Biomedical Signal Processing and Control 8 (2013) 755–763

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Noninvasive diagnosis of melanoma with tensor decomposition-based feature extraction from clinical color image

Ante Jukić[a,*,1], Ivica Kopriva[a], Andrzej Cichocki[b,c]

[a] Division of Laser and Atomic Research and Development, Ruđer Bošković Institute, Bijenička cesta 54, 10000 Zagreb, Croatia
[b] Laboratory for Advanced Brain Signal Processing, Brain Science Institute, RIKEN2-1 Hiroshawa, Wako-shi, Saitama 351-0198, Japan
[c] Warsaw University of Technology and Systems Research Institute, PAN, Poland

ARTICLE INFO

Article history:
Received 28 February 2013
Received in revised form 29 May 2013
Accepted 9 July 2013

Keywords:
Multidimensional signal processing
Tensor decomposition
Feature extraction
Noninvasive diagnosis
Melanoma

ABSTRACT

We propose a method for feature extraction from clinical color images, with application in classification of skin lesions. Proposed feature extraction method is based on tensor decomposition of the clinical color image of skin lesion. Since color image is naturally represented as a three-way tensor, it is reasonable to use multi-way techniques to capture the underlying information contained in the image. Extracted features are elements of the core tensor in the corresponding multi-way decomposition, and represent spatial-spectral profile of the lesion. In contrast to common methods that exploit either texture or spectral diversity of the tumor only, the proposed approach simultaneously captures spatial and spectral characteristics. The procedure is tested on a problem of noninvasive diagnosis of melanoma from the clinical color images of skin lesions, with overall sensitivity 82.1% and specificity 86.9%. Our method compares favorably with the state of the art results reported in the literature and provides an interesting alternative to the existing approaches.

1. Introduction

Noninvasive diagnosis of tumor is a procedure of identification and discrimination among various types of tumors by exploiting data that is not collected invasively, e.g., by biopsy. The aim of the noninvasive diagnosis is to detect malignant tumors with high accuracy and to simultaneously reduce the number of biopsies of the innocuous benign tumors. In general, methods for automated noninvasive diagnosis rely on sophisticated processing techniques applied on the collected data. The data can be acquired by various imaging modalities, such as multispectral (MSI) or hyperspectral imaging (HSI) [1,2]. Diagnosis is usually obtained by classifying a set of features extracted from the image of the tumor. Consequently, extraction of tumor-specific features is of central importance for accurate diagnosis. This is in line with the reasoning in the machine learning community that feature extraction matters more than the method used for classification [3,4].

Malignant melanoma is presently among the leading cancers in the white-skinned population, with rapidly increasing incidence and mortality rates over the last decades [5–7]. While advanced form of the cutaneous melanoma is still practically incurable, early diagnosis can significantly increase probability of survival. In fact, very high degree of curability can be achieved if the surgical excision is performed early enough [8]. The increased occurrence, along with the high lethality in case of an advanced melanoma, implies a demand for a simple and accurate screening test as an alternative to biopsy.

In spite of best efforts of researchers, the accuracy of the noninvasive diagnosis of the melanocytic lesions is far from ideal. Even binary problem of distinguishing between malignant melanoma and benign melanocytic lesion without histological examination remains a challenge [9]. One of the most widely used methods for preliminary diagnosis based on visual inspection is the so called ABCDE rule. It is a semi-quantitative diagnosis scheme based on naked eye inspection (i.e., a clinical image) of the skin lesion. The scheme includes examination on the asymmetry (A), border sharpness (B), color variation (C), number of differential structures (D) present in the lesion, and evolution (E) of the lesion in time. Unfortunately, it has shown a limited sensitivity in melanoma diagnosis [7], with overall accuracy depending on the level of expertise of dermatologist. In case of a well-trained dermatologist typical accuracy is around 75% [10]. Since the naked eye inspection achieves such low performance, epiluminescence light microscopy (ELM) or

* Corresponding author. Current address: Signal Processing Group, University of Oldenburg, 26111 Oldenburg, Germany. Tel.: +49 441 798 3377; fax: +49 441 798 3902.
E-mail addresses: ante.jukic@uni-oldenburg.de (A. Jukić), ikopriva@irb.hr (I. Kopriva), cia@brain.riken.jp (A. Cichocki).
[1] This work was done while the author was with the Ruđer Bošković Institute, Zagreb, Croatia.

1746-8094/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.bspc.2013.07.001

dermoscopy was suggested. The idea of ELM is to improve diagnostic performance by evaluating morphological features of lesions [11]. It was reported that the accuracy of the diagnosis obtained by dermoscopy in case of an expert dermatologists is around 75–84% [5,12]. Despite the formal training of dermatologists and availability of comprehensive atlases, interpretation of the features acquired by dermoscopy is often subjective and not reproducible, especially for an inexperienced clinician. Still, standard approach for classification of skin lesions in clinical practice is dermoscopy-based visual inspection, followed by biopsy and tissue analysis if needed [7].

An accurate computer-based automatic diagnostic system is required to reduce the need for highly trained dermatologist and time required for diagnosis. Such system with high levels of sensitivity and specificity (typically above 90%) could provide a second-opinion to dermatologist and also reduce the number of unnecessary biopsies. In [9] a meta-analysis of several studies was performed, and showed that computer-based diagnosis is comparable with performance of a human expert. Also, it was noted that better performance was achieved using dermoscopic images than with clinical images [9,13–16].

Opposed to inspection of morphological features, another line of research was focused on the MSI and HSI systems, with discrimination among tumors based on their spectral profiles. In [1], a MSI system with 10 channels in range 430–950 nm was designed and used for automatic discrimination between benign nevus and malignant melanoma by using auto-fluorescence of the tumor. A HSI system with 21 channels between 440 and 660 nm was used in [2] for automatic discrimination among benign and malignant skin tumors on laboratory mice based on fluorescence induced by fluorophore.

While automatic systems based on dermoscopic and high spectral resolution MSI/HSI systems show great potential, it was demonstrated in [9], as well as recently in [17], that automatic diagnosis of melanoma is possible using auto-fluorescent macroscopic (clinical) color images. These methods use clinical red-green-blue (RGB) color images, and achieve sensitivity in range 80–94% and specificity in range 46–95%. While dermoscopic images captures subsurface structure of the lesion, clinical capture what a clinician sees in naked eye inspection [18]. Results of the twelve studies that used clinical images [9] as well as results from [17] are summarized in Table 1. Practical importance of these results is in demonstration that accurate automatic diagnosis of melanoma is possible from clinical images using affordable RGB auto-fluorescence imaging. A more detailed methodological review of computerized analysis of pigmented skin lesions can be found in [18], and references therein.

Aim of this paper is to present a novel method for feature extraction, suitable for analysis of MSI data, and to demonstrate it on automated noninvasive diagnosis of cutaneous melanoma from clinical color images. The paper proposes a method for feature extraction from multi-way data using tensor decomposition [19,20], used for analysis of RGB color images. For this purpose experimental multispectral medical image is represented by its Tucker3 decomposition. Dimensionality analysis yields that the extracted features simultaneously contain spatial and spectral information about the image. To account for possible nonlinear nature of the acquired data, images are nonlinearly transformed prior to decomposition. The proposed scheme is demonstrated on noninvasive diagnosis of melanoma from clinical auto-fluorescent color images, and sensitivity and specificity are estimated through a two-fold cross-validation procedure.

The rest of the paper is organized as follows. In Section 2 we provide preliminaries on tensor algebra, and define a nonlinear transformation based on kernel methods. In Section 3 we propose a method for feature extraction. The experimental results are presented in Section 4, with concluding remarks contained in Section 5.

2. Preliminaries

This section contains basics of multi-way analysis, as well as definition of nonlinear transformation related to kernels. Also, interpretation of multispectral image as tensor is presented, with focus on RGB color image. In the following, scalars will be denoted by lowercase italic letters (e.g., x), vectors by bold lowercase letters (e.g., $\mathbf{x}$), matrices by bold capital letters (e.g., $\mathbf{X}$), and tensors by underlined bold letters (e.g., $\underline{\mathbf{X}}$).

