Ann Math Artif Intell (2015) 75:217–229
DOI 10.1007/s10472-014-9424-8

Skin lesion feature vectors classification in models of a Riemannian manifold

Nikolay Metodiev Sirakov · Ye-Lin Ou · Mutlu Mete

Published online: 15 June 2014

Abstract This study is a continuation of a work published by Mete, Ou, and Sirakov (2012), where a model of a 4D manifold of feature vectors was developed. The present paper introduces an improved metric in the 4D manifold first and then extends both the size of the sample space and the dimension (to 6D) of the manifold model in which the sample space lies. As a result, we not only overcame the issue of one single vector representing multiple skin lesions, which occurred in the work of Mete, Ou, and Sirakov (2012), but also improved the accuracy of classification. Furthermore, a statistical evaluation of our support vector machine (SVM) classification method was performed. The intervals of confidence were calculated for the mean of classification of a large sample set in the 6D model. Comparison results of classification with our SVM in 4D and 6D models using 10-fold cross-validation are given at the end of the paper. It is found that the 6D model improves the classification results of the previous study suggesting that two newly introduced features contributed to the increase of the classification accuracy.

Keywords Feature vectors · Riemannian manifold · Support vector machines · Classification · Intervals of confidence

Mathematics Subject Classifications (2010) 62H35 · 68U10 · 68P10

N. M. Sirakov (✉)
Department of Mathematics, Department of Computer Science, Texas A & M University-Commerce, Commerce, Texas, USA
e-mail: Nikolay.Sirakov@tamuc.edu

Y.-L. Ou
Department of Mathematics, Texas A & M University-Commerce, Commerce, Texas, USA
e-mail: Yelin.Ou@tamuc.edu

M. Mete
Department of Computer Science, Texas A & M University-Commerce,Commerce, Texas, USA
e-mail: Mutlu.Mete@tamuc.edu

1 Introduction and motivation

This paper applies Support Vector Machines (SVM) to classify feature vectors (FV) in a 6D model space with a metric. The 6D Riemannian manifold is an extension of the 4D manifold presented in [1]. The 4D FV (A_B, A_C, B, C) is generated by embeding most of the ABCD features used by dermatologists for the purpose of skin lesion diagnosis to benign and malignant [2–4]. With A_B we have denoted in [1] the lesions asymmetry, A_C the asymmetry of the color regions present in the lesion, B the number of the skin lesion abrupt ending arcs, C the number of the color regions present in the lesion.

In [1], the weights from the *Total Dermoscopy Score* were used to develop a metric for the 4D manifold. Furthermore, the 4D vectors were mapped into a 3D space via generalized cylindrical and spherical coordinates in order to determine pattern of the malignant vectors. It was observed that these vectors are located around a four-degree polynomial curve in the 3D spherical spaces, which suggested the use of such a kernel by $\nu - SVM$ in this space. A set of 70 vectors was used to train different types of SVM and the same vectors were classified with an f-measure of 89 % [1].

Despite its popularity among dermatologists ABCD rule poses some disadvantages such as the one presented hereafter. As we extended the size of the sample space from 70 to 90 images we discovered that 18 % of the FVs have the same coordinates in the 4D manifold [1]. A reason for this repeated labeling of lesions is that each vector component takes only a few choices. For example, the A_B component takes the values of 0, 1, and 2 only. This very disadvantage makes a classifier fails many cases since learning from many cross-labeled same feature vectors causes more intricate classifier.

To avoid the problem above, in this paper we split A_B into A_M^B and A_m^B, where $A_M^B \in [0, 1]$ represents the asymmetry of the skin lesion about the major axis, and by $A_m^B \in [0, 1]$ the asymmetry of the skin lesion about the minor axis. Applying the same idea to split A_C to A_M^C and A_m^C we generate a new 6D FV $(A_M^B, A_m^B, A_M^C, A_m^C, B, C)$. This technique proved to be successful as it separates, in the 6D model, the vectors which coincided in the 4D manifold.

The asymmetry calculation approach from [1] is also modified to determine the new kind of asymmetries A_M^B, A_m^B, A_M^C, A_m^C. Further, a 6D manifold was developed using a new set of weights suggested by our dermatologist. Another achievement of the present study is that we extended the number of samples in our experiments, from 70 to 90 FVs, which brings more robust generalization into the skin lesion classification. Unlike our previous study in [1], we performed 10-fold cross-validation for the purpose of the 90 FVs classification in the 6D Riemannian manifold using different kernels of SVM.

The rest of the paper is organized as follows: the next section provides review of the most related studies; Section 3 explains the measurement of the features used to construct the 6D FVs and introduces an elaborated formulation of the active contour model employed to extract these features. Section 4 develops the 6D manifold as the model space and introduces the basic tools of Riemannian geometry we will use to study geometric structures and distribution of the lesion vectors. Next, Section 5 describes the experiments and the classification results obtained with SVM classifier. The paper ends with Section 6 giving a discussion on the obtained results, a list of the contributions, comparison with a recent study, and the future work.

2 Related work

The skin lesion cancer is one of the most deadly cancers in the US and in the world. An automatic imaging system that extracts the features of skin lesions, measures them and

classifies the lesion as malignant or benign would greatly improve clinical practice and patient care. Thus this problem attracts the attention of experts from science fields like mathematics, computer science, and image processing.

