Original Article

www.jmss.mui.ac.ir

A Novel Unsupervised Approach for Minimally-invasive Video Segmentation

Toktam Khatibi[1], Mohammad Mehdi Sepehri[1,2], Pejman Shadpour[2]

[1]Department of Industrial Engineering, Tarbiat Modares University, [2]Department of Hospital Systems and Engineering Hospital Management Research Center, Iran University of Medical Sciences, Tehran, Iran

Submission: 17-04-2013 Accepted: 24-11-2013

ABSTRACT

Temporal segmentation of laparoscopic video is the first step toward identifying anomalies and interrupts, recognizing actions, annotating video and assessing the surgeons' learning curve. In this paper, a novel approach for temporal segmentation of minimally-invasive videos (MIVS) is proposed. Illumination variation, shadowing, dynamic backgrounds and tissue respiratory motion make it challenging to extract information from laparoscopic videos. These challenges if not properly addressed could increase the errors of data extraction modules. Therefore, in MIVS, several data sets are extracted from laparoscopic videos using different methods to alleviate error effects of data extraction modules on MIVS performance. Each extracted data set is segmented temporally with Genetic Algorithm (GA) after outlier removal. Three different cost functions are examined as objective function of GA. The correlation coefficient is calculated between objective values of the solutions visited by GA and their corresponding performance measures. Performance measures include detection rate, recognition rate and accuracy. Cost functions having negative correlations with all mentioned performance measures are selected. Finally, a multi-objective GA is executed on the data sets to optimize the selected cost functions. MIVS is tested on laparoscopic videos of varicocele and ureteropelvic junction obstruction surgeries collected from hasheminejad kidney center. Experimental results demonstrate that MIVS outperforms the state-of-the-art methods in terms of accuracy, detection rate and recognition rate.

Key words: *Minimally invasive surgery, multi-objective genetic algorithm, surgical instruments, video segmentation*

INTRODUCTION

Laparoscopy or minimally invasive surgery is a surgical procedure which uses a few small incisions.[1] Therefore, it has many advantages over open surgery like shorter recovery time.[2] The surgeon insert laparoscope into the abdominal cavity to see through the laparoscope.[1] The video output of the laparoscope can be recorded. Minimally-invasive video (MIV) can be divided into temporal segments. In previous, several researchers study temporal segmentation of endoscopic/laparoscopic videos.[3-6] These methods aim to segment medical video based on tissue boundaries[4,5] or type of active instruments identified by signals of sensors, that are installed on surgical tools.[3,6] The last class of segmentation methods considers types of active surgical instruments for surgical workflow detection.[3,6]

In this paper, a novel approach for temporal segmentation of MIVS is proposed. MIVS aims to segment the laparoscopic video based on several data sets, describing motion and number of surgical tools.

Extracting data from laparoscopic videos faces various challenges such as illumination variation, shadowing, dynamic backgrounds and tissue respiratory motion. These challenges if not properly addressed could increase the errors of data extraction modules. Therefore, in MIVS, several data sets are extracted from laparoscopic videos using different methods to alleviate error effects of data extraction modules on MIVS performance. Each extracted data set is segmented temporally with Genetic Algorithm (GA) after outlier removal. Three different cost functions are examined as objective function of GA. Moreover, the solutions visited by GA are compared to the solution segmented by human experts and the performance measures are calculated for the visited solutions.

The correlation coefficient is calculated between objective values of the solutions visited by GA and their corresponding performance measures. Performance measures include detection rate, recognition rate and accuracy. Cost functions having negative correlations with all mentioned performance measures are selected. Finally, a

Address for correspondence:
Prof. Mohammad Mehdi Sepehri, Department of Industrial Engineering, Tarbiat Modares University, Tehran, Iran. E-mail: mehdi.sepehri@modares.ac.ir

multi-objective GA (MOGA) is executed on the data sets to optimize the selected cost functions.

The novelty of this paper is three fold. Firstly, unlike the previous methods, the segmentation obtained by MIVS can discriminate surgical actions so that each segment includes at most one surgical activity. Thus, temporal segmentation of laparoscopic videos by MIVS has different applications, such as detecting laparoscopic anomalies and interrupts. Moreover, it is the prerequisite step for action recognition, video tagging, training purposes and surgeon's learning curve assessment.

Secondly, unlike the previous methods which segment video based on single data set, MIVS aims to segment video based on several data sets. Moreover, type of data extracted in MIVS is different from those extracted in the previous studies. MIVS aims to segment video based on the number and motion of instruments while the previous studies do that by extracting tissue properties. Moreover, unlike some previous studies, MIVS does not require any additional information collected from external sensors.

Thirdly, a novel approach is proposed in this paper to find the most appropriate cost functions, considering the correlations between cost functions and performance measures. The cost functions having negative correlation with all mentioned performance measures are selected. The reasons for that are as follows:

- The mentioned performance measures are positive indicators. Therefore, maximizing them is desired
- Cost functions are negative indicators and minimizing them is desired
- If a cost function is negatively correlated with the mentioned performance measures, decreasing this cost function make the mentioned performance measures increase.

The rest of the paper is organized as follows. Section 2 reviews the related works. In section 3 the methodology of this research is described and then the results and findings are reported in section 4. Finally, section 5 summarizes and concludes the paper.

RELATED WORKS

Motion data of surgical instruments can be analyzed for skill assessment and surgical activity recognition with reasonable accuracy.[7] In this paper, motion and number of surgical instruments will be extracted from laparoscopic video with image processing techniques. Then, the extracted data sets, which are time-series data, will be segmented temporally. Therefore, the previous studies about extracting surgical instrument descriptors and segmentation of temporal data are reviewed.

Previous Studies about Extracting Descriptors of the Surgical Instruments

First of all, the video frames have to be segmented to identify the visible surgical instruments. Afterwards, the motion descriptors of the identified instruments will be estimated. Surgical instrument segmentation and tracking have been attracted in the previous researches and various methods are proposed for surgical instrument detection.[8-10]

These methods are designed specifically for use in virtual reality and simulations. Videos of simulated surgeries face a few challenging situations. For example, respiratory motions of tissues and illumination variations may not be considered in the videos of the simulated surgeries.

In this paper, we have to face more challenges to extract data from videos of real laparoscopic procedures compared to virtual reality and simulations. Therefore, the previous methods proposed for processing videos of the simulated surgeries may not be robust to the mentioned challenges. Therefore, an offline rule-based segmentation algorithm is proposed in this paper for laparoscopic instrument detection.

After segmenting the surgical instruments, their motion patterns must be described. Many previous studies introduce approaches for motion detection and estimation. Some researchers use optical flow,[11-14] motion history image (MHI) and motion energy image (MEI),[15-17] motion history histogram (MHH),[16] directional motion history image and directional motion energy image[17] for detection or estimation of motions in video frames. MHI does not provide explicit motion estimation. MEI can be used as a mask for detection of moving regions in the video frames. MHH is introduced to solve overwriting older motions with newer ones in MHI. However, it doesn't show the sequence of motions in time. Optical flow vectors measure the replacements of corresponding pixels of two consecutive frames. Global and local motion vectors can be estimated from the optical flow. Some studies use histogram of optical flow for this purpose,[18,19] but this requires quantization of optical flow vectors. The major disadvantage of motion quantization is the risk of information loss. Thus, motion quantization may lead to over segmentation of the quantized data. For this reason, the continuous motion vectors are considered in this paper and no quantization is applied to the extracted data sets.