2.1. Tensor algebra and Tucker3 decomposition

Tensors are generalization of matrices and vectors. A tensor can be represented by a multi-way array with arbitrary number of indices, for example, an N-mode tensor has N indices. For clarity, and motivated by application in RGB image analysis, we will focus on tensors with three indices. In this paper an image RGB color image is represented by tensor $\underline{\mathbf{X}} \in \mathbb{R}_{0+}^{I_1 \times I_2 \times I_3}$, that consists of elements $x_{i_1 i_2 i_3} \in \mathbb{R}_{0+}$, with $\mathbb{R}_{0+}$ denoting the set of nonnegative real numbers. Each index in tensor is called way or mode, and number of levels on a mode represents dimension of that mode, e.g., dimension of mode-1 is I_1. This is in line with the standard notation used in multi-way analysis [21]. RGB image $\underline{\mathbf{X}}$ is a set of $I_3 = 3$ spectral images, corresponding to red, green and blue color channels

Table 1
Comparative performance analysis of fifteen studies related to image analysis based automated diagnoses of melanoma from clinical images.[a]

Source	Sample size	Melan. [%]	Method of analysis	Sens. [%]	Spec. [%]
Green et al., 1991	70	7	Single set	80	91
Cascinelli et al., 1992	88	49	Single set	83	60
Claridge et al., 1992	88	49	Single set	91	69
Schindewolf et al., 1993	353	61	Ten-fold CV	94	88
Green et al., 1994	164	11	Single set	89	89
Ercal et al., 1994	214	56	Single set	80	86
Schindewolf et al., 1994	404	59	Ten-fold CV	90	88
Bono et al., 1996	43	42	Single set	83	72
Cristofolini et al., 1997	176	20	Single set	78	46
Seidenari et al., 1998	90	34	Single set	93	95
Smith et al., 2000	60	47	Single set	86	94
Farina et al., 2000	237	28	Single set	80	46
Tabatabaie et al., 2008, [17]	160	50	Single set	82.5	92.5
Tabatabaie et al., evaluated herein	180	50	Two-fold CV	79.9	79.1
Proposed method	180	50	Two-fold CV	82.1	86.9

[a] Full references to first twelve studies are given in [9]. (Method of analysis refers to a procedure used to validate the method. Single set means that the cross-validation procedure was not used, and that the reported performance was based on a result of prediction on a single test set.)

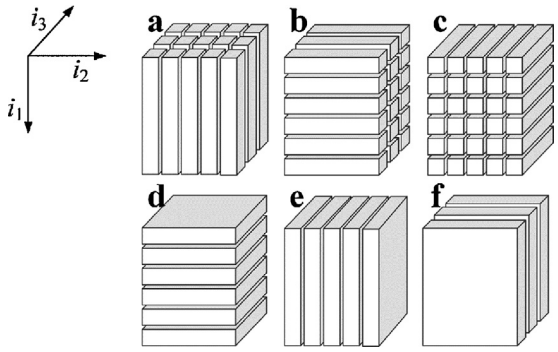

Fig. 1. Fibers in (a) mode-1; (b) mode-2; (c) mode-3; and (d) horizontal; (e) lateral; and (f) frontal slices.

[22]. Each spectral image consists of $I_1 \times I_2$ pixels, with $x_{i_1 i_2 i_3}$ representing brightness intensity at pixel (i_1, i_2) in spectral band i_3. Consequently, two modes of tensor $\underline{\mathbf{X}}$ are used for spatial information, i.e., rows and columns of the color image, and third mode is for spectral band. By fixing one or two indices of a three-way tensor we can define subtensors. Tensor slice is a matrix (two dimensional array) obtained by fixing a single index in a three-way tensor. For example, frontal slice of three-way tensor $\underline{\mathbf{X}}$ is obtained by fixing index i_3, and its usually denoted as $\mathbf{X}_{::i_3}$ or $\mathbf{X}_{i_3}$. By fixing all indices except i_n we obtain an I_n dimensional vector called nth mode fiber. For example, mode-3 fiber of $\underline{\mathbf{X}}$ at position (i_1, i_2) is a vector denoted as $\mathbf{x}_{i_1 i_2}$ or $\mathbf{x}_{i_1 i_2 :}$. For illustration, fibers and slices of a three-way tensor are displayed in Fig. 1.

Tensor is often transformed to matrix and vice versa. Procedure of rearranging the elements of a tensor into a matrix is known as unfolding or matricization. The mode-n unfolding of tensor $\underline{\mathbf{X}}$ is a matrix $\mathbf{X}_{(n)} \in \mathbb{R}^{I_n \times \prod_{k,k\neq n} I_k}$ that consists of mode-n fibers stacked in the matrix as columns. The particular ordering of mode-n fibers into columns of $\mathbf{X}_{(n)}$ is not important as long it is consistent through all computations [19,23]. Important notion is the n-rank of a tensor, defined as the dimension of the space spanned with columns of $\mathbf{X}_{(n)}$. If three-way tensor has n-ranks equal to J_1, J_2, J_3 we say that it is a rank-(J_1, J_2, J_3) tensor, i.e., it has multilinear rank (J_1, J_2, J_3).[2] Mode-n product of a tensor $\underline{\mathbf{Y}}$ and matrix $\mathbf{A}$ is defined when number of columns of matrix is equal to the dimension of the tensor in mode n. The result is a new tensor $\underline{\mathbf{Z}} = \underline{\mathbf{Y}} \times_n \mathbf{A}$, expressed in unfolded form as $\mathbf{Z}_{(n)} = \mathbf{A}\mathbf{Y}_{(n)}$ [19]. For example, mode-1 product of a three-way tensor $\underline{\mathbf{Y}} \in \mathbb{R}^{J_1 \times J_2 \times J_3}$ and a matrix $\mathbf{A} \in \mathbb{R}^{I_1 \times J_1}$ is a three-way tensor $\underline{\mathbf{Z}} = \underline{\mathbf{Y}} \times_1 \mathbf{A} \in \mathbb{R}^{I_1 \times J_2 \times J_3}$ with elements

$$z_{i_1 j_2 j_3} = \sum_{j_1=1}^{J_1} y_{j_1 j_2 j_3} \cdot a_{i_1 j_1},$$

where $a_{i_1 j_1} \in \mathbb{R}$ denotes element of matrix $\mathbf{A}$ on the position (i_1, j_1).

Tensor decompositions are valuable in analysis of multi-way data, and applications in image and signal analysis, neuroscience, chemometrics and psychometrics (for detailed survey see [23]). Basic decomposition of general N-way tensor is the Tucker decomposition [24]. In case of three-way tensors it is often referred to as Tucker3 decomposition, and it is illustrated in Fig. 2. Three-way tensor $\underline{\mathbf{X}}$ is decomposed into three factor matrices $\mathbf{A} \in \mathbb{R}^{I_1 \times J_1}$, $\mathbf{B} \in \mathbb{R}^{I_2 \times J_2}$, $\mathbf{C} \in \mathbb{R}^{I_3 \times J_3}$, and a three-way core tensor $\underline{\mathbf{G}} \in \mathbb{R}^{J_1 \times J_2 \times J_3}$. For a single element, Tucker3 decomposition can be written as

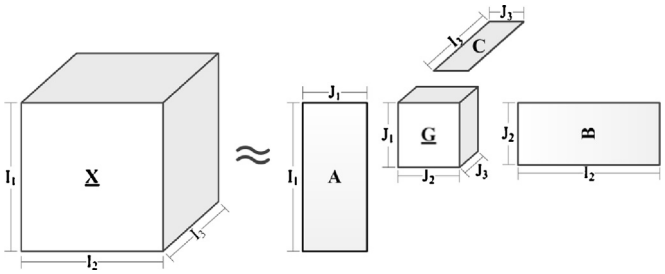

Fig. 2. Tucker3 decomposition of a three-way tensor.