A number of papers has been devoted to the problem of skin lesion features extraction. One of the most targeted feature is the lesion's boundary. To extract the boundary of a lesion from noisy image the authors of [7] applied level set approach, whereas [8] used a gradient vector flow. The active contour presented in [22] was successfully applied to skin lesion boundaries extraction, but the method is not capable of accurately defining color regions boundaries. A completely different approach based on neural networks and generic algorithms were employed in [10] for the same purpose. Another feature of interest was the pigmented networks determined in [9] employing graph theory.

The ultimate aim for features extraction and quantification is to provide the necessary strong basis for automatic skin lesion classification to benign and malignant. An overview of set of features useful for classification and methods to do so are discussed in [11]. One of the methods most related to our is given in [12], where SVM classifier has been applied on vectors whose components are color related features. The reported f-measure accuracy for this study is 80 % using 199 images.

Another study close to the present one in the sense of applying the ABCD rule was published by Alcon et al [13]. Unlike our method (which uses SVM) they employed logistic model trees. The accuracy of automatic classification to malignant and benign is 86 %. To achieve this accuracy the authors used "context knowledge" about the risk group and the body part where the lesion is located on.

3 Lesion features

In the present section we first list the set of features used to construct the FVs and explain the basic approaches applied to quantify these features. Next we introduce a shrinking active contour elaborated to extract these features from noisy skin lesion images.

3.1 Construction of the FVs

In this section we describe the quantification of the three main lesion features used to generate a FV from every image of our dataset.

(A)symmetry – As a shape descriptor, asymmetry has been considered in [2] as the uppermost important in evaluation of skin lesions. The lesions posing asymmetric border sign for uncontrolled growth and are high likely to be diagnosed as malignant. Extending our previous calculation in [1], we go beyond simple thresholding and opt to use direct asymmetry measures from the lesions. To calculate asymmetry values for each lesion we start with finding major and minor axes of the lesion boundary [17] (Fig. 1). Note that in the clinical practice dermatologist has to estimate A (first component in the ABCD rule) of the lesion investigating the major and minor axes of asymmetry [2] in a quick manner. Similarly, we modeled this procedure using a threshold after calculation of asymmetry values with respect to major and minor axes. However, this process ends up in [1] with only binary outcomes, 0 or 1, meaning that loss of information is inevitable due to discretization. To relax this stage in the decision making model, instead of using a predefined threshold, we use both asymmetry values of the major and minor axes, as direct inputs to FVs.

In order to quantify asymmetry [1, 15, 21] with respect to a given axis, the area of the lesion or color mask, R, is flipped around this axis. The overlapping region is area of

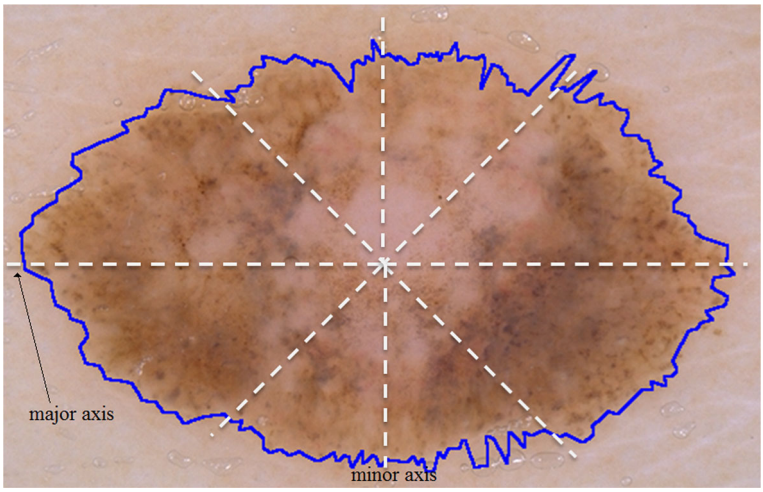

Fig. 1 Lesion boundary (blue, dark in gray-scale) extracted by S-ACES with BC 4 and used to calculate the area S_a

interest, which is denoted by T_s and called true-symmetry. In a similar manner, the non-overlapping region is called false-symmetry and denoted by F_s. Then asymmetry is given by F_s/R for the axis of interest. In contrast with previous studies, in which asymmetry component of FV takes either 0, 1 or 2, we extent A_B to $A_M^B \in [0, 1]$, and $A_m^B \in [0, 1]$.

Similar to asymmetry calculation for lesion mask, asymmetricity of each color region is also calculated replacing A_C - asymmetry of colors in [1] with asymmetry of colors with respect to the lesion's major axis A_M^C and with respect to the lesion's minor axis A_m^C. These components are determined as the average of the asymmetricity of all color regions. For more detailed calculation of the asymmetry components please see [1, 15]. It is obvious from $X^i = (A_M^{Bi}, A_m^{Bi}, A_M^{Ci}, A_m^{Ci}, B^i, C^i)$ that the feature A of the ABCD rule is extended into four features using asymmetry values.

(B)order – To calculate B in [1] we developed a model for border evaluation. This is again to mimic the naked eye abrupt lesion edges assessment performed by dermatologists (ABDC rule [2]). To determine the value taken by B our approach uses the active contour model introduced in Section 3.2. Employing the boundary condition (BC) given in (4) the active contour determines the lesions boundary. The area bounded by this boundary is denoted by S_a. Applying the BC defined in (5) the active contour extracts the lesion's color regions. The union of these regions is denoted by S_b. Further, the lesion is divided into eight sectors by the help of eight straight segments with a join initial point, at the mass center, and $\pi/4$ radians apart from each other (Fig. 1). Thus, in every sector $j = 1, ..., 8$ we define an area S_a^j and an area S_b^j. Then we test if $1 - \tau \leq S_a^j/S_b^j \leq 1 + \tau$ holds [15], where τ is an empirically found threshold. If this test holds for a current section, we add one point to the total score for abrupt border.