Some previously proposed features include scale-invariant feature transform (SIFT),[20] speeded up robust features (SURF)[21] and Harris corner points.[22-24] Matched pairs of features extracted from two consecutive video frames are identified and the mapping function is estimated from the matching pairs. The coefficients of the mapping function are considered as the motion parameters.[25]

Our experiment shows that very few SURF and SIFT features

are corresponding to the identified surgical instruments in many video frames. It could mislead the motion estimation method because of not enough features. Hence, SURF and SIFT features are not used for motion estimation in this paper.

The number of Harris corner points, detected in each surgical instrument image, is more than its corresponding SURF and SIFT features. Another advantage of Harris corner detector over SIFT detector is its shorter running time.[26] Therefore, the instrument motion estimation based on its corner points is more accurate than based on SURF or SIFT features.

In this paper, the corner points are extracted from consecutive video frames based on Harris corner detector that is a mathematical operator. It is simple to compute and fast[22-24] After extracting corner points, the corner points of two consecutive frames are matched. Affine parameters of the instrument motion are estimated based on matching corner points of the surgical instruments.

For more confidence, the motion magnitude of the surgical instruments is estimated too. For this purpose, the optical flow vectors are computed for every pixel inside the identified instruments. Optical flow is the apparent motion of pixels between frames. The motion magnitude is estimated based on the optical flow vectors. We extract motion parameters and motion magnitude based on two different methods, considering corner points and optical flow vectors. The reason of considering two different data extraction method is to alleviate the effect of data extraction error on the performance of our proposed method.

Previous Methods for Temporal Segmentation

Temporal segmentation of time-series data is performed to find internally homogeneous segments.[27] There is 2^n-1 possible segmentation for a sequence of length n. Even if some constraints like minimum allowable segment length are added to the temporal segmentation problem, the number of all feasible segmentations is yet an exponential number of order n.[28]

Many different algorithms and methods are introduced for temporal segmentation.[27,29-31] Temporal segmentation methods are classified into three classes, including grammar-based methods, segment boundary detection and sliding windows (SW).[32]

Grammar-based models segment data by learning statistical models of motion data. Some popular grammar-based segmentation methods include hidden Markov models, conditional random fields and semi-Markov models.[32] Grammar-based methods are supervised while two others are unsupervised. Using grammar-based methods requires prior knowledge about the segment properties.

If temporal segmentation aims to cut motion stream into single action instances, grammar based methods can be used for temporal segmentation. In a time series of local motion parameters, the segments can be considered as actions. Grammar-based methods are used for action segmentation and recognition simultaneously. The grammatical model of each action must be defined previously and all possible actions must be known in prior. These methods are not robust against new action types and speed and style variations of the same action.[32]

Boundary detection methods segment the motion sequences based on time derivatives of local motion parameters.[33,34] They are not robust against data extraction error.[32]

Moreover, they are highly dependent to the specified threshold for detecting the segment boundaries. A state-of-the-art method of this group is change detection method.[34] Change detection method[34] tries to estimate the data distribution and use it for segmenting time series. The performance of this method depends on its parameter setting.

In SW approach, a segment is grown until its cost exceeds the predefined threshold. This process is repeated for the next segment.[27] SW is a very fast method and it is more generalizable than boundary detection methods.[32]

Boundary detection and SW methods use criteria for finding the segment boundaries, including dynamic time warping (DTW),[30] direct point to point distance (DPPD)[35] and some other measures.[35] DTW and DPPD are used in a supervised manner because they compare the query time series with some predefined patterns. Please refer to[30,35] for more details about the discussed criteria. A state-of-the-art method of this group is dynamic principal component analysis (DPCA)-based segmentation.[36] This method measures the segment homogeneity in an unsupervised manner.

We will perform temporal segmentation with unsupervised methods in this paper. Because unsupervised methods are more generalizable and they don't require any predefined patterns. Moreover, different surgeons perform the same activities in different styles with different speeds. Therefore, temporal segmentation of laparoscopic videos using predefined patterns is not robust against different styles of surgical activities.

Temporal segmentation tries to detect homogeneous segments in time-series data. For this purpose, the homogeneity of the segments must be measured quantitatively. Therefore, a homogeneity measure must be used. A good segmentation method tries to optimize the homogeneity measure. For this reason, temporal segmentation is actually an optimization problem[30] and meta-heuristics can be used for time-series segmentation[29-31] similar to other optimization problems.

Temporal segmentation using meta-heuristics requires defining and using an appropriate objective function (cost function). Objective function is defined for determining the homogeneity of generated segments. Previous studies use several measures of segment homogeneity, including summation of the segment variances[27,37,38] and summation of covariance of segments.[27,36]

Finding the border points of the current segment highly depends on the last border point of the previous segment in SW and boundary detection methods. If the borders of a segment are misidentified, the border points of its adjacent segments may be misrecognized too. As mentioned before, SW and boundary detection methods are sensitive to the predefined thresholds. Utilizing meta-heuristics for choosing the segment borders can alleviate the mentioned error propagation in segments. Meta-heuristics choose the borders of the next segments independently and finally, the best segmentation is a solution with minimum total cost among the candidate solutions.

In this paper, temporal segmentation is performed with meta-heuristic algorithms as an optimization problem. For this purpose, several cost functions are examined to find the best objective function for the Meta-heuristic algorithm.

MATERIALS AND METHODS

The main steps of the research methodology are presented in Figure 1.

As shown in Figure 1, MIVS method consists of four steps, including data extraction from laparoscopic video, outlier detection and removal, finding appropriate cost functions and finally temporal segmentation with MOGA. The more details about steps of MIVS will be described in the following subsections.

Data Extraction

In this paper, five different data sets are extracted from laparoscopic video frames. One of them includes the number of the identified instruments and four other data sets describe the motion of the instruments. From these four data sets, one estimate affine parameters of the aggregated motion of the identified instruments. Another one estimates the average magnitude of motion vectors of the identified instruments. The third one estimates the affine motion parameters of the instrument having the largest motion magnitude. The last one estimates the magnitude of motion vector of this instrument.

The main steps of our data extraction method are illustrated in Figure 2.

As shown in Figure 2, first of all, camera motion compensation is performed.[39] Camera motion is detected and compensated in order to recognize surgical instrument motions with high accuracy.[39]

Then, a novel surgical instrument segmentation algorithm (SISA) is applied to the input video frames to identify the visible instruments. IS in Figure 2 is the union of instruments identified by SISA from input video frames. IS includes one or more individual connected components IS_i.

Therefore, the number of individual connected components is stored in DS_1. Moreover, the affine parameters of the aggregated motion of all the identified instruments are stored in DS_2. Average magnitude of the aggregated motion of instruments is stored in DS_3. On the other hand, the motion magnitude for each surgical instrument (IS_i) is estimated and stored in temporary variable M_i. Affine motion parameters of instrument IS_j having the largest motion magnitude (M_j) is stored in A_j. A_j and M_j are stored in DS_4 and DS_5, respectively.

