted by R_A and R_B in the matrices $\mathbf{V}_A$ as well as $\mathbf{V}_B$, respectively, is a non-trivial one. In our approach a simple threshold value was used, i.e., all singular values falling below this threshold are assumed to be 0. Thus, for each $\mathbf{A}_{(j)}$, its $R_{\mathbf{A}(j)}$ is a number of singular values above the experimentally chosen threshold.

When comparing computations of D_{ch} with (10) and (18) we see a significant difference. The first is that the matrix $\mathbf{G}_{(j)}$ is of dimensions $R_{\mathbf{B}(j)}$ x $R_{\mathbf{A}(j)}$, whereas each product in the difference in (10) is of dimensions N_j x

N_j which is always larger. The second benefit of using (18) is that Tr can be directly computed from the matrix $\mathbf{G}$ given in (19) with no further matrix multiplications.

For a matrix of dimensions r x c, the computational complexity of determining only the matrices $\mathbf{V}$ and $\mathbf{D}$ from the SVD decomposition is of order $4c^2(r + 2c)$[18]. This can seem prohibitive for large tensors. However, in many pattern recognition tasks, such as computation of the nearest neighbors, SVD for the prototype patterns can be computed beforehand and stored in a database. This greatly simplifies computations since once a test pattern is SVD decomposed. Then only (18) needs to be determined which requires one matrix multiplication given in (19), as well as one inner product to determine the third term in (18).

4. Experimental Investigations

In this section, we will present the experimental evaluation of the SVM classifier with the chordal distance based (CDB) kernel for analysis of complex multi-dimensional data. We want to establish, if the tensor representation of complex data can boost the quality of the kernel classifier. We compare the proposed method with a popular SVM with RBF kernel, that is widely used in many practical applications. We run two kinds of experiments:

- Binary classification task, in which we analyse the performance of chordal distance based kernel for two-class problems. This is a first choice due to the binary nature of SVM classifier.

- Multi-class classification task, as it reflects many real-life problems with a set of possible class labels. Here, a reconstruction scheme must be applied in order to aggregate local binary decisions of SVM classifiers into a single multi-class output.

4.1. Datasets

Chordal distance based kernel was designed in order to efficiently represent and analyze complex, multi-dimensional data such as images or video sequences. In order to evaluate its usefulness, we have selected two popular datasets from the face recognition domain: Georgia Tech Faces and AT&T (ORL). They both have a high number of classes and are characterized by a high interclass and low intraclass variances. Therefore, they are most suitable for comparing the tensor representation with a standard vector one for the purpose of training SVM classifier.

Georgia Tech Faces[4] dataset consists of images of 50 people taken during 1999. Each person from the dataset is described by 15 corresponding JPEG color images with cluttered background taken at resolution 640x480 pixels. The average size of the faces in these images is 150x150 pixels. The pictures show frontal and/or tilted faces with different facial expressions, lighting conditions and scale. A sample of Georgia Tech Faces dataset is presented in Figure 3.

AT&T (ORL)[5] dataset consists of images of 40 people taken between April 1992 and April 1994. Each person from the dataset is described by 10 corresponding 8-bit grayscale PGM images taken at resolution 92 x 112 pixels. Images in this database were taken at different times varying the

[4] `http://www.anefian.com/research/face_reco.htm`

[5] `http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html`

Figure 3: Sample of different objects and classes from the Georgia Tech Faces dataset.

lighting, facial expression and facial details (e.g., with and without glasses). All the images were taken against a dark homogeneous background with the subjects in an upright, frontal position (with tolerance for some side movement). A sample of Georgia Tech Faces dataset is presented in Figure 4.

4.2. Experimental set-up

As a base classifier, we use Support Vector Machine trained with the SMO procedure [35]. In our experiments, we use both RBF and CDB kernels and tune their parameters according to the parameter selection procedures implemented in LIBSVM package [4]. Tensor representation and decompositions were implemented in C++ using the DeRecLib software package [2].

RBF kernel uses vector representations of classifier images, while CDB kernel uses tensor representation.

For binary classification experiments, used multi-class datasets are decomposed into a set of binary problems. This is done with one-versus-all scheme [14] (in which one of the classes is compared against the remaining ones).