In each segment abrupt ending of the lesion [6] is investigated and inputed as fifth component in FVs. [1, 15] detailed calculation of this component. The larger value of B component, the higher likelihood for melanoma.

(C)olors – Since the presence of many colors in a lesion increases the likelihood of melanoma, the last component C of a FV regards number of colors seen in the lesion.

As defined by the ABCD rule, the following are considered major lesion colors: white, red, light-brown, dark-brown, blue-gray, and black. If all 6 colors are found in a lesion the maximum color score is 6. The minimum score is 1 since at least 1 color is expected in skin lesions. Melanomas are usually characterized by presence of three or more colors (Fig. 3).

3.2 Automatic extraction of the FVs components

This section elaborates a shrinking active contour model (S-ACES), whose core version was first [6]. This model is employed to automatically extract the FV components: lesion boundary, lesion colored regions, number of colored regions, and boundary used to calculate abrupt lesion endings.

For the proper skin lesion classification to benign and malignant, the accurate extraction of the above features is an important requirement. To study the abilities of the S-ACES model to comply with the above requirement experiments have been performed in [14] with 50 skin lesion images with ground truth of their boundaries. A statistics on these experiments showed that the mean of the border error is 6.6 % while the mean of the recall and the precision are 85 % and 93 % respectively [14]. In [21] tuning the parameters the border error was reduced to 5.4 %, and consistently, precision moved up to 99% but the recall dropped to 75%. 51 images have been used for the second set of experiments.

The above experiments confirmed that S-ACES is a suitable tool for skin lesion boundary extraction if proper parameters are used in the basic equations of the model. These equations are as follows: evolution equation (EE); initial conditions (IC); and boundary conditions (BC). The EE guides the active contour toward the target (Fig. 2b) and is formulated with the following parametric vector function [1]:

$$r(q,t) = e^{aq-4a^2(t_0+u\partial t)}(C_1 cos(c.a.q), C_2 sin(c.a.q)) \quad (1)$$

In (1), q is a space angular parameter that describes a particular curve and $q \in [0, \frac{2\pi}{ac}]$, $t = t_0 + u\partial t$ defines a family of curves for $u \in [0, \infty)$ (a curve moving through the plane), t_0 denotes the initial time, ∂t is a time period, u is the consecutive time step. Further, $a = |dq|/2$, C_1, C_2, and c are real constants such that $r : R^2 \rightarrow R^2$. The symbol dq denotes the rate of change of q.

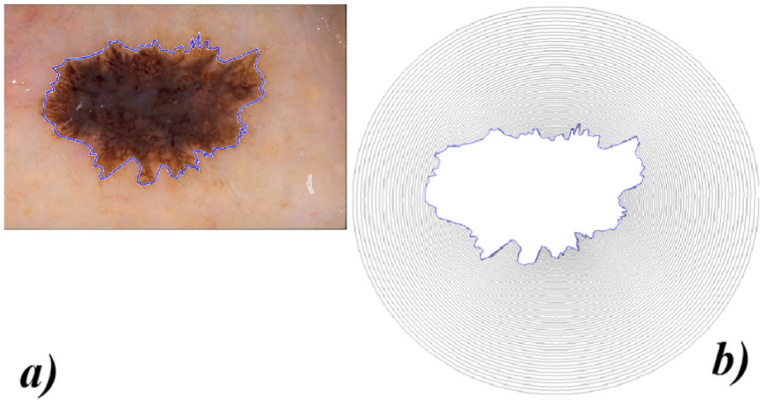

Fig. 2 **a)** Skin lesion along with its boundary determined by S-ACES; **b)** the evolving contour guided by the EE along with the lesion boundary extracted by the contour and its BC. The most outer curve is defined by the IC to envelops the image

The IC are developed to envelop the entire image [6, 15, 21] and to keep the curve closed during the time of evolution as shown in (Fig. 2b)

$$u = 0, t_0 = 0.1, c = 1000, C_1 = C_2 = R = \sqrt{n_c^2 + n_r^2}, \tag{2}$$

$$a = |dq|/2 \geq 0.1, q \in [0, 0.002\frac{\pi}{a}], \tag{3}$$

In (2) n_c denotes the number of columns in the image, n_r the number of rows. One may tell that (1) describes a circle with a radius $R \to 0$ if $a^2 t \to \infty$. In practice R becomes a point (pixel) if $a^2 t \approx 2.5$.

In order to halt the active contour on the lesion's boundary (Fig. 2) the following BC is applied [15]:

$$r(q,t) = r(q, t + \partial t) \;\; if \;\; \frac{\partial f}{\partial t}(r(q,t)) > \epsilon, \;\; for \;\; 2.5 > ta^2 > 0.001 \tag{4}$$

To define color regions and extract their boundaries, the following BC is applied along with evolution (1), and IC (2) and (3):

$$\begin{array}{l} r(q,t) = r(q, t + dt) \;\; if \;\; f(r(q, t + dt)) \in [\epsilon_1, \epsilon_2] \\ r(q,t) \neq (q, t + dt) \quad otherwise. \end{array} \tag{5}$$

The bounds ϵ_1 and ϵ_2 represent image intensities and their values are selected by the user to define the six major colors considered under the ABCD rule [2]: white, red, light-brown, dark-brown, blue-gray, and black. Suggestions about a set of bounds for these colors in RGB space are given by [15]. $f(x, y)$ denotes the image function such that $f : R^2 \to R^+$.