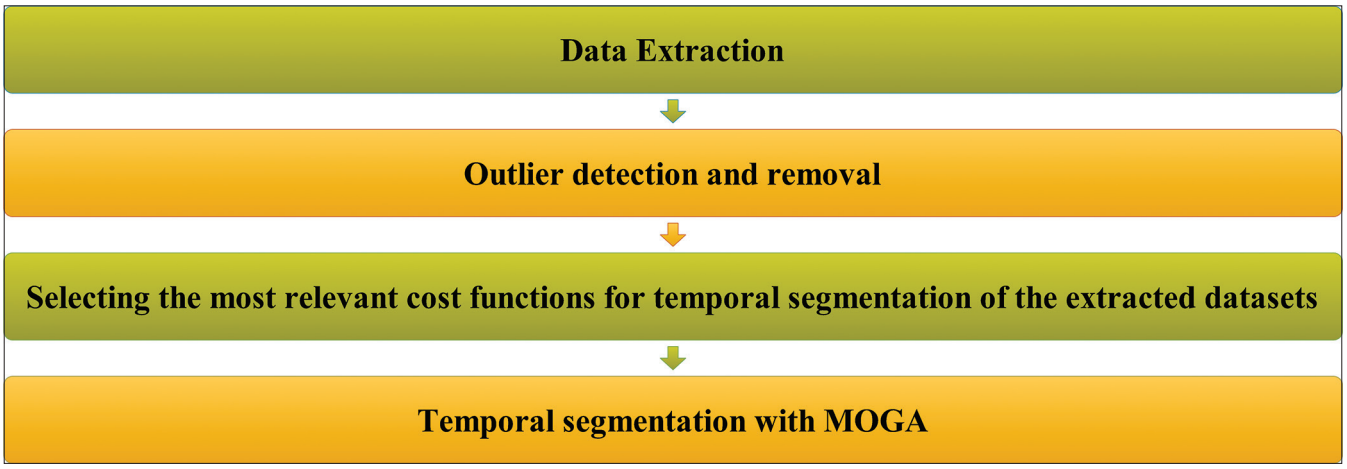

Figure 1: The main steps of minimally-invasive videos (MIVS) method for MIV segmentation

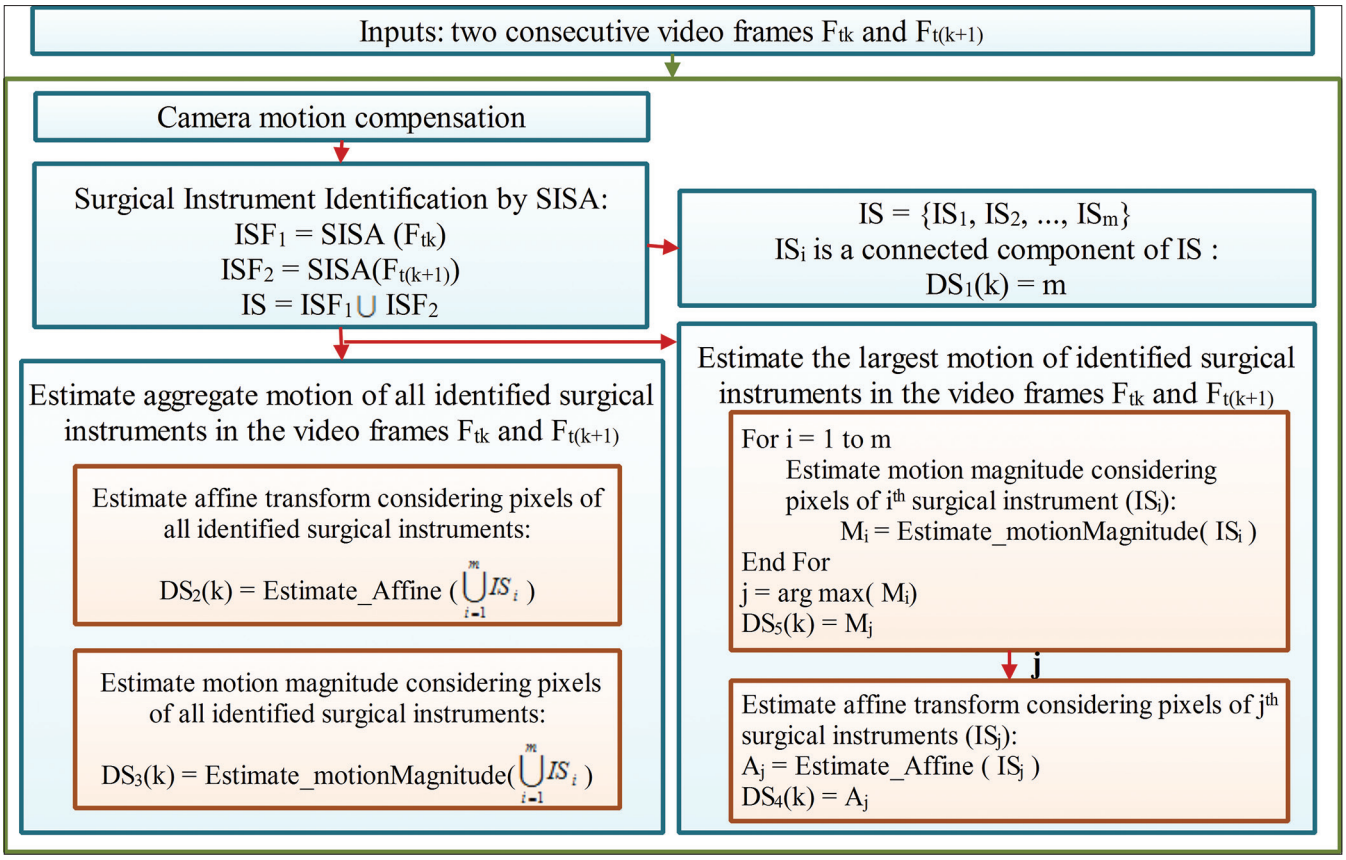

Figure 2: The main steps of data extraction method from laparoscopic videos (DS_i: i^{th} extracted data set)

For performing the mentioned steps illustrated by Figure 2, three modules, including SISA, Estimate_Affine and Estimate_Motion_Magnitude will be executed. We will introduce these three modules in the following subsections.

SISA Module

As discussed in Section 2.A, existing methods for surgical instrument detection are more appropriate for virtual reality and simulated applications. They may have poor performance in segmenting laparoscopic videos taken from the surgical activities. Therefore, SISA as a novel SISA is proposed in this paper. The main steps of SISA are illustrated in Figure 3.

As illustrated by Figure 3, SISA executes an offline rule extraction algorithm for describing surgical instrument color (REDSIC).

In REDSIC, Positive (negative) class indicates surgical instrument (soft tissue) pixels.

REDSIC extracts the required rules for classifying pixels of each video frame. For this purpose, some representative regions of tissues and instruments are selected manually. Color average of each region is extracted. The extracted data set is partitioned to train and test set. Train set is used for training the classifier and test set is used for evaluating the classifier performance.[40] The classification rules can be extracted from decision trees more conveniently than other classifiers like k-nearest neighbor classifier. Therefore, REDSIC classifies data using decision tree.

After extracting discriminating rules by REDSIC, video frames are segmented by Steps 1-3 of SISA. In SISA, 3 × 3 average filter is applied to the video frame in RGB color space and creates a filtered image. Pixels of the filtered image are classified based on the output rules of REDSIC. Pixels of positive class are considered as candidate pixels of instruments. The candidate pixels compose different connected components (regions). The large enough regions will be considered as surgical instruments.

Evaluating SISA is performed in Appendix A via comparing to other state-of-the-art methods.

Estimate_Affine Module

Estimate_Affine module is used for estimating motion of surgical instruments as affine parameters. The reason for that is affine transformation can recognize many surgical actions like opening or closing the tool tip and changing the shape of the instrument. Furthermore affine transformation can recognize any combination of translational, rotational and rescaling transformations.