For using SVM on multi-class problems, one needs a reconstruction scheme in order to aggregate local binary decisions of SVM classifiers into a single multi-class output. In our experiments, we use the one-versus-one Pairwise Coupling method [21], as it was experimentally proved that this is among best choices for aggregating binary classifiers [15].

For simultaneous training/testing and pairwise statistical analysis, we use a 5x2 CV combined F-test [1]. It repeats five-time two fold cross-validation so that in each of the folds the size of the training and testing sets is equal. The combined F-test is conducted by comparison of all versus all. As a test score

Figure 4: Sample of different objects and classes from the AT&T (ORL) dataset.

the probability of rejecting the null hypothesis is adopted, i.e. that classifiers have the same error rates. As an alternative hypothesis, it is conjectured that tested classifiers have different error rates. A small difference in the error rate implies that the different algorithms construct two similar classifiers with similar error rates; thus, the hypothesis should not be rejected. For a large difference, the classifiers have different error rates and the hypothesis should be rejected.

Additionally, one should note that for binary classification we have a large number of comparisons. To analyze them, we use the Shaffer post-hoc test [16] to find out which of the tested methods are distinctive among an n x n comparison. The post-hoc procedure is based on a specific value of the significance level α. Additionally, the obtained p-values should be examined in order to check how different given two algorithms are.

We fix the significance level $\alpha = 0.05$ for all comparisons.

4.3. Experimental Results

The results for binary classification problems for both datasets are given in Tables 1 - 4, with respect to accuracy and p-values from pairwise statistical analysis. Table 3 presents the results of Shaffer test for binary problems. Results for multi-class analysis are given in Table 2. Computational time is presented in Table 5.

4.4. Discussion

Experimental investigations allow us to draw some interesting conclusions about the usefulness of the proposed SVM with chordal distance-based kernel.

Table 1: Accuracy [%] and statistical analysis for Georgia Tech Faces dataset, decomposed into 50 binary problems. Symbol '+' stands for a situation in which the proposed method is superior, '-' for vice versa, and '=' represents a lack of statistically significant differences.

Person	SVM (CDB kernel)	SVM (RBF kernel)	p-value	Person	SVM (CDB kernel)	SVM (RBF kernel)	p-value
Person 1	87.24	81.30	+ (0.0064)	Person 26	89.76	69.16	+ (0.0046
Person 2	91.20	83.19	+ (0.0045)	Person 27	83.29	78.16	+ (0.0261
Person 3	78.38	69.29	+ (0.0106)	Person 28	86.43	78.49	+ (0.0328
Person 4	74.90	74.11	= (0.0743)	Person 29	89.10	78.49	+ (0.0296
Person 5	82.18	64.29	+ (0.0023)	Person 30	91.20	82.46	+ (0.0183
Person 6	85.87	73.77	+ (0.0280)	Person 31	87.04	80.24	+ (0.0348
Person 7	92.46	74.06	+ (0.0185)	Person 32	82.65	75.89	+ (0.0416
Person 8	94.82	82.38	+ (0.0132)	Person 33	84.58	71.90	+ (0.0088
Person 9	85.39	83.51	= (0.1056)	Person 34	78.72	76.84	= (0.0893
Person 10	91.05	76.39	+ (0.0019)	Person 35	87.82	81.29	+ (0.0368
Person 11	94.28	75.20	+ (0.0120)	Person 36	79.85	77.93	= (0.7006
Person 12	93.78	78.49	+ (0.0206)	Person 37	81.28	69.48	+ (0.0062
Person 13	87.93	73.98	+ (0.0326)	Person 38	83.78	78.30	+ (0.0421
Person 14	92.30	85.29	+ (0.0392)	Person 39	88.56	72.19	+ (0.0086
Person 15	86.58	71.18	+ (0.0158)	Person 40	85.39	74.20	+ (0.0146
Person 16	87.38	72.29	+ (0.0187)	Person 41	80.68	68.90	+ (0.0073
Person 17	86.83	64.98	+ (0.0065)	Person 42	88.36	80.48	+ (0.0250
Person 18	88.32	75.83	+ (0.0382)	Person 43	84.93	68.39	+ (0.0102
Person 19	89.38	75.30	+ (0.0109)	Person 44	87.03	79.12	+ (0.0184
Person 20	84.39	73.94	+ (0.0306)	Person 45	84.37	72.58	+ (0.0097
Person 21	87.39	86.12	= (0.2503)	Person 46	79.17	78.60	= (0.1837
Person 22	90.32	78.39	+ (0.0128)	Person 47	83.58	72.16	+ (0.0113
Person 23	86.37	80.34	+ (0.0407)	Person 48	86.12	70.84	+ (0.0074
Person 24	88.72	76.38	+ (0.0168)	Person 49	85.28	67.89	+ (0.0038
Person 25	87.29	74.39	+ (0.0094)	Person 50	90.06	77.05	+ (0.0072