$$x_{i_1 i_2 i_3} \approx \sum_{j_1=1}^{J_1} \sum_{j_2=1}^{J_2} \sum_{j_3=1}^{J_3} g_{j_1 j_2 j_3} \cdot a_{i_1 j_1} \cdot b_{i_2 j_2} \cdot c_{i_3 j_3} \tag{1}$$

with J_1, J_2, and J_3 denoting dimensions of the core. Decomposition (1) can be expressed more compactly as

$$\underline{\mathbf{X}} \approx \underline{\mathbf{G}} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C}. \tag{2}$$

In practice, it is common that the core tensor $\underline{\mathbf{G}}$ is smaller than the original tensor $\underline{\mathbf{X}}$, i.e., $J_n < I_n$, and if not stated otherwise we will assume this is fulfilled in remainder of the paper. In this case, tensor $\underline{\mathbf{X}}$ is approximated by a rank-(J_1, J_2, J_3) tensor on the right side of (2), and (assuming the factor matrices have full rank) dimensions of the core tensor $\underline{\mathbf{G}}$ determine the multilinear rank of the approximation. By approximating the original tensor $\underline{\mathbf{X}}$ with a rank-(J_1, J_2, J_3) tensor we effectively perform model reduction. This is the reason why this type of decomposition is of practical interest. Decomposition (2) of $\underline{\mathbf{X}}$ can be seen as a composition of directional bases $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ in modes 1, 2 and 3, connected through a set of weights contained in $\underline{\mathbf{G}}$. The elements in the core can then effectively be considered as features, that describe a sample contained in tensor $\underline{\mathbf{X}}$. For example, when $\underline{\mathbf{X}}$ represents RGB image the elements of the core connect spatial information (in modes 1 and 2) and spectral information (in mode 3).

The Tucker3 decomposition is flexible in modeling complex interactions within the data, since the core tensor allows interaction of factors from all modes. However, this decomposition is, in general, not unique. Proof of uniqueness exists in the special case when the core tensor is diagonal (i.e., $g_{j_1 j_2 j_3} \neq 0$ only for $j_1 = j_2 = j_3$), resulting in the canonic polyadic decomposition (CPD), also known as parallel factor analysis (PARAFAC) model, that is essentially unique (up to scale and permutation) under mild conditions [25]. For Tucker3, constraints such as nonnegativity, sparsity and orthogonality are commonly imposed on the factor matrices and the core tensor in order to attain meaningful decomposition. Particular constraints are used to narrow down the solution space and yield a virtually unique decomposition [23,24].

An example of orthogonality constrained decomposition is the higher order singular value decomposition (HOSVD), a multilinear generalization of the matrix singular value decomposition[3] (SVD) [26]. The HOSVD decomposition of the data tensor has a form (2) with $J_n = I_n$ and orthogonal factor matrices in all modes, e.g., for mode-1 factor $\mathbf{A}^T\mathbf{A} = \mathbf{I}$ holds. The factor matrix in nth mode is estimated by performing SVD of the nth mode unfolding of the data tensor, e.g., $\mathbf{A}$ is equal to the left singular vectors of the unfolding

[2] There are other definitions of rank for tensors besides the multilinear rank, for details see [23].

[3] Note that the HOSVD does not give a decomposition of the data tensor with a reduced dimensions of the core tensor. It is an exact decomposition in a set of orthonormal bases, with core tensor of equal size as the original tensor.

$\mathbf{X}_{(1)}$. After obtaining the factor matrices, the core tensor is calculated as

$$\underline{\mathbf{G}} = \underline{\mathbf{X}} \times_1 \mathbf{A}^T \times_2 \mathbf{B}^T \times_3 \mathbf{C}^T. \quad (3)$$

Often we are interested in an optimal rank-(J_1, J_2, J_3) approximation of the data tensor $\underline{\mathbf{X}}$ in the least-squares sense with $J_n < I_n$. The optimal approximation can be found by minimizing the Frobenius norm between the data tensor $\underline{\mathbf{X}}$ and its rank-(J_1, J_2, J_3) Tucker3 decomposition:

$$D[\underline{\mathbf{X}} \,||\, (\underline{\mathbf{G}}, \mathbf{A}, \mathbf{B}, \mathbf{C})] = \left\| \underline{\mathbf{X}} - \underline{\mathbf{G}} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C} \right\|_F^2 \quad (4)$$

while at the same time imposing orthogonality constraints on factors **A**, **B**, **C**. Actually, it is enough to find factor matrices that maximize function

$$g(\mathbf{A}, \mathbf{B}, \mathbf{C}) = \left\| \underline{\mathbf{X}} \times_1 \mathbf{A}^T \times_2 \mathbf{B}^T \times_3 \mathbf{C}^T \right\|_F^2,$$

and the core tensor is given with (3) [27]. Several algorithms, such as the higher order orthogonal iteration (HOOI) [27] and algorithms based on the Newton method on Grassmannians (NG) [28,29], optimize (4) through an iterative procedure. The HOOI has proved to be a "workhorse" algorithm in various applications of tensor decompositions [20]. However, note that minimization of (4) is a non-convex optimization problem with multiple local minima [27,29], and there is no guarantee of attaining the global optimum. However, in most situations, orthogonality constraints result in a virtually unique decomposition [30].

Another way to obtain rank-(J_1, J_2, J_3) approximation of the data tensor $\underline{\mathbf{X}}$ is to relax the optimality condition. In this case we can use the truncated HOSVD (trHOSVD) decomposition. This decomposition can be easily computed in two steps: (i) for each n unfold tensor to $\mathbf{X}_{(n)}$, and calculate the standard SVD, i.e., $\underline{\mathbf{X}}_{(n)} = \mathbf{U}^{(n)}\mathbf{S}^{(n)}\mathbf{V}^{(n)T}$. The orthogonal factor **A** is formed by J_1 leading left singular vectors of $\mathbf{X}_{(1)}$, i.e., the first J_1 columns of $\mathbf{U}^{(1)}$. Analog procedure is performed for other modes; (ii) compute the core tensor using (3). Since this decomposition is calculated using matrix SVD for each of the modes, it can be very fast, even for large problems. Since it is non-iterative there are no problems with local minima. Although truncated HOSVD (as opposed to truncated SVD in matrix case) does not give the optimal rank-(J_1, J_2, J_3) decomposition, it is a good approximation with upper bound on error given in [26]. This type of low-rank approximation was useful in various applications. For comparison, we performed several tests comparing trHOSVD, HOOI and NG for feature extraction, with no significant difference in overall classification performance. Therefore, we used trHOSVD in the experimental section, as reported in Section 4.

2.2. Dimensions of the core tensor

One of the essential problems when performing Tucker decomposition is the size of the core tensor. Since in our approach elements of the core tensor $\underline{\mathbf{G}} \in \mathbb{R}^{J_1 \times J_2 \times J_3}$ will represent features, the number of features directly depends on dimensions of the core tensor in (2). Estimation of multilinear rank of a tensor is essential, but still very difficult problem in multi-way analysis. This problem is also known as the intrinsic dimensionality problem or model order selection. Several methods have been developed recently, with various applications in signal processing and data analysis [31]. Estimation of rank in each of the modes can be based on the gap in a sequence of parameters, e.g., eigenvalues of the covariance or correlation matrix. Some of the methods in this group are GAP [32,33], RAE (ratio of adjacent eigenvalues) [34], SORTE (second order statistic of the eigenvalues) [32,33], RAESORTE [34], and EIF (empirical indicator function) [35,36]. These methods estimate mode-n rank from the unfolding $\mathbf{X}_{(n)}$ by determining the number of dominant eigenvalues of the Gramm matrix $\mathbf{X}_{(n)}\mathbf{X}_{(n)}^T$, e.g., by searching for a gap between adjacent eigenvalues. Also, information theoretic avenue can be used, resulting in model order selection methods based on Akaike's information criterion (AIC) [37], Kullback–Leibler information criterion (KIC) [38], and minimum description length (MDL) [39]. When applied to tensors, these methods analyze corresponding unfolding $\mathbf{X}_{(n)}$ to estimate the intrinsic dimension in mode-n. Other approaches can be derived using Bayesian estimation, an example being automatic relevance determination (ARD) method for multi-way models [40]. The ARD method directly estimates dimensions of the core for the Tucker decomposition.