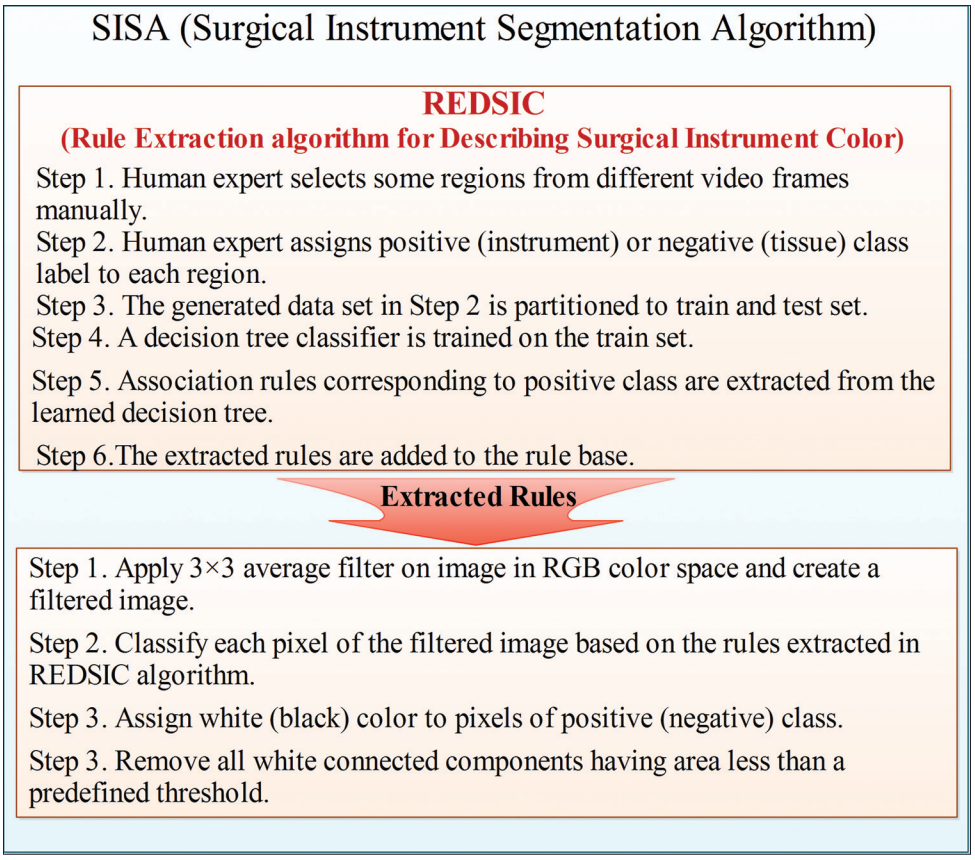

Figure 3: Steps of surgical instrument segmentation algorithm

Steps of Estimate_Affine module are as follows:

Step 1: Corner points of the target object are extracted from video frames F_{tk} and $F_{t(k+1)}$ via Harris corner detector as described in.[22]

Step 2: Matching pairs of corner points extracted from two consecutive frames are identified as described in.[23,24]

Step 3: Affine transform parameters are estimated by Random Sample Consensus algorithm as described in.[25] This transform tries to map corner points of F_{tk} to their corresponding corner points of $F_{t(k+1)}$.

Estimate_MotionMagnitude Module

Estimate_MotionMagnitude module is used for estimating the average motion magnitude of one or more objects. The steps of this module are as follows:

Step 1: Optical flow vectors are calculated for the boundary pixels of the target object as described in.[11] For this purpose, the corresponding boundary pixels of the target object in video frames F_{tk} and $F_{t(k+1)}$ are considered

Step 2: The average of optical flow magnitude is calculated for boundary pixels of the target object.

One instrument may have more than one dominant motion direction. Therefore, all optical flow vectors of its boundary pixels are considered for estimating its motion magnitude regardless of their motion direction.

Figure 4 indicates how our proposed method of data extraction is applied to some real laparoscopic video frames.

As illustrated by Figure 4, two consecutive video frames (F_i and F_j) are segmented using SISA method. Since two disjoint connected components are identified by SISA, as illustrated by Figure 4c and d, the value of the corresponding member of DS_1 will be 2. Then the optical flow vectors and corner points are extracted from F_i and F_j. The matching pairs of corner points of each surgical instrument are identified. The motion pattern of each surgical instrument is determined based on its matched corner points. Moreover, the motion magnitude of each surgical instrument is determined based on the optical flows of its boundary points. Since I_1 has the larger motion magnitude than I_2, the affine motion parameters and motion magnitude of I_1 are stored in DS_4 and DS_5 respectively. As we told before, the affine parameters and magnitude of the aggregated motion of surgical instruments are estimated and stored in DS_2 and DS_3, respectively.

Now, we are ready to explain the reason of extracting five different data sets from laparoscopic videos. Let us see the