Table 2: Accuracy [%] and statistical analysis for AT&T (ORL) dataset, decomposed into 40 binary problems. Symbol '+' stands for a situation in which the proposed method is superior, '-' for vice versa, and '=' represents a lack of statistically significant differences.

Person	SVM (CDB kernel)	SVM (RBF kernel)	p-value	Person	SVM (CDB kernel)	SVM (RBF kernel)	p-value
Person 1	93.72	77.28	+ (0.0106)	Person 21	87.98	79.24	+ (0.0125
Person 2	94.87	79.40	+ (0.0089)	Person 22	90.07	82.18	+ (0.0246
Person 3	89.38	82.89	+ (0.0174)	Person 23	91.28	81.52	+ (0.0283
Person 4	90.07	82.37	+ (0.0218)	Person 24	94.52	86.88	+ (0.0139
Person 5	90.93	82.78	+ (0.0236)	Person 25	86.34	85.75	= (0.2501
Person 6	91.85	79.88	+ (0.0168)	Person 26	89.74	81.29	+ (0.0202
Person 7	90.19	68.70	+ (0.0036)	Person 27	87.26	82.68	+ (0.0408
Person 8	93.28	82.78	+ (0.0138)	Person 28	82.93	80.90	= (0.3770
Person 9	90.92	83.74	+ (0.0274)	Person 29	83.18	81.06	= (0.1390
Person 10	83.89	81.27	= (0.1602)	Person 30	88.46	76.19	+ (0.0054
Person 11	87.92	81.63	+ (0.0149)	Person 31	90.71	82.05	+ (0.0118
Person 12	92.37	79.36	+ (0.0107)	Person 32	92.18	85.38	+ (0.0247
Person 13	93.10	84.28	+ (0.0138)	Person 33	90.86	80.19	+ (0.0094
Person 14	91.18	78.49	+ (0.0064)	Person 34	84.26	82.37	= (0.3118
Person 15	92.23	84.15	+ (0.0237)	Person 35	85.49	81.23	= (0.0736
Person 16	91.37	86.38	+ (0.0398)	Person 36	93.58	82.97	+ (0.0163
Person 17	92.03	82.17	+ (0.0196)	Person 37	91.28	77.47	+ (0.0304
Person 18	87.46	85.12	= (0.1592)	Person 38	85.46	77.05	+ (0.0248
Person 19	90.24	85.49	+ (0.0387)	Person 39	90.63	79.48	+ (0.0379
Person 20	94.12	83.78	+ (0.0109)	Person 40	84.28	82.97	= (0.0719

Table 3: Results for Shaffer test over 90 different binary classification experiments (50 from Georgia Tech Faces dataset and 40 from AT&T (ORL) dataset). Symbol '+' stands for a situation in which the proposed method is superior, '-' for vice versa, and '=' represents a lack of statistically significant differences.

Methods	p-value
SVM (CDB kernel) vs. SVM (RBF kernel)	+ (0.0082)

Table 4: Accuracy [%] and statistical analysis for Georgia Tech Faces and AT&T (ORL) datasets with Pairwise Coupling for multi-class classification. Symbol '+' stands for a situation in which the proposed method is superior, '-' for vice versa, and '=' represents a lack of statistically significant differences.