All of the mentioned methods have been used in diverse applications, such as detection of number of clusters [33], dimensionality reduction using principal component analysis (PCA) [41], and the choice of number of sources in linear model [42], as well as in feature extraction [20]. However, due to different theoretical assumptions and noise in the real-world data, they often return significantly different estimations when applied to the same problem [34]. Another related problem arises from the fact that clinical images of skin lesions used in our experiments do not have equal sizes, i.e., dimensions in modes 1 and 2 are not the same for all images. This causes large variation among estimated dimensions of the core tensor for different samples. Additionally, the mentioned methods were designed to find a low-rank approximation with small n-ranks and good fit (small residual between the original tensor and its low-rank approximation). This aim is not of direct importance in feature extraction, since we seek for a decomposition that results in good classification performance. Therefore, dimensions of the core tensor are often determined through cross-validation procedure by estimating classification performance for various decompositions. However, we tested several methods for automatic n-rank estimation, with results presented and commented in more detail in the experimental section of the paper.

2.3. Nonlinear transformation

Representation of the color image by multilinear decomposition (2) can be questioned on the basis of several arguments, see [43,44] for multi- and hyperspectral remote sensing. Taking into account possible nonlinear relations in the data is expected to improve final classification accuracy. This assumption is supported by the Cover's theorem [45]. In a nutshell, this theorem states that a set of samples, that are not linearly separable in the original (low-dimensional) space, are more likely to be linearly separable after being nonlinearly mapped into some higher-dimensional space. Consequently, it is expected that linear relations hold with high probability for the data nonlinearly mapped in a higher-dimensional space [45].

In this regard, an implicit nonlinear transformation of the original image into reproducible kernel Hilbert space (RKHS) is performed, based on kernel techniques. Description of the kernel-based nonlinear transformation will be given here due to completeness, but for a detailed theoretical treatment interested reader is referred to [46]. Let $k : S \times S \to \mathbb{R}$ be a real, positive definite kernel, and $S \subseteq \mathbb{R}^I$ a nonempty set. Then map $\Phi : S \to \mathbb{R}^S$ can be defined as $\mathbf{s} \mapsto \Phi(\mathbf{s}) := k(., \mathbf{s})$, where $\mathbb{R}^S := \{f : S \to \mathbb{R}\}$ is a set of all functions from S to $\mathbb{R}$. In this way each input sample $\mathbf{s} \in S$ is mapped to a function $\Phi(\mathbf{s})$ defined on the input space, which is possibly an infinite dimensional object. However, it is possible to obtain an approximation of the map Φ by evaluating it only on a set of points, since the transformed data is embedded in a subspace of RKHS [46,47]. Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_D\} \subset S$ be a set of points in the input space. Then transformation $\Phi_D : S \to \mathbb{R}^D$ defined as

$$\mathbf{s} \mapsto \Phi_D := [k(\mathbf{v}_1, \mathbf{s}), \ldots, k(\mathbf{v}_D, \mathbf{s})]^T$$

is called the empirical kernel map with respect to a set of points $\{\mathbf{v}_1, \ldots, \mathbf{v}_D\}$ [46]. In this way, nonlinearly transformed data is projected into a D dimensional subspace of RKHS, while at the same time all calculations are performed on the original data from the input set $S \subseteq \mathbb{R}^I$ [46]. Obviously, $D \gg I$ must hold to exploit benefits stipulated by the Cover's theorem. Complex problem of selecting the points $\{\mathbf{v}_1, \ldots, \mathbf{v}_D\}$ is known as basis selection. Various basis selection methods were previously proposed, such as kernel PCA [48] and feature vector selection [49]. In experiments we use *k-means* clustering, with basis vectors estimated as centroids of $k = D$ clusters, as in [47].

In this paper we perform nonlinear transformation of the sample tensor $\underline{\mathbf{X}}$ along the mode 3, i.e., each mode-3 fiber (corresponding to a single pixel), is nonlinearly mapped to a new D dimensional vector, $\mathbf{x}_{i_1 i_2 :} \mapsto \Phi_D(\mathbf{x}_{i_1 i_2 :}) = [k(\mathbf{v}_1, \mathbf{x}_{i_1 i_2 :}), \ldots, k(\mathbf{v}_D, \mathbf{x}_{i_1 i_2 :})]^T$. Nonlinear mapping for tensor objects

$$\Phi_D : \mathbb{R}^{I_1 \times I_2 \times I_3} \to \mathbb{R}^{I_1 \times I_2 \times D}$$

$$\underline{\mathbf{X}} \mapsto \Phi_D(\underline{\mathbf{X}})$$

that maps a three-way tensor $\underline{\mathbf{X}}$ to a new three-way tensor $\Phi_D(\underline{\mathbf{X}})$, is defined by replacing each mode-3 fiber $\mathbf{x}_{i_1 i_2 :}$ with its image $\Phi_D(\mathbf{x}_{i_1 i_2 :})$.[4] As mentioned, basis points $\{\mathbf{v}_1, \ldots, \mathbf{v}_D\}$ are estimated by *k-means* clustering of mode-3 fibers $\{\mathbf{x}_{i_1 i_2 :}\}$ into D clusters. Now, Tucker3 decomposition can be used to represent nonlinearly transformed data as:

$$\Phi_D(\underline{\mathbf{X}}) \approx \underline{\mathbf{G}} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C}. \tag{5}$$

Since $D \gg I_3$ the core tensor in decomposition (5) can have greater dimension J_3 than the core tensor in (2). In the experimental section we use Gaussian kernel $k(\mathbf{x}, \mathbf{y}) = \exp(-\|\mathbf{x} - \mathbf{y}\|^2 / \sigma^2)$. Also, we experimented with other kernels, such as Laplacian $k(\mathbf{x}, \mathbf{y}) = \exp(-\|\mathbf{x} - \mathbf{y}\| / \sigma^2)$ and polynomial $k(\mathbf{x}, \mathbf{y}) = (1 + \langle \mathbf{x}, \mathbf{y} \rangle)^d$, but this did not result in better results than with the Gaussian kernel.

3. Proposed feature extraction method

The proposed feature extraction scheme is summarized in Algorithm 1. Let us assume that our dataset consists of K samples (e.g., RGB images of skin lesions), each represented by a three-way tensor $\underline{\mathbf{X}}^{(k)}$, $k = 1, \ldots, K$. Also, let a class label $c^{(k)}$ be assigned to each sample, e.g., $c^{(k)} = 1$ if kth sample belongs to class 1 (benign nevus), and $c^{(k)} = 2$ if it is from class 2 (melanoma). The proposed feature extraction procedure consists of nonlinear transformation of each sample $\underline{\mathbf{X}}^{(k)}$, followed by tensor decomposition using trHOSVD algorithm. Core tensor is calculated directly from the nonlinearly transformed sample by projecting the tensor $\Phi_D(\underline{\mathbf{X}}^{(k)})$ onto factors $\mathbf{A}^{(k)}$, $\mathbf{B}^{(k)}$ and $\mathbf{C}^{(k)}$ obtained by the trHOSVD. Finally, features are obtained by rearranging elements of the core $\underline{\mathbf{G}}^{(k)}$ into a vector $\mathbf{g}^{(k)}$, since standard classifiers use vector representation of the samples.[5] Note that feature extraction is performed on each sample $\underline{\mathbf{X}}^{(k)}$ separately, but with the same parameters (σ, D) for the nonlinear transform and fixed dimensions of the core (J_1, J_2, J_3). In this way, for each image we obtain the same number of features, equal to the number of elements of the core tensor.