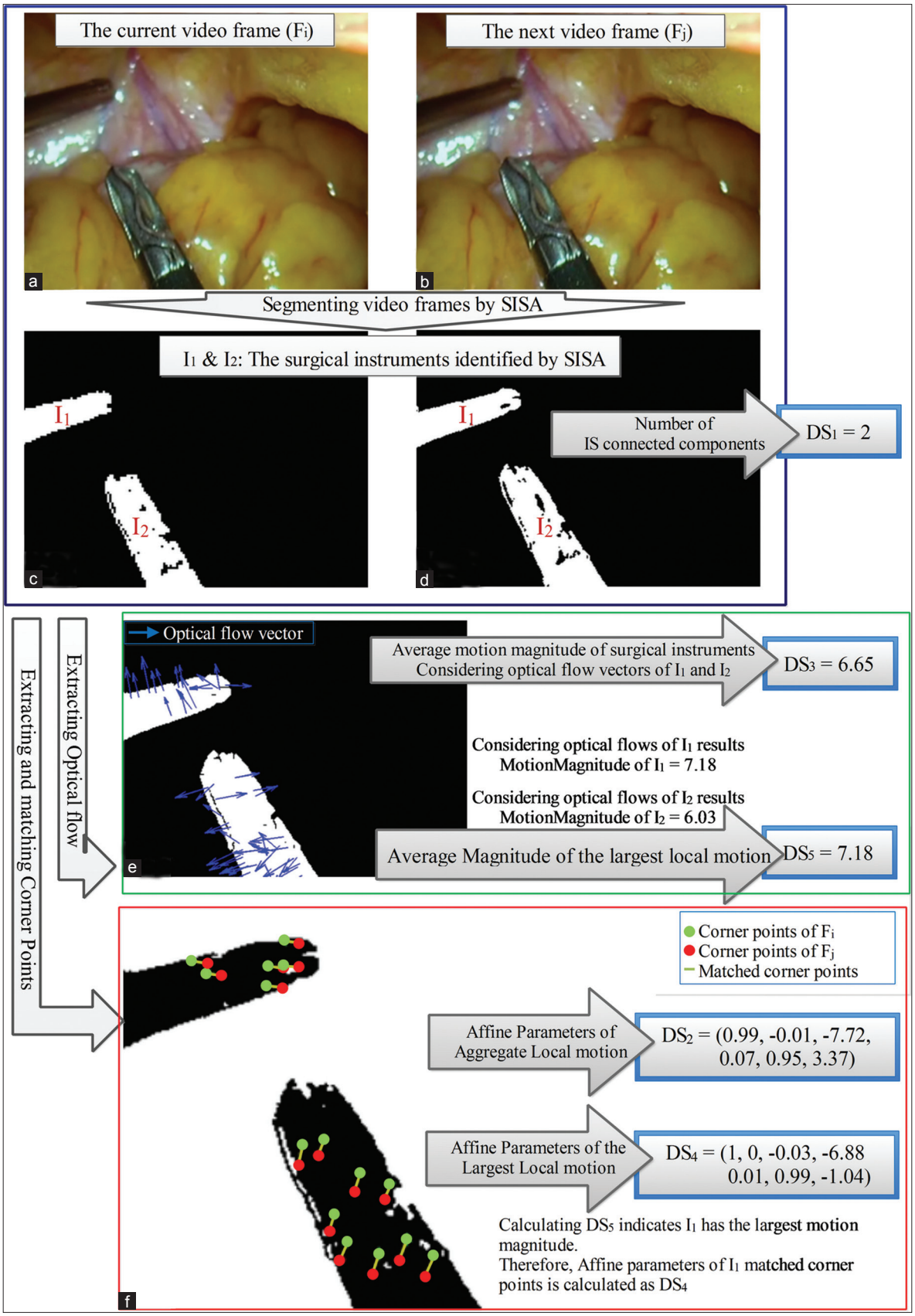

Figure 4: Applying the proposed method of data extraction to some real laparoscopic video frames: (a) A sample video frame (F_i). (b) The next video frame (F_j). (c) The segmented image of F_i. (d) The segmented image of F_j. (e) Extracting optical flows from F_i and F_j and calculating data sets DS_3 and DS_5. (f) Extracting and matching corner points from F_i and F_j and calculating data sets DS_2 and DS_4

results of segmenting some sample video frames with SISA in Figure 5.

As depicted in Figure 5, SISA segments video frames and labels some regions as surgical instruments. SISA algorithm is applied to the video frames illustrated by Figure 5 (a-1), (a-2), (a-3), (a-4) and (a-5) and outputs their corresponding segmented image depicted in Figure 5 (b-1), (b-2), (b-3), (b-4) and (b-5) respectively.

Figure 5 (a-1), (a-2) and (a-3) indicate two surgical instruments. Figure 5 (b-1) can detect different surgical instruments as two disjoint regions. However, the segmented regions of two instruments in Figure 5 (b-2) are not disjoint and constitute one connected component. Therefore, the number of white connected components is not a good representative of the number of instruments in this frame. Moreover, the largest motion may be estimated incorrectly.

As illustrated by Figure 5 (b-3), dark shadow of tissue is misclassifies as instrument. This may underestimate the aggregated motion and overestimate the instrument numbers. However, the largest motion can be detected correctly because the dark shadow of tissue indicates no significant motion after compensating camera motion.

Figure 5 (a-4) indicates only one instrument. Figure 5 (b-4) illustrates that the segmented image of this instrument includes more than one connected component. This may cause error in estimating the number of instruments and the largest motion parameters.

Hence, it is concluded that using only one data set may be misleading because of its estimation errors in real laparoscopic videos. Therefore, segmenting video based on multiple data sets can decrease the error rate of video boundary detection.

Outlier Detection and Removal

Outliers can be misrecognized as segment boundaries. It may increase the segmentation error. Thus, outlier detection and removal are required before segmenting the extracted data sets temporally. Many approaches have been presented for outlier detection in previous studies.[41] These methods can be divided into statistical tests, depth-based approaches, deviation-based methods, distance based and density based methods.[42]

Statistical tests have to assume certain data distributions which may be incorrect sometimes. Depth-based methods have high computational complexity, specifically for multivariate data with more than three dimensions.

In this paper, we used distance-based outlier detection. For this purpose, the distances to the k nearest neighbors of each data record are accumulated. Data records having high distance score are considered as outliers. For reducing the computational costs, an improved K-NN based outlier detection method is used as described in.[43] This method clusters data records and removes clusters having no outlier. It reduces the computational efforts to find k-nearest neighbors of each data record.

After detecting outliers, they are replaced with their adjacent members of the time series. Final step of preprocessing data is normalizing it to [0, 1]. Objective functions will not work properly without normalization because attributes with large ranges may dominate attributes with small ranges.[40] Time series extracted from the laparoscopic video, are verified for outlier detection and removal as shown in Figure 6.

Samples of the extracted data sets are shown in Figure 6 (a-1, a-2, a-3, a-4 and a-5 respectively). Moreover, the distances to three nearest neighbors of each data record

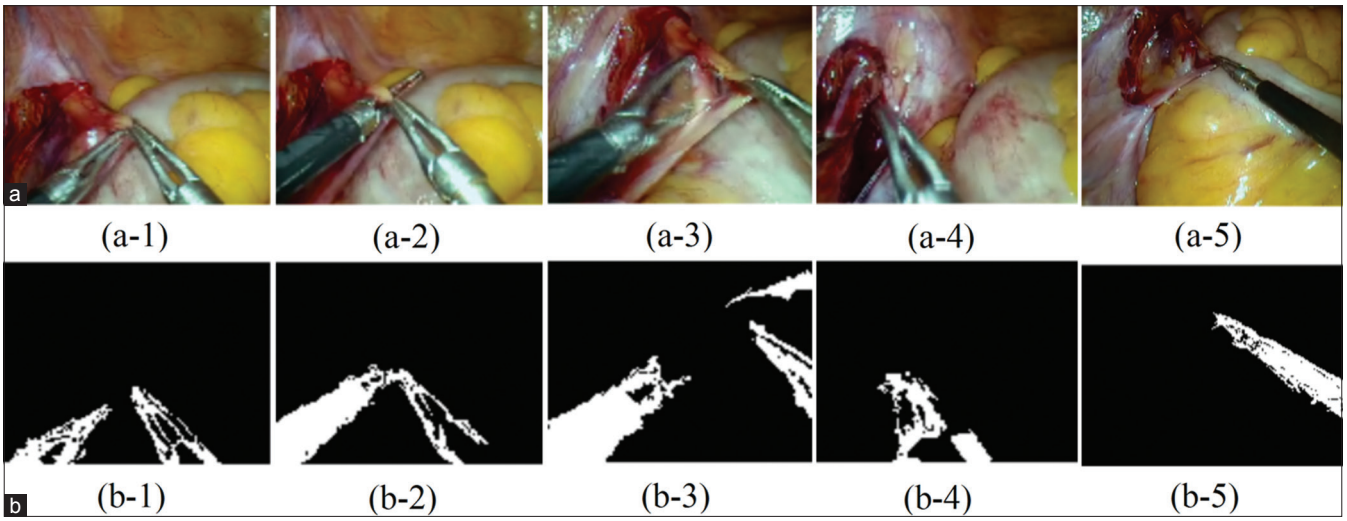

Figure 5: (a) Some laparoscopic video frames. (b) Their corresponding segmented image generated by surgical instrument segmentation algorithm algorithm