The BC (5) is designed to detect the color's boundaries in images with a high rate of noise (Fig. 3). A BC that let the active contour pass through noise, whose area is a half of a window with sizes $q \times k$ pixels, is developed in [16]. The window is traversing the active contour at every time step u and the BC (5) is checked at every space step Δq (digital analog of dq). A number of experiments were preformed in [16] to validate the capabilities of the new BC.

In [21] the extracted lesion boundaries were used to measure the lesion's asymmetry and calculate the lesion's area. Thus 51 2D FVs were generated in that study. Further BC (4)

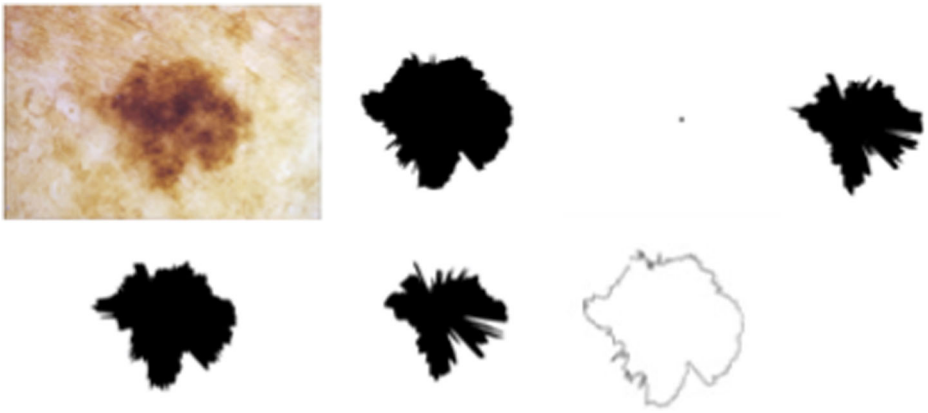

Fig. 3 Upper left to down right: image of malignant lesion; masks from – the lesion's boundary, the black, the blue gray, the light brown, and the red color regions. The down right polygon represents the abrupt boundary of the lesion. All boundaries are extracted by S-ACES

was applied in [15] along with an area based approach to calculate the lesions abrupt endings of 64 skin lesions from images. A BC (5) that allows the active contour passes through noisy regions and extract lesion colored regions is developed in [15] and implemented in [16]. The noise robust implementation was applied to extract the boundaries of the colored lesion regions from total of 90 skin lesion images. Ninety FVs were generated from these images. Thus, our approach automatically extracts and quantifies the asymmetry of the lesion's boundary with respect to (w.r.t.) the major axis of asymmetry A_M^B and w.r.t. the minor axes A_m^B; the asymmetry of lesion's colored regions w.r.t. the major and minor axes, respectively, A_M^C, A_m^C, lesion's abrupt boundary B, and the number of colors in a lesion C. Thus a 6D FV is generated: $X^i = (A_M^{Bi}, A_m^{Bi}, A_M^{Ci}, A_m^{Ci}, B^i, C^i)$ for $i = 1, ..., n$; where n is the number of images in the sample space where we developed our classification framework. For example, the 6D FV for the lesion in Fig. 3 has the following values for its components (0.82, 0.83, 0.79, 078, 1, 3). In this vector the black color was not considered (three colors were counted, Fig. 3), because its area is too small.

In this paper we present the sample space $S_n = \{X^i : i = 1, ..., n\}$ as a subspace of a 6D Riemannian manifold to study the geometric structures and distribution pattern of the sets defined by malignant and benign FVs.

4 Sample space geometry

By extracting the lesion features of an image we generate a FV associated with this image. The set of all FVs of the images studied is the sample space. One of our goals is to embed this sample space into a Riemannian manifold so that the tools of Riemannian geometry can be applied to help identify and classify the malignant images. To the effect, the results of extraction is a process to label images quantitatively. In [1] we used 4-feature labeling and hence the sample space is naturally embedded into a 4D manifold $R^4 = \{(X_1, X_2, X_3, X_4) |\ X_i \in R\}$. The weights in the ABCD rule, used by most dermatologists [2, 3], lead us to choose the metric defined by:

$$g(X, Y) = 1.3X_1Y_1 + 1.3X_2Y_2 + 0.5X_3Y_3 + 0.1X_4Y_4, \tag{6}$$

where $X = (X_1, X_2, X_3, X_4)$ and $Y = (Y_1, Y_2, Y_3, Y_4)$ are any two vectors in R^4. Thus the norm of $X = (X_1, X_2, X_3, X_4)$ is given by

$$|X| = (g(X, X))^{1/2} = [1.3X_1^2 + 1.3X_2^2 + 0.5X_3^2 + 0.1X_4^2]^{1/2}, \tag{7}$$

which yields the *distance* $\rho(X)$ of the terminal point of the vector X to the origin.