From feature extraction point of view it is important that the core tensor $\underline{\mathbf{G}}^{(k)}$ connects two spatial modes with spectral mode of the transformed image. In this way, the core tensor represents a spatial-spectral profile of the lesion present in the image $\underline{\mathbf{X}}^{(k)}$ while each element $g^{(k)}_{j_1 j_2 j_3}$ of the core tensor can be viewed as a spatial-spectral characteristic of the lesion. This can be seen from the following expression

$$g^{(k)}_{j_1 j_2 j_3} = \Phi_D(\underline{\mathbf{X}}^{(k)}) \times_1 (\mathbf{a}^{(k)}_{j_1})^T \times_2 (\mathbf{b}^{(k)}_{j_2})^T \times_3 (\mathbf{c}^{(k)}_{j_3})^T,$$

where $\mathbf{a}^{(k)}_j$ is jth column of $\mathbf{A}^{(k)}$. Feature $g^{(k)}_{j_1 j_2 j_3}$ is obtained by projecting tensor $\Phi_D(\underline{\mathbf{X}}^{(k)})$ onto appropriate columns of factor matrices. Therefore, factor matrices in (4) can be seen as directional projection matrices for feature extraction in each of the modes. Consequently, we conjectured that the core tensor viewed as a set of features can be used for robust and accurate noninvasive diagnosis of melanoma.

Proposed feature extraction method differs from the one recently published in [20], where simultaneous tensor decomposition is performed for all samples in the training set to find projection matrices. Consequently, it is required that all images in the data set have the same dimensions (both in spatial and spectral modes). However, this requirement is not easily met in practice, and method proposed here allows greater flexibility when acquiring clinical images for analysis. Also, separate decomposition of each sample enables us to easily update training set when new labeled sample is available, since it is decomposed independently of other samples previously present in the training set.

Introduction of nonlinear transformation prior to tensor decomposition was motivated by practical experience. We performed several experiments with direct decomposition of the original images and results were not satisfactory. Since spectral dimension of RGB images is very low, we introduced a nonlinear transformation that increases dimension in the mode-3, as described in Section 2.3. This resulted in a significant increase of performance, with results reported in experimental section.

Algorithm 1. Proposed method for feature extraction.

Input: three-way tensor $\underline{\mathbf{X}}^{(k)} \in \mathbb{R}_{0+}^{I_1 \times I_2 \times I_3}$, autofluorescent RGB image
Parameters: kernel map (σ, D), core size (J_1, J_2, J_3)
Do:
• nonlinear transformation
$\underline{\mathbf{X}}^{(k)} \mapsto \Phi_D(\underline{\mathbf{X}}^{(k)})$
• tensor decomposition
$\Phi_D(\underline{\mathbf{X}}^{(k)}) \approx \underline{\mathbf{G}}^{(k)} \times_1 \mathbf{A}^{(k)} \times_2 \mathbf{B}^{(k)} \times_3 \mathbf{C}^{(k)}$
• feature extraction
$\underline{\mathbf{G}}^{(k)} = \Phi_D(\underline{\mathbf{X}}^{(k)}) \times_1 (\mathbf{A}^{(k)})^T \times_2 (\mathbf{B}^{(k)})^T \times_3 (\mathbf{C}^{(k)})^T$
• vectorization of the core tensor
$\hat{\mathbf{g}}^{(k)} = vec(\underline{\mathbf{G}}^{(k)})$
Output: features $\hat{\mathbf{g}}^{(k)} \in \mathbb{R}^{J_1 J_2 J_3}$

Classification system built around the proposed method is displayed in Fig. 3. Feature extraction is performed for each sample in the data set with the same values of parameters σ, D, and J_n, $n = 1, 2, 3$. In this way all of the K core tensors have the same dimensions and for each sample we obtain $L = J_1 \cdot J_2 \cdot J_3$ features used for classification. Validation step consists of partitioning the K samples $(\mathbf{g}^{(k)}, c^{(k)})$ into two sets: K_{TR} training samples $(\mathbf{g}^{(k)}_{TR}, c^{(k)}_{TR})$ and K_{TE} test samples $(\mathbf{g}^{(k)}_{TE}, c^{(k)}_{TE})$. In the experimental section we used linear support vector machine (linSVM) classifier, and nonlinear SVM with the Gaussian kernel (rbfSVM) and polynomial kernel (polySVM), although other classifiers can be used as well.

In order to obtain a reliable estimate of classification performance on an independent data set, we performed a two-fold cross-validation (CV). In general, CV is performed by using some samples from the data set to train the classifier, and remaining samples to test it [4]. In K-fold CV the data set is divided into K equally sized parts so that each contains the same proportion of all class labels. In the kth step classifier is trained on $K - 1$ parts of the data

[4] This is actually a slight abuse of notation, since map Φ_D was originally defined for vectors, not tensors. The notation $\Phi_D(\underline{\mathbf{X}})$ can also be interpreted in the following sense: we take mode-3 unfolded tensor $\mathbf{X}_{(3)}$, perform nonlinear map of each column, and then rearrange columns back to a three-way tensor.

[5] The vectorization procedure was performed by stacking mode-1 fibers of the core tensor. However, ordering of the elements of the core tensor is not important, as long as the same procedure is used for each sample.

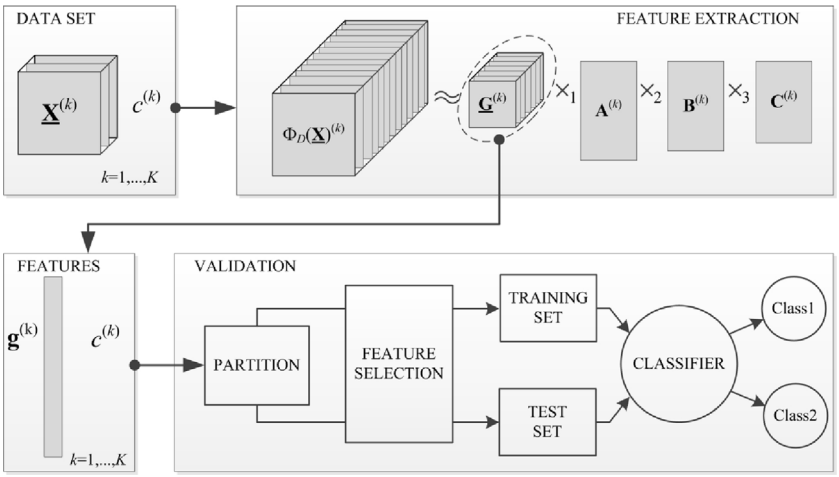

Fig. 3. Flow chart diagram for classification system based on the proposed method.

Fig. 4. Examples of clinical color images: three benign nevi (left) and three malignant melanomas (right). Border of the cropped region used for analysis is also displayed.

set and tested on kth part of the data set. The procedure is repeated for $k = 1, \ldots, K$ and the K results are combined to calculate the final estimate of the classification performance. For example, in case of a two-fold CV samples are divided into two equal sized data sets, each consisting of 50% of samples from class 1, and 50% of samples from class 2. In order to obtain a more reliable estimate of prediction performance, CV is repeated several times and the validation results are averaged.

Even though the number of features L is significantly smaller than the number of elements in each of the tensors $\Phi_D(\underline{\mathbf{X}}^{(k)})$, it can still be large. Thus it is reasonable to perform a feature selection (FS) step to identify significant features obtained by the proposed method. We ranked features based on the Fisher score [20], albeit other information criteria can also be used for ranking [50]. For each feature Fisher score is calculated using samples in the training set as:

$$\varphi(i) = \frac{K_{TR}^1(\bar{g}_{TR,i}^1 - \bar{\bar{g}}_{TR,i})^2 + K_{TR}^2(\bar{g}_{TR,i}^2 - \bar{\bar{g}}_{TR,i})^2}{\sum_{k=1}^{K_{TR}}(g_{TR,i}^{(k)} - \bar{g}_{TR,i}^{c_{TR}^{(k)}})}$$

where K_{TR}^c denotes the number of training samples belonging to the cth class ($c = 1, 2$), $g_{TR,i}^{(k)}$ is ith feature of kth training sample, $c_{TR}^{(k)}$ label of the kth sample, $\bar{g}_{TR,i}^c$ is the mean of the ith feature calculated over training samples from class c, and $\bar{\bar{g}}_{TR,i}$ is the mean of the ith feature calculated over all training samples. Larger value of Fisher score means that the feature is more discriminative. Feature selection is performed by selecting a number of features with largest Fisher score.