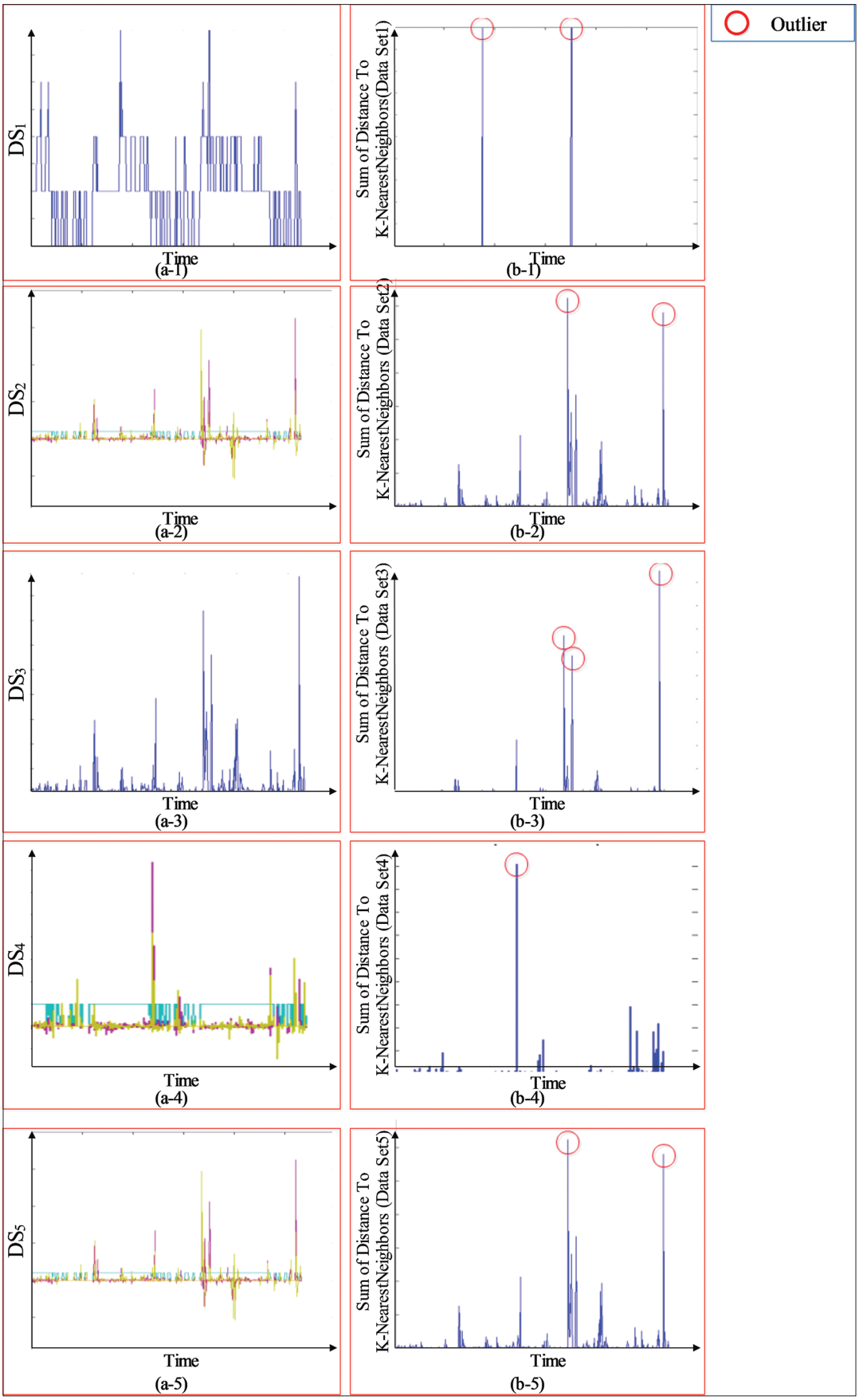

Figure 6: (a) Samples of extracted data sets. (b) Their corresponding outliers found by outlier detection method

are accumulated and displayed in Figure 6 (b-1, b-2, b-3, b-4 and b-5 respectively). Outliers detected in the data sets are illustrated by red circles in Figure 6 (b-1, b-2, b-3, b-4 and b-5 respectively).

Outliers of the data set DS_1 can be identified based on expert knowledge, too. DS_1 indicates the number of visible surgical instruments. The maximum number of visible surgical instruments in each kind of laparoscopic procedure is limited and known. Therefore, this knowledge can be used for outlier detection in DS_1.

For example, the time series shown in Figure 5 is extracted from a varicocele laparoscopic surgery. Expert knowledge indicates that four instruments cannot be inside body at the same time in varicocele.

Selecting the Most Relevant Cost Functions for Temporal Segmentation of Extracted Data Sets

In this paper, the laparoscopic video will be temporally segmented based on the data sets extracted to describe motion and number of surgical instruments.

Temporal segmentation aims to detect homogeneous segments in time-series data. Therefore, temporal segmentation is an optimization problem in which the objective (cost) function measures homogeneity of the segments. The results of segmentation are highly dependent to the selected cost function. Finding the appropriate cost function for temporal segmentation is a challenging task, specifically when it is performed in an unsupervised manner. In this paper, a novel algorithm is proposed to find the appropriate cost functions for measuring the segment homogeneity.

In the proposed algorithm, GA is used for segmenting time series.

An individual of GA is represented by n chromosomes which n is the length of the time series. Each chromosome of GA is a binary variable. If the value of a chromosome is 1, this chromosome is a border point of a segment. Otherwise, the chromosome is an internal point of a segment. Cross-over and mutation operators are defined as usual. However, a constraint is defined on the number and length of allowed segments. It prevents GA from generating a few long segments or too many short segments.

The proposed algorithm is executed several times. Each time, GA attempts to optimize a candidate cost function, selected among various cost functions used in the previous studies.

The correlation coefficient is calculated between the cost function of the solutions visited by GA and their corresponding performance measures.

Previous studies investigate the performance of their proposed methods for temporal segmentation of videos using detection rate, recognition rate and accuracy.[44] Hence, in this paper, performance measures, including detection rate, recognition rate and accuracy, are considered.

Since, all candidate objective functions used in time-series segmentation are cost functions, minimizing them is desired. On the other hand, the performance measures used in this paper are positive indicators and maximizing them is desired. For this reason, candidate objective functions having negative correlation with all mentioned performance measures are the appropriate cost functions. Therefore, these cost functions are selected as objective functions of MOGA, used in the next step of MIVS.

The main steps of the proposed algorithm to find the appropriate cost functions are as follows:

Segmentation Module with GA (DS_j, CF_i):

Step 1: Segment DS_j using GA to minimize CF_i. It is supposed that CF_{ij} is the cost function CF_i applied to the data set DS_j

Step 2: Calculate detection rate, recognition rate and accuracy of each solution visited by GA are calculated

Step 3: Compute the correlation coefficients between the mentioned performance measures and their corresponding values of CF_{ij}

Step 4: Add CF_{ij} to the list SelectedCostFunctions if it has negative correlation with all the mentioned performance measures.

Let us assume C and D are the number of candidate cost functions and data sets respectively. Segmentation module with GA is executed C × D times, each time with a different combination of parameter settings.

Previous researches introduced and used several cost functions for time-series segmentation. In this paper, three cost functions, CF_1, CF_2 and CF_3, are used as candidate cost functions:

CF_1

Cost function CF_1 is the sum of the costs of the individual segments. Cost of a segment is defined as the sum of variances of the variables in the segment.[27,37,38] CF_1 is calculated as:

$$CF_1 = \sum_{i=1}^{SegNo} \frac{\sum_{k=a_i}^{b_i} || x_k - m_i ||^2}{b_i - a_i + 1} \quad (1)$$

Where $X = \{x_k: 1 \leq k \leq n\}$ is a time-series data set

extracted from the laparoscopic video. m_i denotes the mean of the segment *i*. a_i and b_i are borders of the segment *i*.