Dataset	SVM (CDB kernel)	SVM (RBF kernel)	p-value
Georgia Tech Faces	97.03	76.28	+ (0.0048)
AT&T (ORL)	97.89	81.24	+ (0.0102)

Table 5: Average time [s.] required for training a Support Vector Machine with given kernel for considered problems.

Dataset	SVM (CDB kernel)	SVM (RBF kernel)
Georgia Tech Faces (binary)	119.38	93.20
Georgia Tech Faces (multi-class)	288.24	241.39
AT&T (ORL) (binary)	98.34	72.26
AT&T (ORL) (multi-class)	204.38	158.23

One can observe a significant gain in accuracy for both binary and multi-class problems, when using the chordal distance-based kernel. RBF kernel is inferior in almost all of the cases, and never achieves statistically better results than the proposed tensor-based SVM. RBF kernel requires a vector input of data, and therefore loses valuable information about the images under consideration. Chordal distance-based kernel allows to process the entire image (regardless of the fact, if it is color or gray-scale) as a single object, thus preserving the spatial relations between pixels and maintaining additional information for the classification process. This is especially vivid in case of Georgia Tech Faces dataset, which consists of color images. Using RBF kernel forces the user to present it as a very long vector (three times the number of pixels), while the presented SVM with chordal distance-based kernel is able to efficiently represent each color matrix. Therefore, the proposed method can highly useful for very complex data, such as hyperspectral images or video sequences. Both pairwise and multiple comparisons statistical tests prove, that the proposed tensor version of SVM classifier achieves statistically superior results over all of the datasets.

The proposed classification system suffers the same limitations, as standard SVM methods - it is binary by nature. Therefore, we have used an efficient pairwise coupling aggregation method in order to reconstruct the original multi-class problem from a number of binary outputs. Experiments show, that the proposed algorithm significantly outperforms RBF kernel and returns highly competitive results in comparison to the ones presented in literature. Please note, that we do not use any kind of feature extraction or preprocessing - the implemented kernel allows us to efficiently process complex images with high final accuracy.

Finally, we must report that the proposed SVM version requires a longer training time than its RBF counterpart. However, for most of the classification task this is not a problem, as we may train the classifier beforehand without any time constraints. In such scenarios, the gain in accuracy is far more important than increase of the training time. We should note, that the response times of both RBF-based and chordal distance-based SVMs are almost identical - so it is suitable for real-time operation. For cases, in which one would require a re-trainable / adaptive classifier (e.g., data streams with concept drift), one of our future goals is to introduce a distributed version of our SVM classifier suitable to be run on GPU (e.g.,CUDA) architecture.

5. Conclusions

In this paper, we have presented a novel approach for handling complex and multidimensional data. It was based on processing data structures as tensors and classifying them with a Support Vector Machine with chordal distance-based kernel. By handling data as tensors, we were able to process multidimensional data (such as color images) as single objects, preserving the spatial relations between pixels. The used kernel allowed to efficiently tackle the tensor-based objects and compute distances between them. Support Vector Machine with this kernel was trained with Sequential Minimal Optimization procedure in order to provide a highly efficient pattern recognition algorithm. Necessary details for efficient implementation of the chordal distance-based kernel were provided.

We have examined our method on a number of experiments from face recognition domain, both binary and multi-class. We observed a significant increase in accuracy, when using chordal distance-based kernel. Statistical analysis, both pairwise and multiple comparison, allowed us to conclude that the proposed version of SVM classifier is highly more effective than its popular version with RBF kernel.

As future works, we plan to use the tensor-based kernel for one-class classification task and to formulate ensembles of SVM with chordal distance-based kernels for efficient decomposition of massive and multi-class data. We also plan to use this approach for handling hyperspectral images.

Acknowledgment

The financial support from the Polish National Science Centre NCN in the year 2014, contract no. DEC-2011/01/B/ST6/01994, is greatly acknowledged

References

[1] E. Alpaydin. Combined 5 x 2 cv f test for comparing supervised classification learning algorithms. *Neural Computation*, 11(8):1885–1892, 1999.

[2] Cyganek B. DeRecLib. `http://www.wiley.com/go/cyganekobject`, 2013.