4. Experiments

The proposed method for feature extraction is demonstrated on noninvasive diagnosis of melanoma. More specifically, in the experiments[6] we test discrimination between benign nevus and melanoma based on autofluorescent clinical RGB images. The images of benign nevus and superficial melanoma were made available by courtesy of [51–54]. The data set for the experiments consisted of 180 clinical RGB images with 90 samples of benign nevus and 90 samples of superficial melanoma. For illustration, few examples of analyzed images are displayed in Fig. 4.

Since images were collected from various archives, we had no control over parameters of the image acquisition. Images in the data set differed in number of pixels and aspect ratio, depending on the size of the lesion and parameters of the acquisition procedure. Before performing experiments, each image was manually cropped around the lesion, to prevent extracting features from the non-tumor related areas of the image. The cropped images were used, to show that the proposed method is robust when a small amount of image is occupied by healthy skin. A more sophisticated approach would be to use segmentation to extract the part of the image corresponding to the lesion, as in [17], but it would introduce additional computational cost, and remains to be done in future experiments. Additionally, it is important to note that no hair removal was performed, since the images did not contain a large amount of hair. Also, the feature extraction should be robust to imperfect preprocessing, e.g., to a small amount of healthy skin or hair in the image. At the end, mode-1 and mode-2 dimensions of the images were in range $I_{1,\min} = 38$ to $I_{1,\max} = 912$ and $I_{2,\min} = 36$ to $I_{2,\max} = 671$.

To demonstrate performance of the proposed method we compared it against an approach based on independent component analysis (ICA) [17]. As reported in [9], the ICA-based scheme performed better than the twelve studies given in Table 1. This approach uses separate spectral and spatial features of skin lesion

[6] All calculations were carried out in MATLAB environment running on a Windows 7 PC.

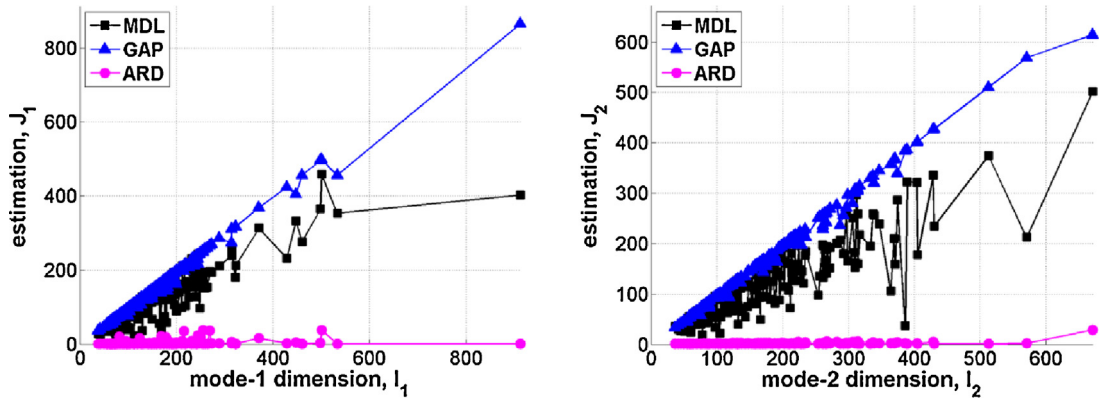

Fig. 5. Estimation of mode-1 and mode-2 rank for images in the data set using MDL (square), GAP (triangle) and ARD (circle): estimated rank vs. mode-1 dimension (left); and estimated rank vs. mode-2 dimension (right).

present in the image. Spectral features for each sample were calculated as mean and variance of red, green and blue channel from RGB image of the lesion. In order to obtain spatial characteristics of the lesion, original RGB image is converted to grayscale. Then spatial features are extracted from the grayscale image using ICA-learned filters. Firstly, a set of patches (16×16 pixels) was selected from the images in the training set. The preprocessing step was performed, including mean removal for each patch, followed with data whitening and dimensionality reduction using PCA. After preprocessing, as described in [17], ICA was used to obtain spatial filters as rows of the unmixing matrix obtained from a set of patches, for nevus and melanoma independently. Specifically, 100 ICA filters were obtained for melanoma and 100 filters for benign nevus, resulting in a filter bank with 200 filters. Finally, spatial features were calculated as the energy of the response of each grayscale image to a filter. The original paper by Tabatabaie and Esteki [17] used FastICA algorithm [55] with *tanh* nonlinearity to estimate spatial filters. However, there is a possible problem with stability of the FastICA algorithm. Since the algorithm uses a random initial point for the optimization of the cost function, it is possible to end up in a different local optimum, depending on the initial point. To alleviate this problem we used *Icasso* software package [56] that runs FastICA algorithm M times with different initial points, and the unmixing matrix (i.e., filters) is obtained by clustering the estimated unmixing matrices from all runs.[7] In our experiments for ICA-based method we used $M = 30$ restarts. To make the comparison fair, ICA approach has been applied on the same images as the proposed tensor-based feature extraction method, i.e., no hair removal preprocessing has been done. The reason is that the presence of hair was very limited and we believe that a good feature extraction method should exhibit robustness without requiring ideal preprocessing. Also, no image segmentation has been performed to separate lesion from the healthy skin. The reason is that the images of the melanoma and nevus were already cropped around the lesion area and, thus, area occupied by the healthy skin was small.

The classification performance was measured in terms of sensitivity and specificity, which are standard metrics for binary classification. Higher sensitivity and specificity mean better classification performance. For our experiments accuracy of classification can be calculated as the arithmetic mean of the two.

For start, we tested several methods for automatic rank selection (i.e., dimensions of the core tensor) described in Section 2.3. Estimations of rank in mode-1 and mode-2 for images in the data set using MDL, ARD and GAP method are displayed in Fig. 5. As expected, estimation results highly depend on the dimensions of the clinical image. This is not acceptable since the proposed method requires that the core tensors have the same dimensions for all samples. Additionally, different methods produced very different results for the same image. As can be seen in Fig. 5, GAP method estimation is very close to the dimension of the image, while ARD method gives very small estimation. Results for other methods fall in between these two extreme cases (only MDL is displayed in the image).

Due to this we could not find optimal dimensions of the core tensor either by a single method or as consensus of several methods. Thus, we employed wrapper-like procedure for determination of the optimal dimensions: performance of the classification was estimated through cross-validation for different dimensions of the core tensor and dimensions that produced the best result were selected as optimal.[8] To be more precise, feature extraction was performed for dimensions of the core from a grid $1 \leq J_1 \leq I_{1,\min}$, $1 \leq J_2 \leq I_{2,\min}$ and $1 \leq J_3 \leq D$, with classification performance for each point estimated using CV with 30 random partitions. Dimensions that produced the smallest CV error were selected for further analysis. Experiments also showed that feature extraction method is not sensitive to choice of parameter D, as long as it is large enough (in comparison with $I_3 = 3$) so we fixed it to $D = 40$, while kernel parameter was set to $\sigma = 0.8$ through CV. These parameters for nonlinear transformation were used in further experiments.

This procedure led to the best classification performance for dimensions of core tensor (10,12,12) and (11,12,10). We selected these dimensions as optimal, but there were several other combinations with just slightly inferior performance. The proposed method was thoroughly evaluated for core tensor with selected optimal dimensions through two-fold CV with 100 random partitions. The obtained results can be found in Table 2 in terms of sensitivity and specificity (mean $\pm$ standard deviation). Overall, equal accuracy of 84.5% using the whole core tensor was obtained with dimensions (10,12,12) and (11,12,10).

Since the number of features for selected dimensions of the core was still quite large (above thousand), we experimented with an additional feature selection step. Elements of the feature vector were ranked in descending order based on their Fisher score, and only significant features were selected for classification. Results of this analysis for core dimensions (10,12,12) and (11,12,10) are displayed in Fig. 6. It can be seen that in the beginning adding more features improved performance significantly, but after a certain number accuracy was not drastically increased. Thus, selection step can be used to reduce the number of features without affecting the

[7] Although we used *Icasso*, the obtained results did not differ significantly compared with application of FastICA without stability analysis.

[8] To measure performance we used rbfSVM as classifier, with parameter σ_C tuned through CV.