CF_2

Cost function CF_2 considers the relations between variables of multivariate time series. CF_2 is defined based on covariance of variables in each segment as:[27,36]

$$CF_2 = \sum_{i=1}^{SegNo} \frac{1}{b_i - a_i + 1} \sum_{k=a_i}^{b_i} s_{cov}(P_k, P_{s_i}) \quad (2)$$

Where P_k is covariance matrix describing the relationship between the variables around the k^{th} data point (x_k). P_{si} is the covariance matrix of the segment *i* with borders a_i and b_i.[27] Details about how to calculate P_k and P_{si} have been presented elsewhere.[27]

CF_3

Another cost function CF_3 for multivariate data is introduced, too. CF_3 is defined based on the homogeneity measure of temporal segments presented by Cavallo.[45] Cavallo have defined the homogeneity measure of the segment as the ratio of the second component of singular value decomposition to the first one (σ_2/σ_1).[45] If this ratio is not less than a very low threshold (about 0.05), the segment is not homogeneous enough. This measure is appropriate for recognition of nonlinear patterns occurring in time series and it is useful for detection of trend changes in time series. It is used for segmentation based on SW, too.[45] Therefore, the ratio of (σ_2/σ_1) is considered in a new cost function CF_3 as:

$$CF_3 = \sum_{s=1}^{SegNo} (b_i - a_i + 1).\frac{\sigma_2(x_{a_i} : x_{b_i})}{\sigma_1(x_{a_i} : x_{b_i})} \quad (3)$$

Since CF_3 is zero for the segments with length of 1 and segments with length of 1 are not desired, a constraint is defined on the allowed total number of segments as:

$$0.06.n \le segNo \le 0.2.n \quad (4)$$

Where n is the length of the time series. This constraint is defined based on the expert knowledge. It is verified for 25 laparoscopic videos.

After executing GA to minimize the cost function, the performance measures of the visited solutions are calculated. The performance measures include detection rate, recognition rate and accuracies.

Detection rate measures the ratio of correct detections of segment boundaries in a solution visited by GA to the total number of segments in human segmented time series. Detection rate is computed as:

$$\text{Detection rate} = \frac{\#\text{Correct detections}}{\#\text{Ground-truth boundaries}} \quad (5)$$

For determining the ground-truth solution, human experts analyze laparoscopic video frames manually and determine the segment boundaries. Therefore, a binary time-series is generated in which 0 and 1 elements are corresponding to internal and border points of the segments, respectively. The data set will be used as the ground-truth data to evaluate the performance of our proposed method.

Recognition rate measures the ratio of correct detections of segment boundaries in a solution visited by GA to its total number of segments. Recognition rate is computed as:

$$\text{Recognition rate} = \frac{\#\text{Correct detections}}{\#\text{Boundaries of a GA solution}} \quad (6)$$

Accuracy is a performance measure defined as the harmonic average of detection rate and recognition rate. Accuracy is calculated as:

$$\text{Accuracy} = 2.\left(\frac{1}{\text{Detection rate}} + \frac{1}{\text{Recognition rate}}\right)^{-1} \quad (7)$$

Let CF_{ij} be the cost function CF_i computed from data set DS_j. Among all CF_{ij}'s, those having negative correlations with all mentioned performance measures are considered as the objective functions of MOGA method. For this purpose, Pearson's correlation coefficient is computed as:

$$R = \frac{\text{cov}(X,Y)}{\sigma_X.\sigma_Y} \quad (8)$$

R is the Pearson's correlation coefficient which is always between -1 and $+1$. In this paper, X and Y are values of the cost function CF_{ij} and one of the mentioned performance measures computed from the solution of temporal segmentation, respectively.

Segmentation Module with GA algorithm is executed several times. Each time a data set and a candidate cost function are passed to the algorithm as input parameters. Table 1 summarizes the results of executing Segmentation Module with GA algorithm with different parameter settings.

For each data set and each cost function, the best solution visited by GA is evaluated. For this purpose, the value of the cost function is determined for the best solution of GA and the ground-truth segmented solution. Moreover, the maximum correlation coefficient between the cost function of solutions visited by GA and their corresponding performance measures is listed in Table 1.

As mentioned before, the cost functions having negative correlation with all the mentioned performance measures

Table 1: Comparing the ground-truth solution with GA best solution in terms of the cost function and calculating the maximum correlation between the cost function and performance measures of GA visited solutions

Data	Cost function	CF of ground-truth segmentation	CF of GA best solution	Max (R_{ijk})
DS_1	CF_1	0.1389	0.0355	0.66
DS_2	CF_1	5.65	2.21	0.5443
	CF_2	0.09	0.0533	0.2103
	CF_3	0.1205	0.1321	−0.3585
DS_3	CF_1	0.1655	0.0138	0.1995
DS_4	CF_1	5.82	1.07	0.7131
	CF_2	2.52	1.55	0.5044
	CF_3	0.1553	0.1598	−0.5378
DS_5	CF_1	0.3113	0.06	0.0569
DS_1,	CF_1	0.269	0.09	0.015
DS_3	CF_2	0.2342	0.2187	0.4153
	CF_3	0.0676	0.0721	−0.4038
DS_1,	CF_1	0.3926	0.0918	0.6578
DS_5	CF_2	0.2123	0.2291	0.2424
	CF_3	0.1109	0.1164	−0.4925
DS_3,	CF_1	0.444	0.0694	0.02
DS_5	CF_2	0.247	0.0459	0.4343
	CF_3	0.1631	0.1337	−0.5259
DS_1,	CF_1	0.5166	0.1533	0.474
DS_3,	CF_2	0.3142	0.152	0.633
DS_5	CF_3	0.1253	0.1378	−0.7171

GA – Genetic algorithm; CF – Cost function

are desired. Among the verified cost functions, CF_3 has negative correlation with the data sets. However, CF_3 can be calculated only for multivariate data sets. Moreover, the only cost function examined on univariate data sets is CF_1 which has positive correlation for all univariate data sets. Therefore, univariate data sets are combined column-wise to generate multivariate ones. Then, CF_3 is computed from these new data sets.

As shown in Table 1, CF_3 values of the ground-truth solutions are less than those of the best solutions of GA. It is concluded that the ground-truth segmentation is an optimal or near-optimal solution in terms of CF_3.

Temporal Segmentation with MOGA

In the previous step, the most appropriate cost functions are selected for temporal segmentation of the extracted data sets. In this step, a multi objective GA is designed to segment data sets with the aim of minimizing the selected cost functions. The objective function of MOGA is the linear weighted combination of the selected cost functions as:

$$\text{MOGA cost function} = \sum_{j=1}^{D}\sum_{i=1}^{C}\left(W_{ij}.\delta_{ij}\right) \quad (9)$$

Where C and D are the number of cost functions and data sets respectively. W_{ij} is the weight assigned to cost function CF_{ij}. The value of W_{ij} in (9) is specified as:

$$W_{ij} = \begin{cases} 0 & \max\limits_{1\le k\le 3}\{R_{ijk}\} \ge 0 \\ \min\limits_{1\le k\le 3}\{|R_{ijk}|\} & \max\limits_{1\le k\le 3}\{R_{ijk}\} < 0 \end{cases} \quad (10)$$

R_{ijk} is the correlation coefficient between CF_{ij} and the kth performance measure.

δ_{ij} in (9) is computed as:

$$\delta_{ij} = CF'_{ij} - CF''_{ij} \quad (11)$$

Where CF'_{ij} is the value of CF_{ij} of the best solution visited by GA. CF''_{ij} is the value of CF_{ij} of the ground-truth segmented solution.

Cost functions having non-negative correlation with some of the performance measures are not considered in the MOGA objective function. Because contributing them to objective function of MOGA has a negative impact on the performance of temporal segmentation.