We compute the angle between the two vectors $X = (X_1, X_2, X_3, X_4)$ and $Y = (Y_1, Y_2, Y_3, Y_4)$ by

$$\theta = \cos^{-1} \frac{g(X, Y)}{|X||Y|}. \tag{8}$$

For example, the directional angles of a vector $X = (X_1, X_2, X_3, X_4)$ are computed by

$$\theta_i = \cos^{-1} \frac{g(X, e_i)}{|X|} \text{ for } i = 1, ..., 4, \tag{9}$$

where $e_1 = \frac{1}{\sqrt{1.3}}(1, 0, 0, 0,)$, $e_2 = \frac{1}{\sqrt{1.3}}(0, 1, 0, 0,), e_3 = \frac{1}{\sqrt{0.5}}(0, 0, 1, 0,)$, $e_1 = \frac{1}{\sqrt{0.1}}(0, 0, 0, 1,)$ is an orthonormal basis of the Riemannian manifold (R^4, g).

In [1] we also made an attempt to understand the geometric structure of the malignant set in the sample space by projecting 4D space into a 2D plane via a generalized polar coordinates

$$(\rho(X), \theta_m) \quad \longrightarrow \quad (|X| \cos\theta_m, \quad |X| \sin\theta_m), \tag{10}$$

where $\theta_m = \frac{\theta_1+\theta_2+\theta_3+\theta_4}{4}$ is the average of the directional angles.

We also obtained a visualization of our 4D sample space in 3D space by projecting a 4D point to a 3D point via the generalized cylindrical and spherical coordinates [1]:

$$X \quad \longrightarrow \quad (\rho(X)\cos\theta_m, \quad \rho(X)\sin\theta_m, \ \theta_\sigma), \tag{11}$$

$$X \quad \longrightarrow \quad (\rho(X)\sin\theta_m\cos\theta_\sigma, \quad \rho(X)\sin\theta_m\sin\theta_\sigma, \ \rho(X)\cos\theta_m), \tag{12}$$

where $\theta_\sigma = \frac{1}{4}\sum_{i=1}^{4}(\theta_m(X) - \theta_i(X))^2$ is the variation of the directional angles.

In this paper, we improve the 4D metric by changing the weights of the metric given in (6). After consulting our medical colleagues about the importance of the features in diagnosing malignant lesions we choose to use the new metric

$$\bar{g}(X, Y) = 0.3^2 X_1 Y_1 + 0.5^2 X_2 Y_2 + 0.4^2 X_3 Y_3 + 1.5^2 X_4 Y_4 \tag{13}$$

for the manifold R^4.

Replacing the previous metric g by the new metric $\bar{g}$ in formulas (7) and (9) we obtain the distance function and the directional angles of a vector X under the new metric $\bar{g}$. A further computation gives the generalized cylindrical and spherical coordinates using the new metric $\bar{g}$. The comparison of the 10-fold classification results of the SVM classifier applied to the previous cylindrical and spherical coordinates and the new cylindrical and spherical coordinates show that the new metric is a rewarding improvement increasing classification accuracy. More precisely, in the sample space of 90 images, the accuracy of SVM classifier applied to cylindrical coordinates using the new metric is 82.2 % versus 68.8 % of using the previous coordinates; the accuracy of SVM applied to spherical coordinates using the new metric is 83.3 % versus 76.7 % of using the previous coordinates.

In our further attempt to use geometric tools and to enlarge the sample space we discover that there was a drawback in our 4D FVs construction [1]. More precisely, in the 4D sample space constructed from the 90 FVs (A_B, A_C, B, C) there were 18 % of points in which one had exactly same coordinates with another. In other words, there were 18 % of lesions in the data set having repeated labels.

In the present paper, we solved this problem by changing the 4-feature extraction approach into a 6-feature extraction one. As a result, our sample space consists of FVs with 6 components $(A_x^B, A_y^B, A_x^C, A_y^C, B, C)$. Thus, the sample space is now naturally embedded in a 6D manifold $R^6 = \{(X_1, \cdots, X_6) | \ X_i \in R\}$. Based on the new ranking of the features importance for diagnosis in 4D, suggested by our dermatologist, we propose the following Riemannian metric to be used by the SVM for FVs classification in the 6D space:

$$g(X, Y) = 0.3^2 X_1 Y_1 + 0.3^2 X_2 Y_2 + 0.5^2 X_3 Y_3 + 0.5^2 X_4 Y_4 + 0.4^2 X_5 Y_5 + 1.5^2 X_6 Y_6, \tag{14}$$

where $X = (X_1, \cdots, X_6)$ and $Y = (Y_1, \cdots, Y_6)$ are two vectors in R^6. The distance function (determined by the metric given by (14)) of 90 FVs in the 6D sample space is plotted in Fig. 4.

The advantages of replacing the 4D sample space by a 6D sample space include: (1) there are no coinciding points in the new 6D sample space, i.e., the 6-feature extraction gives no repeated labeling of the lesions; (2) the best found 10-fold accuracy with SVM classification, using (14), of the 6D sample space shows an improvement from 80 % to 89 % compared to the result applying the SVM to the 4D sample space.