[3] R.L. Bishop and S.I. Goldberg. *Tensor Analysis on Manifolds.* Dover Books on Mathematics. Dover Publications, 2012.

[4] C.-C. Chang and C.-J. Lin. LIBSVM: A library for support vector machines. *ACM TIST*, 2(3):27, 2011.

[5] Andrzej Cichocki, Rafal Zdunek, Anh Huy Phan, and Shun-ichi Amari. *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation*. Wiley Publishing, 2009.

[6] John H. Conway, Ronald H. Hardin, and Neil J.A. Sloane. Packing lines, planes, etc.: Packings in grassmannian spaces. *Experimental Mathematics*, 5(2):139–159, 1996.

[7] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Mach. Learn.*, 20(3):273–297, September 1995.

[8] B. Cyganek. *Object Detection and Recognition in Digital Images: Theory and Practice*. Wiley, 2013.

[9] B. Cyganek. Pattern recognition framework based on the best rank-(r1,r2,...,rk) tensor approximation. In *Computational vision and medical image processing IV: Proceedings of VipIMAGE 2013*, pages 301–306, 2013.

[10] Lieven De Lathauwer. *Signal Processing Based on Multilinear Algebra*. PhD thesis, Katholike Universiteit Leuven, 1997.

[11] Thomas G. Dietterich and Ghulum Bakiri. Solving multiclass learning problems via error-correcting output codes. *J. Artif. Int. Res.*, 2(1):263–286, January 1995.

[12] Y.I. Dimitrienko. *Tensor Analysis and Nonlinear Tensor Functions.* Springer, 2002.

[13] Richard O. Duda, Peter E. Hart, and David G. Stork. *Pattern Classification (2Nd Edition).* Wiley-Interscience, 2000.

[14] M. Galar, A. Fernandez, E. Barrenechea, H. Bustince, and F. Herrera. An overview of ensemble methods for binary classifiers in multi-class problems: Experimental study on one-vs-one and one-vs-all schemes. *Pattern Recognition*, 44(8):1761–1776, 2011.

[15] M. Galar, A. Fernández, E. Barrenechea Tartas, H. Bustince Sola, and F. Herrera. Dynamic classifier selection for one-vs-one strategy: Avoiding non-competent classifiers. *Pattern Recognition*, 46(12):3412–3424, 2013.

[16] S. García, A. Fernández, J. Luengo, and F. Herrera. Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: Experimental analysis of power. *Inf. Sci.*, 180(10):2044–2064, 2010.

[17] Navin Goel, George Bebis, and Ara Nefian. Face recognition experiments with random projection. volume 5779, pages 426–437, 2005.

[18] G.H. Golub and C.F. Van Loan. *Matrix Computations.* Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, 2013.

[19] Jihun Hamm and Daniel D. Lee. Grassmann discriminant analysis: A unifying view on subspace-based learning. In *Proceedings of the 25th*

International Conference on Machine Learning, ICML '08, pages 376–383, New York, NY, USA, 2008. ACM.

[20] David Hardoon and John Shawe-Taylor. Decomposing the tensor kernel support vector machine for neuroscience data with structured labels. *Machine Learning*, 79(1):29–46, May 2010.

[21] T. Hastie and R. Tibshirani. Classification by pairwise coupling. *Annals of Statistics*, 26(2):451–471, 1998.

[22] Xudong Jiang, Bappaditya Mandal, and Alex Kot. Eigenfeature regularization and extraction in face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(3):383–394, 2008.

[23] D. Kay. *Schaum's Outline of Tensor Calculus*. Schaum's Outline Series. McGraw-Hill Companies,Incorporated, 2011.

[24] Jaepil Ko and Hyeran Byun. Combining svm classifiers for multiclass problem: Its application to face recognition. In Josef Kittler and Mark S. Nixon, editors, *AVBPA*, volume 2688 of *Lecture Notes in Computer Science*, pages 531–539. Springer, 2003.

[25] Tamara G. Kolda and Brett W. Bader. Tensor decompositions and applications. *SIAM Rev.*, 51(3):455–500, August 2009.

[26] Lieven De Lathauwer, Bart De Moor, and Joos Vandewalle. On the best rank-1 and rank-(r1,r2,. . .,rn) approximation of higher-order tensors. *SIAM J. Matrix Anal. Appl.*, 21(4):1324–1342, March 2000.