Table 2
Performance of the proposed method (without feature selection) vs. ICA-based method.[a]

	linSVM		polySVM		rbfSVM	
	Sens. [%]	Spec. [%]	Sens. [%]	Spec. [%]	Sens. [%]	Spec. [%]
Proposed method Core(10,12,12)	20.0 ± 6.2	96.2 ± 3.1	8.3 ± 4.1 $d = 2$	99.4 ± 1.1	81.8 ± 5.2 $\sigma_C = 18$	87.2 ± 4.1
Proposed method Core(11,12,10)	18.4 ± 6.7	95.0 ± 3.2	8.2 ± 4.1 $d = 2$	99.8 ± 0.6	82.1 ± 5.2 $\sigma_C = 19$	86.9 ± 4.2
ICA-based method	79.9 ± 5.6	79.1 ± 6.8	57.2 ± 13.2 $d = 5$	58.9 ± 17.4	75.7 ± 7.4 $\sigma_C = 3.4$	72.4 ± 6.5

[a] Performance estimated by two-fold CV with 100 random partitions (mean ± standard deviation). Parameters d (degree of the polynomial) and σ_C were selected through CV.

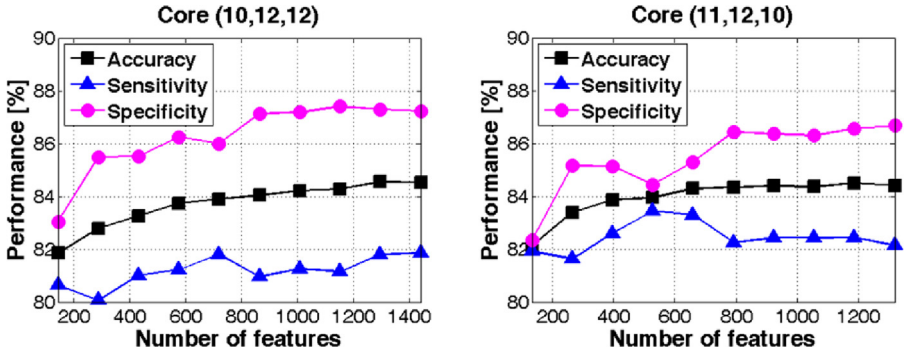

Fig. 6. Classification performance vs. number of features. Feature selection based on Fisher score was applied on the core tensor with dimensions: (10,12,12) (left) and (11,12,10) (right). SVM classifier with Gaussian kernel (rbfSVM) was used with width of the kernel tuned through CV.

overall classification performance. For dimensions of the core in Fig. 6 we obtained almost the same performance with only 40% of the most significant features.

The ICA-based approach yielded sensitivity of 79.9% and specificity of 79.1%, exhibiting a significant decrease in performance compared to the original paper [17]. The reasons for that could possibly be a larger and more diverse dataset used here, and the rigorous two-fold CV used in the experiments. Overall, the proposed tensor decomposition-based approach outperformed the ICA-based approach. Although estimated accuracy depends on the data set, we believe that the two-fold performance assessment presented here provides a quite realistic estimate. Especially with diversity of the data set taken into account, and the fact that there was no control over the acquisition procedure. As opposed to the studies presented in Table 1, we have estimated performance in statistically rigorous manner, by performing two-fold cross-validation procedure. Additionally, performance of the proposed approach is also comparable with the one reported in [57], that uses melanoma specific marker expressions extracted from biopsy tissues, where (depending on the type of gene marker) area under the curve varied between 0.94 and 0.5.

5. Conclusions

Tensor representation of a multi-spectral image enables extraction of features that simultaneously capture spatial and spectral characteristics of object present in the image. Proposed feature extraction is performed through Tucker3 decomposition of nonlinearly mapped image. The method relies on well-known computational procedures, such as singular value decomposition, that can be efficiently implemented in embedded devices. As demonstrated, the proposed method compares favorably with the state of the art results reported in the literature. In perspective, an automated system based on a cost-effective, and nowadays ubiquitous, RGB camera could possibly be used for preliminary screening in distant areas and areas without appropriate medical care. The software with diagnostic algorithm could be implemented for some of the widespread smartphone platforms or for an application-specific low-cost device.

However, several directions are opened for future research. One possibility is to include additional post-processing of the extracted features. Replacing the feature selection step with an additional feature extraction step could possibly result in better performance. In this case one could apply linear techniques (e.g., linear discriminant analysis), or tensor-based feature extraction [20]. It would be interesting to perform additional experiments using a larger dataset, consisting of images acquired under controlled conditions. Classification performance in different (controlled) scenarios would give a better assessment of the real-world performance, and could provide valuable insight and directions for future development. An important contribution to computer-based analysis of pigmented skin lesions would be to build a benchmark data set with clinical images, which could be used for direct comparison of diagnostic methods.

Acknowledgments

Work of A. Jukić and I. Kopriva has been supported through grant 098-0982903-2558 funded by the Ministry of Science, Education and Sports, Republic of Croatia.

References

[1] M. Elbaum, A.W. Kopf, H.S. Rabinovitz, R.G. Langley, H. Kamino, M.C. Mihim Jr., et al., Automatic differentiation of melanoma from melanocytic nevi with multispectral digital dermoscopy: a feasibility study, Journal of the American Academy of Dermatology 44 (2) (2001) 207–218.
[2] S.G. Kong, L.-J. Park, Hyperspectral image analysis for skin tumor detection, in: R.I. Hammoud (Ed.), Augmented Vision Perception in Infrared: Algorithms and Applied Systems, Springer-Verlag, London, 2009 (Chapter 7).
[3] I. Guyon, J. Weston, S. Barnhill, V. Vapnik, Gene selection for cancer classification using support vector machine, Machine Learning 46 (1-3) (2002) 389–422.