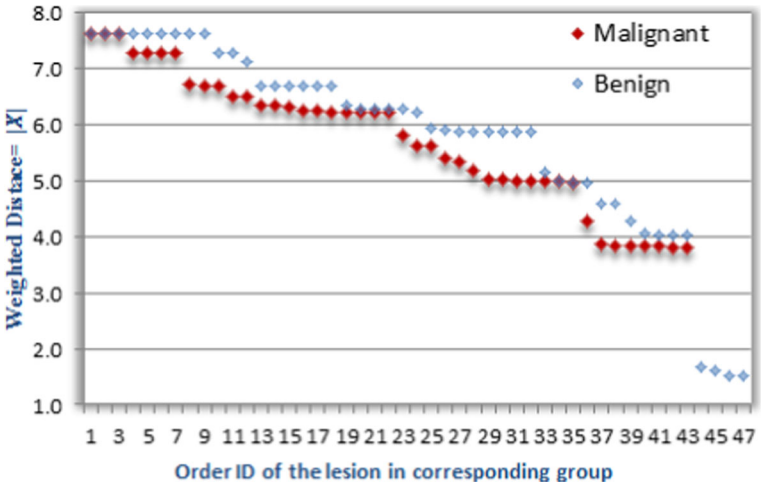

Fig. 4 The graph of the distance function for the 90 6D FVs

5 Experiments and discussion

We used 90 FVs to verify the advantages of utilizing 6D manifold for image lesions classification to benign and malignant. A data set of 90 (47 benign and 43 malignant) lesion images were selected from [5]. To ensure fair experiments and statistical calculation the images have the same resolution with respect to which the image size varies between 1891x1261 and 707x484. The images are produced under the same conditions and three dermatologist anonymously diagnosed each lesion as benign or malignant and manually draw the boundary of each lesion. Next, they agreed on the best boundary and diagnosis. The above defined ground truth was used to determine the accuracy of boundary extraction and classification.

To separate malignant FVs from benign ones we used SVM [23] in the 6D space. Recall that in [1] we did compare SVM classification results in 4D with TDS results.

Throughout the classification experiments, performed in the 6D space, we have applied the LibSVM [24], where in the SVM was used the Riemanian metric represented by (14). 10-fold cross-validation (CV) and model accuracies are reported in Table 1. In 10-fold CV, at each fold non-overlapping quasi-equal class distribution one-tenth of the samples was

Table 1 10-fold cross validation (CV) and model accuracies in 6D of the utilized 90 FVs. 4D data-set is included for a comparison

E	A	Re	Pre	T	K	D	γ	C	Co	Norm
4*D* Model	0.83	0.79	0.85	C-SVM	Poly	4	0.11	1	15	Yes
4*D* CV	0.80	0.72	0.84	C-SVM	Poly	5	0.001	1	10	Yes
6*D* Model	0.93	0.93	0.93	C-SVM	Poly	5	0.11	5	12	Yes
6*D* CV	0.89	0.84	0.92	C-SVM	Radial	N/A	0.01	6.3	1	Yes

The parameters are the ones used to obtain the accuracy in the corresponding experiment.

Norm shows whether data-set is normalized or not.

E: experiment type, A: accuracy, Re: recall, Pre: precision, T: type of SVM, K: kernel, D: degree, C: cost, Co: coefficient

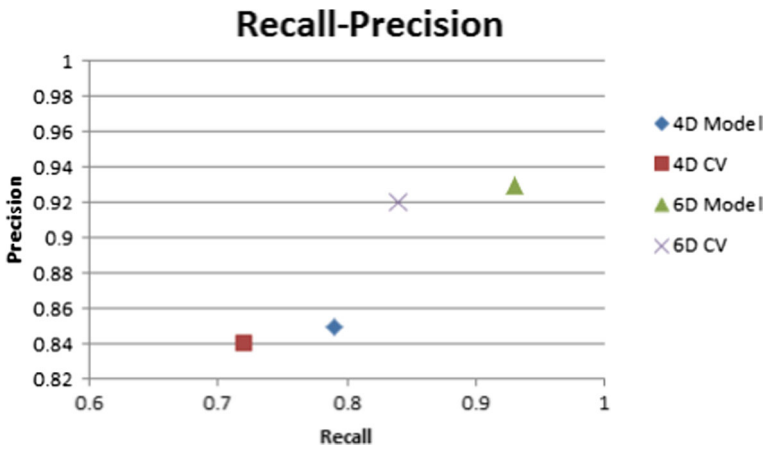

Fig. 5 Recall and precision are plotted, on the interval [0,1] for each of the four kind of classification

excluded from the dataset, and called test set. A model was built with the remaining ones (nine-tenth), which are called training samples. Then, each lesion in test set was classified with this SVM model. This process was repeated ten times having each of the lesions in test set in turn.

A second set of experiment was also conducted to find model accuracy, which is another measure of a classifier accuracy. In model accuracy, none of lesions was removed from training set. All of them are used as training sample and the accuracy is reported.

While experimenting, to reach best classification parameters, we compare candidate models over f-measure, which is harmonic mean of precision and recall. Note that recall is the percentage of positive (malignant) labelled instances that were predicted as positive and found by $TP/(TP + FN)$. Precision is defined as the percentage of positive predictions that are correct, and calculated by $TP/(TP + FP)$.[1]

We employed both, $C - SVM$, the most common type of SVM used in the literature, and $\nu - SVM$ [20] to build up a hyper-surface that will separate malignant from benign FVs. Similar to $C - SVM$, $\nu - SVM$ is also a soft margin classifier using ν-parameterizations instead of C. A wide range of parameters are tested in order to select the optimal set. A grid search was performed for the 6D dataset within the two SVM types. The parameters were varied in the following set1: $\gamma = \{0.001, 0.005, 0.01, 0.015, 0.02, 0.11\}$; $d = \{1, 2, 3, 4, 5, 6\}$; $\nu = \{0.29, 0.4, 0.5\}$ and applying the following kernels - linear, polynomial, radial basis function, and sigmoid. We determined that the best CV accuracies are found using $C - SVM$. The parameter used to calculate the accuracy, recall and precision, reported in Table 1, are listed in the right part of the same table. In addition, the average for recall and precision are plotted in Fig. 5 for every kind of classification performed on the data set of 90 lesion images.