[27] Yang Li, Yangzhou Du, and Xueyin Lin. Kernel-based multifactor analysis for image synthesis and recognition. In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, volume 1, pages 114–119 Vol. 1, Oct 2005.

[28] Zhao Lihong, Song Ying, Zhu Yushi, Zhang Cheng, and Zheng Yi. Face recognition based on multi-class svm. In *Proceedings of the 21st Annual International Conference on Chinese Control and Decision Conference*, CCDC'09, pages 5901–5903, Piscataway, NJ, USA, 2009. IEEE Press.

[29] Dahua Lin, Yingqing Xu, Xiaoou Tang, and Shuicheng Yan. Tensor-based factor decomposition for relighting. In *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, volume 2, pages II–386–9, Sept 2005.

[30] Tanaya Mandal, Angshul Majumdar, and Q. M. Jonathan Wu. Face recognition by curvelet based feature extraction. In *Proceedings of the 4th International Conference on Image Analysis and Recognition*, ICIAR'07, pages 806–817, Berlin, Heidelberg, 2007. Springer-Verlag.

[31] Carl D. Meyer, editor. *Matrix Analysis and Applied Linear Algebra.* Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2000.

[32] A.V. Nefian. Embedded bayesian networks for face recognition. In *Multimedia and Expo, 2002. ICME '02. Proceedings. 2002 IEEE International Conference on*, volume 2, pages 133–136 vol.2, 2002.

[33] Sung Won Park and M. Savvides. Individual kernel tensor-subspaces for robust face recognition: A computationally efficient tensor framework without requiring mode factorization. *Trans. Sys. Man Cyber. Part B*, 37(5):1156–1166, October 2007.

[34] Bo Peng and Gang Qian. Online gesture spotting from visual hull data. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(6):1175–1188, June 2011.

[35] J. C. Platt. Sequential minimal optimization: A fast algorithm for training support vector machines. Technical report, Advances in kernel methods - support vector learning, 1998.

[36] Lebby G. Shavers C., Li R. An svm-based approach to face detection. In *Proceedings of the 38th Southeastern Symposium on System Theory*, pages 362–366. Tennessee Technological University, 2006.

[37] Marco Signoretto, Lieven De Lathauwer, and Johan A.K. Suykens. A kernel-based framework to tensorial data analysis. *Neural Networks*, 24(8):861 – 874, 2011. Artificial Neural Networks: Selected Papers from {ICANN} 2010.

[38] Joshua B. Tenenbaum and William T. Freeman. Separating style and content with bilinear models. *Neural Comput.*, 12(6):1247–1283, June 2000.

[39] S. Valuvanathorn, S. Nitsuwat, and Mao Lin Huang. Multi-feature face recognition based on pso-svm. In *ICT and Knowledge Engineering (ICT*

Knowledge Engineering), 2012 10th International Conference on, pages 140–145, Nov 2012.

[40] M. A. O. Vasilescu and Demetri Terzopoulos. Multilinear analysis of image ensembles: Tensorfaces. In *Proceedings of the 7th European Conference on Computer Vision-Part I*, ECCV '02, pages 447–460, London, UK, UK, 2002. Springer-Verlag.

[41] M.A.O. Vasilescu and D. Terzopoulos. Multilinear (tensor) image synthesis, analysis, and recognition [exploratory dsp]. *Signal Processing Magazine, IEEE*, 24(6):118–123, Nov 2007.

[42] Chengbo Wang and Chengan Guo. An svm classification algorithm with error correction ability applied to face recognition. In *Proceedings of the Third International Conference on Advances in Neural Networks - Volume Part I*, ISNN'06, pages 1057–1062, Berlin, Heidelberg, 2006. Springer-Verlag.

[43] ChunJuan Yan. Face image gender recognition based on gabor transform and svm. In Gang Shen and Xiong Huang, editors, *Advanced Research on Electronic Commerce, Web Application, and Communication*, volume 144 of *Communications in Computer and Information Science*, pages 420–425. Springer Berlin Heidelberg, 2011.