[4] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer, New York, 2009 (Chapter 18).
[5] W. Stolz, O. Braun-Falco, P. Bilek, M. Landthaler, W.H.C. Burgdorf, A.B. Cognetta (Eds.), Color Atlas of Dermatoscopy, second ed., Blackwell Publishing, Berlin, 2002.
[6] A. Green, N. Martin, G. McKenzie, J. Pfitzner, F. Quintarelli, B.W. Thomas, et al., Computer image analysis of pigmented skin lesions, Melanoma Research 1 (4) (1991) 231–236.
[7] E.M. Wurm, H.P. Soyer, Scanning for melanoma, Australian Prescriber 33 (2010) 150–155.
[8] A.W. Kopf, B. Welkovich, R.E. Frankel, E.J. Stoppelmann, R.S. Bart, G.S. Rogers, et al., Thickness of malignant melanoma: global analysis of related factors, Journal of Dermatologic Surgery and Oncology 13 (4) (1987).
[9] B. Rosado, S. Menzies, A. Harbauer, H. Pehamberger, K. Wolff, M. Binder, et al., Accuracy of computer diagnosis of melanoma – a quantitative meta-analysis, Archives of Dermatology 139 (2003) 361–367.
[10] T. Schindewolf, W. Stolz, R. Albert, W. Abmayr, H. Harms, Classification of melanocytic lesions with color and texture analysis using digital image processing, Analytical & Quantitative Cytology & Histology 15 (1993) 1–11.
[11] H.P. Soyer, J. Smolle, H. Kerl, H. Steettnre, Early diagnosis of malignant melanoma by surface microscopy, Lancet 330 (8562) (1987) 803.
[12] G. Argenziano, H.P. Soyer, Dermoscopy of pigmented lesions – a valuable tool for early diagnosis of melanoma, Lancet Oncology 2 (7) (2001) 443–451.
[13] H. Ganster, P. Pinz, R. Rohrer, E. Wildling, M. Binder, H. Kittler, Automated melanoma recognition, IEEE Transactions on Medical Imaging 20 (4) (2001) 233–239.
[14] K. Hoffmann, T. Gambichler, A. Rick, M. Kreutz, M. Anschuetz, T. Grünenduck, et al., Diagnostic and neural analysis of skin cancer (DANAOS). A multicentre study for collection and computer-aided analysis of data from pigmented skin lesions using digital dermoscopy, British Journal of Dermatology 149 (4) (2003) 801–809.
[15] S.W. Menzies, L. Bischof, H. Talbot, A. Gutenev, M. Avramidis, L. Wong, et al., The performance of SolarScan – an automated dermoscopy image analysis instrument for the diagnosis of primary melanoma, Archives of Dermatology 141 (11) (2005) 1388–1396.
[16] H. Iyatomi, H. Oka, M.E. Celebi, M. Hashimoto, M. Hagiwara, M. Tanaka, et al., An improved Internet-based menaloma screening system with dermatologist-like tumor area extraction algorithm, Computerized Medical Imaging and Graphics 32 (2008) 566–579.
[17] K. Tabatabaie, A. Esteki, Independent component analysis as an effective tool for automated diagnosis of melanoma, in: Proceedings of the IEEE Biomedical Engineering Conference (CIBEC'08), Cairo, 2008, pp. 1–4.
[18] K. Korotkov, R. Garcia, Computerized analysis of pigmented skin lesions: a review, Artificial Intelligence in Medicine 56 (2) (2012) 69–90.
[19] A. Cichocki, R. Zdunek, A.H. Phan, S.I. Amari, Nonnegative Matrix and Tensor Factorization, John Wiley & Sons, Chichester, 2009.
[20] A.H. Phan, A. Cichocki, Tensor decompositions for feature extraction and classification of high dimensional datasets, IEICE Nonlinear Theory and its Applications 1 (2010) 37–68.
[21] H.A.L. Kiers, Towards a standardized notation and terminology in multiway analysis, Journal of Chemometrics 14 (3) (2000) 105–122.
[22] I. Kopriva, A. Cichocki, Blind multi-spectral image decomposition by 3D nonnegative tensor factorization, Optics Letters 34 (14) (2009) 2210–2212.
[23] T.G. Kolda, B.W. Bader, Tensor decompositions and applications, SIAM Review 51 (3) (2009) 455–500.
[24] L.R. Tucker, Some mathematical notes on three-mode factor analysis, Psychometrica 31 (1966) 279–311.
[25] J.B. Kruskal, Three-way arrays: rank and uniqueness of trilinear decompositions, Linear Algebra and its Applications 18 (2) (1977) 95–138.
[26] L. De Lathauwer, B. De Moor, J. Vandewalle, A multilinear singular value decomposition, SIAM Journal on Matrix Analysis and Applications 21 (4) (2000) 1253–1278.
[27] L. De Lathauwer, B. De Moor, J. Vandewalle, On the best rank-1 and rank-(R1,R2,. . .,RN) approximation of higher-order tensors, SIAM Journal on Matrix Analysis and Applications 21 (4) (2000) 1324–1342.
[28] L. Elden, B. Savas, A Newton–Grassmann method for computing the best multi-linear rank-(r1,r2,r3) approximation of a tensor, SIAM Journal on Matrix Analysis and Applications 31 (2009) 248–271.
[29] B. Savas, L.-H. Lim, Quasi-Newton methods on Grassmannians and multilinear approximations of tensors, SIAM Journal on Scientific Computing 32 (6) (2010) 3352–3393.
[30] D. Luo, C. Ding, H. Huang, Are tensor decomposition solutions unique? On the global convergence HOSVD and ParaFac algorithms, in: J.Z. Huang, L. Cao, J. Srivastava (Eds.), Proceedings of the 15th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining, Shenzhen, China, 2011, pp. 148–159.
[31] P. Stoica, Y. Selen, Model-order selection: a review of information criteria rules, IEEE Signal Processing Magazine 21 (4) (2004) 36–47.
[32] Z. He, A. Cichocki, S. Xie, Efficient method for Tucker3 model selection, Electronics Letters 45 (15) (2009) 805–806.
[33] Z. He, A. Cichocki, S. Xie, K. Choi, Detecting the number of clusters in n-way probabilistic clustering, IEEE Transactions on Pattern Analysis and Machine Intelligence 32 (11) (2010) 2006–2021.
[34] F. Cong, A.K. Nandi, H. Zhaoshui, A. Cichocki, T. Ristaniemi, Fast and effective model order selection method to determine the number of sources in a linear transformation model, in: Proceedings of the 2012 European Signal Processing Conference, Bucharest, Romania, 2012, pp. 1870–1874.
[35] E.R. Malinowski, Theory of error in factor analysis, Analytical Chemistry 49 (1977) 606–612.
[36] E.R. Malinowski, Determination of the number of factors and experimental error in data matrix, Analytical Chemistry 49 (1977) 612–617.
[37] H. Akaike, A new look at the statistical model identification, IEEE Transactions on Automatic Control 19 (6) (1974) 716–723.
[38] J.E. Cavanaugh, A large-sample model selection criterion based on Kullback's symmetric divergence, Statistics & Probability Letters 42 (1999) 333–344.
[39] J. Rissanen, Modeling by shortest data description, Automatica 14 (1978) 465–471.
[40] M. Morup, L.K. Hansen, Automatic relevance determination for multi-way models, Journal of Chemometrics 23 (2008) 352–363.
[41] T.P. Minka, Automatic choice of dimensionality for PCA, in: T.K. Leen, T.G. Dietterich, V. Tresp (Eds.), Advances in Neural Information Processing Systems 13, Denver, 2000, pp. 514–520.
[42] F. Cong, Z. He, J. Hämäläinen, A. Cichocki, T. Ristaniemi, Determining the number of sources in high-density EEG recordings of event-related potentials by model order selection, in: Proceedings of the IEEE International Workshop on Machine Learning for Signal Processing (MLSP), Beijing, China, 2011, pp. 1–6.
[43] J.M.P. Nascimento, J.M. Bioucas Dias, Does independent component analysis play a role in unmixing hyperspectral data? IEEE Transactions on Geoscience and Remote Sensing 43 (2005) 175–187.
[44] T. Han, D.G. Goodenough, Investigation of nonlinearity in hyperspectral imagery using surrogate data methods, IEEE Transactions on Geoscience and Remote Sensing 46 (2008) 2840–2847.
[45] T.M. Cover, Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition, IEEE Transactions on Electronic Computers EC-14 (3) (1965) 326–334.
[46] B. Scholkopf, A.J. Smola, Learning with Kernels, MIT Press, Cambridge, MA, 2002.
[47] S. Hameling, A. Ziehe, M. Kawanabe, Kernel-based nonlinear blind source separation, Neural Computation 15 (2003) 1089–1124.
[48] B. Schölkopf, A.J. Smola, R.K. Müller, Nonlinear component analysis as a kernel eigenvalue problem, Neural Computation 10 (1998) 1299–1319.
[49] G. Baudat, F. Anouar, Kernel-based methods and function approximation, in: Proceedings of International Joint Conference on Neural Networks, Washington, DC, 2001, pp. 1244–1249.
[50] A.K. Jain, R.P. Duin, J. Mao, Statistical pattern recognition: a review, IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (2000) 4–37.
[51] T.L. Diepgen, G. Yihune, et al. Dermatology Online Atlas. [Online]. http://www.dermis.net (accessed: January 2013).
[52] DermAtlas. [Online]. http://dermatlas.org (accessed: August 2012).
[53] An Atlas of Clinical Dermatology. [Online]. http://dandem-pdv.is.kkh.dk/atlas (accessed: August 2012).
[54] DermNetNZ. [Online]. http://dermnet.org.nz/lesions (accessed: August 2012).
[55] A. Hyvärinen, E. Oja, A fast fixed-point algorithm for independent component analysis, Neural Computation 9 (7) (1997) 1483–1492.
[56] J. Himberg, A. Hyvärinen, F. Esposito, Validating the independent components of neuroimaging time series via clustering and visualization, NeuroImage 22 (2004) 1214–1222.
[57] D.T. Alexandrescu, C.L. Kauffman, T.A. Jatkoe, D.P. Hartmann, T. Vener, H. Wang, et al., Melanoma-specific marker expression in skin biopsy tissues as a tool to facilitate melanoma diagnoses, Journal of Investigative Dermatology 130 (7) (2010) 1887–1992.