In order to estimate the accuracy of classification of our method for a large data set we calculated the confidence interval for the CV experiments, in 6D.

For this purpose we calculated the percentile of success for each fold of classification in 6D using 90 LFVs: 0.67, 0.78, 0.89, 1, 0.89, 1, 0.78, 1, 0.89, 1. The mean and the standard deviation of the data are 0.889 and 0.1.

[1] TP: True Positive, TN: True Negative, FP: False Positive, FN: False Negative

Since the above classification data is small in size, the experiments do not present binomial trial, the data set is not normally distributed and the standard deviation for the population is unknown we apply the t-distribution formula, as proposed in [25], to calculate the confidence intervals in 6D:

$$\overline{X} \pm t_{\frac{\alpha}{2},n-1} \frac{S}{\sqrt{n}} \tag{15}$$

In (15) $\overline{X}$ denotes the mean of the sample set, n is the size of the sample set, and S is the standard deviation of the sample data. $t_{\frac{\alpha}{2},n-1}$ is a value determined from t-table corresponding to $n-1$ degree of freedom.

Applying (15) on the classification data in 6D it was found that with 90 % of confidence we may state that the mean of the accuracy of classification of a large sample set will belong to the interval of [83.1 %, 94.7 %].

6 Conclusion

As stated in the beginning, this paper is a natural extension of the study presented in [1]. A drawback in the previous work was that 18 % of the FVs had the same coordinates in 4D sample space due to the very narrow range of values that each component of a FV can take. To overcome this repeated labeling problem, obtain a better understanding of the geometric structure of the sample space, and achieve more accurate classification/identification of FVs, the present work has led to the following contributions:

- The active contour model used for skin lesion features extraction from images was theoretically elaborated;
- The metric of the 4D manifold was improved under the guidelines of our medical expert;
- The dimension of the ambient manifold in which the sample space of FVs lies, has been extended from 4D to 6D. With the new model and sample space of 6D FVs we solved the repeated labeling problem and achieved a better accuracy of classification (Table 1);
  -The intervals of confidence were calculated for the mean of the accuracy of the CV classification, with our SVM, in a large sample set in the 6D space.

In [26] the authors developed an approach for skin lesion classification. In their work a geometric information extracted from skin lesion images along with information from knowledge database were used as input to the classification process. In the present study geometric and color information are used as input to the SVM for skin lesion classification to benign and malignant.

[26] calculates the minimal number of samples needed to receive $E = 1$ % error in the evaluation of the classification method presented there. The following formula is applied for this purpose:

$$Z_{\frac{\alpha}{2}} \leq \frac{E\sqrt{n}}{\sqrt{pq}}. \tag{16}$$

In (16) n denotes the number of samples needed, p - is the percentile of successes, q - is the percentile of un-successes, while $Z_{\frac{\alpha}{2}}$ is a coefficient from a table for a standard normal distribution..

Considering $E = 1$ % and the model classification in 6D, where $p = 0.933$ (Table 1), and $q = 0.067$ we determined that $Z_{\frac{\alpha}{2}} = Z_{0.005} = 2.58$. Applying (16) we found that $4,333 \leq n$, which means that at least 4,333 skin lesion images are needed to perform

experiments in order to evaluate the classification with 1 % error. Thus the lower bound of our method is smaller than 6656 needed by [26] to evaluate their method with the same error.

One of the classification works more close to our is [13] where the ABCD rule is applied along with logistic model tree which uses meta data (age, gender, the part of the body the lesion is located) to facilitate the classification. The authors performed 10 fold CV and obtained an f-accuracy of 86 %. In our method no meta data is used and the 10 fold CV showed an f-accuracy of classification 89 % in the 6D manifold.

Another study to compare with is reported in [12]. This paper applies SVM classifier for the purpose of classification to benign and malignant and reports an f-accuracy of 80 % on a sample space of 199 skin lesion images. As noted above the accuracy of our SVM classification is 89 % but our data-set has 90 images. Recall that we calculated in the previous section we are 90 % confident that the accuracy of our approach on a large population (including 199 images) is in the interval of [83.1 %, 94.7 %].

This work will continue with a further extension of the FV dimensionality through including new features as vector components. Features subject of interest are skin lesion dots and streaks.

Acknowledgements This study is partly supported by Texas A&M University-Commerce Research Grant. Thanks to Dr. Rick Selvaggi for helping us determining the weights of the new metrics. Thanks to the graduate student M. Alabdulhadi for helping with the calculation of the interval of confidence and estimation the minimal number of samples. Thanks to Dr. T. Boucher and Dr. H. Coskun for the useful discussions on the intervals of confidence. We are thankful also to the anonymous reviewers whose comments and suggestions helped in improving the quality of the paper.

References

1. Mete, M., Ou, Y.-L., Sirakov, N.M.: Skin Lesion Feature Vector Space With A Metric To Model Geometric Structures Of Malignancy, R.P. Barneva, V. Brimkov, J. Aggarval(Eds.): IWCIA 2012, LNCS 7655, Springer Verlag, pp. 285–297 (2012)
2. Nachbar, F., Stolz, W., Merkle, T., et al.: The ABCD rule of dermatoscopy: High prospective value in the diagnosis of doubtful melanocytic skin lesions. J.. Am. Acad. Dermatol. **30**(4), 551–559 (1994)
3. Argenziano, G., Fabbrocini, G., Carli, P., De Giorgi, V., Sammarco, E., Delfino, M.: Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions. Comparison of ABCD rule of dermatoscopy and a new 7-point checklist based on pattern analysis. Arch. Dermatol **134**, 1563–1570 (1998)
4. American Cancer Society: Cancer Facts & Figures 2010, http://www.cancer.org/. Accessed July, 26 (2010)
5. Argenziano, G., Soyer, H.P., De Giorgi, V., et al.: Interactive atlas of dermoscopy. Milan. EDRA Medical Pub, Italy (2000)
6. Sirakov, N.M., Ushkala, K.: An integral active contour model for convex hull and bound- ary extraction. In: Bebis, G. et al. (Eds), LNCS, Vol. 5876, Springer, pp. 1031–1040 (2009)
7. Silveira, M., Marques, J.S.: :Level set segmentation of dermoscopy images, 5th IEEE ISBI: From Nano to Macro, Paris, 14-17 May 2008, pp. 173–176, (2008). doi:10.1109/ISBI.2008.4540960
8. Zhou, H., Schaefer, Celebi, M.E., Lin, F.: : Gradient vector flow with mean shift for skin lesion segmentation. Comp. Med. Imaging Graphics **35**(I.2), 121–127 (2011)
9. Sadeghi, M., Razmara, M., Atkins, M.S., Lee, T.K.: :A novel method for detection of pigment network in dermoscopic images using graphs. Comput. Med. Imaging Graphics **35**(2), 137–143 (2011)
10. Xie, F., Bovik, A.: :Automatic segmentation of dermoscopy images using self-generating neural networks seeded by genetic algorithm. Journal of Pattern Recognition **46**(3), 1012–1019 (2013)
11. Maglogiannis, I.: Doukas, G.N.:Overview of Advanced Computer Vision Systems for Skin Lesions Characterization, IEEE Tran. on Information Technology in Biomedicine, V. 13, NO. 5. SEPTEMBER (2009)

12. Gilmore, S., Hofmann-Wellenhof, R., Soyer, P.H.: A support vector machine for decision support in melanoma recognition. Exp Dermatol **830-5**(2010), 19 (2010)
13. Alcon, J.F., Ciuhu, C., Kate, W., Heinrich, A., Uzunbajakava, N., Krekels, G., Siem, D., Haan, G.: :Automatic Imaging System With Decision Support for Inspection of Pigmented Skin Lesions and Melanoma Diagnosis. IEEE J. Sel. Top Signal Proc. **3**(1) (2009)
14. Mete, M., Sirakov, N.M.: Application of active contour and density based models for lesion detection in dermoscopy images. In: BMC Bioinformatics 2010, 11(Suppl 6):S23 (2010). doi:10.1186/1471-2105-11-S6-S23
15. Mete, M., Sirakov, N.M.: Dermoscopic diagnosis of melanoma in a 4D feature space constructed by active contour extracted features. Journal of Medical Imaging and Graphics **2012**(36), 572–579 (2012)
16. Nara, C.: : Active contour on the exact solution of the active convex hull model working with noise, Master Degree Thesis, Texas A and M Univ. Commer **07**, 01 (2011)
17. Mulchrone, K., Choudhury, K.: Fitting an ellipse to an arbitrary shape: implications for strain analysis. J. Struct Geol. **26**(1), 143–153 (2004)
18. O'Neill, B.: Semi-Riemannian geometry with applications to relativity. Academic Press, New York (1983)
19. Petersen, P.: Riemannian geometry, Vol. 171. Springer-Verlag, New York (1998)
20. Scholkopf, B., Smola, A.J., Williamson, R.C., Bartlett, P.L.: New Support Vector Algorithms. Neural Comput. **12**, 1207–1245 (2000)
21. Sirakov, N.M., Mete, M., Nara, S.C.: Automatic boundary detection and symmetry calculation in dermoscopy images of skin lesions, IEEE ICIP2011, Brussels, pp. 1637–1640 (2011)
22. Thieu, Q.T., Luong, M., Rocchisani, J.M., Vienne, E.: A convex active contour region-based model for image segmentation. Proc. CAIP'11, Part I, LNCS, Vol. 6854, Springer-Verlag Berlin, Heidelberg, pp. 135–143 (2011)
23. Vapnik, V.: The Nature of Statistical Learning Theory. Springer, New York (1995)
24. Chang, C.C., Lin, C.-J.: LIBSVM: a library for support vector machines, Vol. 2 (2011). http://www.csie.ntu.edu.tw/~cjlin/libsvm
25. Devore, J., Berk, K.: Modern mathematical statistics with applications. International Student Edition ed. Belmont, CA: Thomson Higher Education (2007)
26. Blackledge, J.M., Dubovitskiy, D.A.: Object Detection and Classification with Applications to Skin Cancer Screening. ISAST Tran. Intiligent Syst. **1**(2) (2008)