The main steps of MOGA algorithm for segmenting time series are as follows:

Step 1: Consider CF_{ij}'s having negative correlation to all mentioned performance criteria
Step 2: For each considered CF_{ij}, determine the weight W_{ij} as (10)
Step 3: Normalize weights
Step 4: Define the objective function as (9)
Step 5: Run GA for minimizing the objective function defined in Step 4.

In MOGA, concepts, including individuals, chromosomes, cross-over and mutation, are defined as described in section 3.C.

RESULTS

Data analyzed in this paper is collected from hasheminejad kidney center and includes varicocele and ureteropelvic junction obstruction (UPJO) laparoscopic videos archived in (2010-2011). Varicocele and UPJO are selected as representatives of simple and complex laparoscopic urological procedures respectively. The collected data sets are analysed with the proposed method in this paper.

Data extraction procedure is applied to two consecutive video frames to extract one data record of each data set. Among four consecutive video frames, one data record is extracted. Therefore, the length of each data set extracted from a video with N frames and frame rate of 25 fps is about N/4. This rate of sampling is determined based on the expert's opinions and the reaction rate of the surgeons.

After finding the most appropriate cost functions among the candidates, MOGA is executed on the extracted data sets

to segment these data sets. The mentioned performance measures of a visited solution are computed via comparing its segment boundaries to those of the ground-truth solution.

Figure 7 illustrates a sample of a data extracted from a laparoscopic varicocele video. [Figure 7a] illustrates the sample segmented by human expert and [Figure 7b] indicates its corresponding best solution obtained by MOGA.

Figure 7a and b are very similar to each other. The segment boundaries are represented by solid vertical lines. Lines denoted in green indicate common segment boundaries between the compared samples. As depicted in Figure 7, only one segment boundary (4th border line denoted in brown) is misrecognized in the sample.

Moreover, the performances of GA or MOGA segmentation methods are compared for different combinations of data sets and cost functions as shown in Table 2.

All data sets are combined column-wise and stored in MOGA data set. MOGA_CF is weighted linear combination of CF_{ij}'s, computed as (9). As shown in Table 2, the best performance

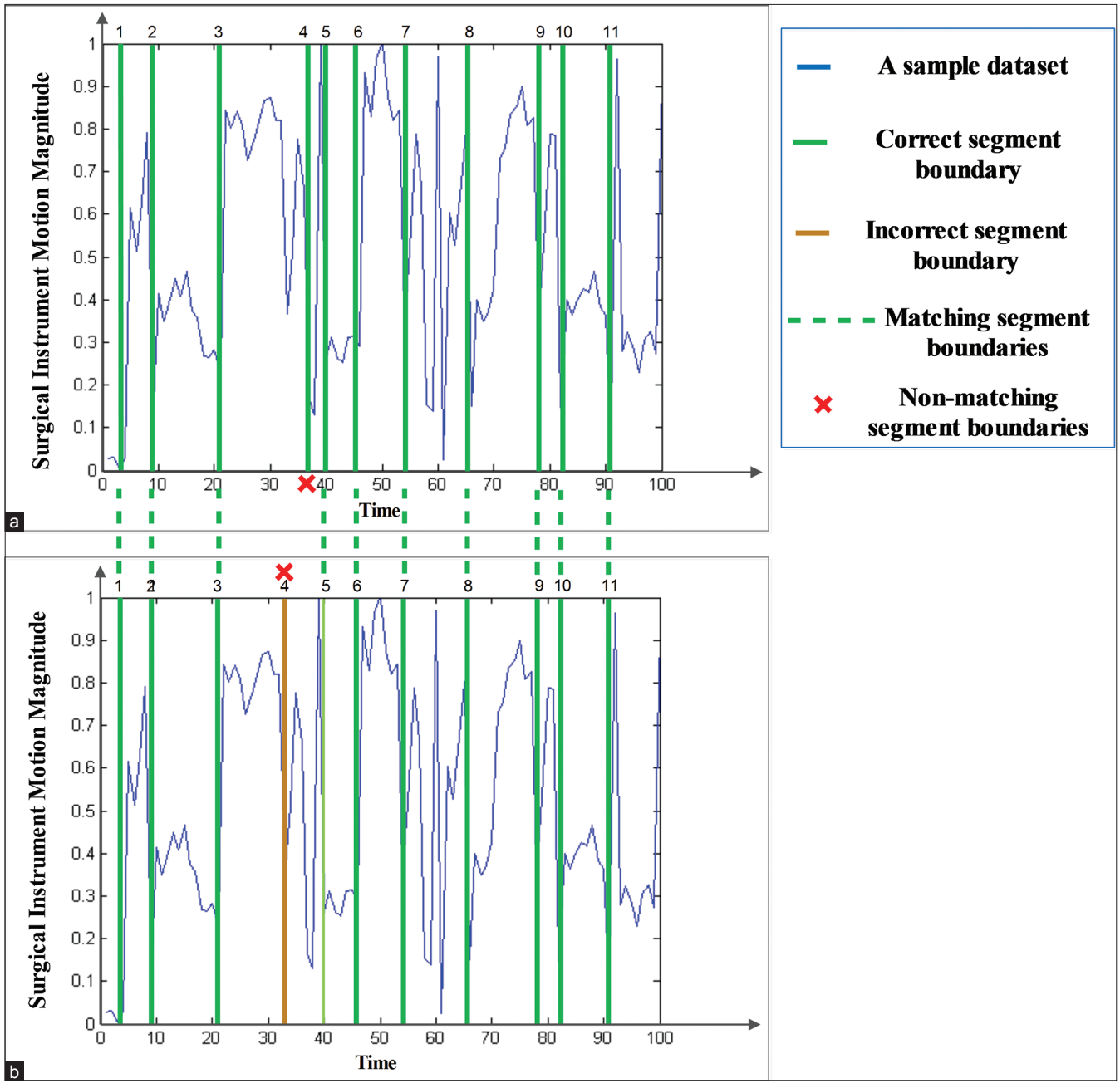

Figure 7: A sample data extracted from varicocele video: (a) Sample segmented by human. (b) Its corresponding best solution obtained by multi-objective genetic algorithm

measures are obtained by segmenting the MOGA data to optimize MOGA_CF.

Figure 8 indicates a sample of temporal segments of a real laparoscopic video. Moreover, the data sets extracted from the sample are illustrated in Figure 8.

The video frames illustrated by Figure 8a are some sample frames of different segments. The sample sequence consists of four different segments, including S_i, S_{i+1}, S_{i+2} and S_{i+3}. Only a few video frames of each segment are illustrated by Figure 8a.

Segment S_i indicates the grasping activity performed by the surgeon. Segment S_{i+1} and S_{i+3} indicate the incision activity. Segment S_{i+2} is a surgical interruption in which no motion and activity are being performed. Data sets DS_2, DS_3, DS_4 and DS_5 have near-zero value in this segment while all data sets have non-zero values in segments S_i, S_{i+1} and S_{i+3}.

As illustrated by Figure 8b, the pattern of all extracted data sets of segment S_{i+1} is different from those of its adjacent segments S_i and S_{i+2}. Furthermore, data sets 2-5 of segment S_{i+2} are different from those of segment S_{i+3}. Therefore, it is concluded that the extracted data sets can discriminate the adjacent segments of the sample.

For evaluating the performance of MIVS, it is compared to some state-of-the-art methods, including change detection method,[34] DPCA-based temporal segmentation[36] and SW method with cost function of (σ_2/σ_1).[45]

Figure 9 indicates the results of several segmentation methods applied to the data sets of Figure 8b.

As illustrated by Figure 9, change detection method[34] splits segment S_{i+1} into two smaller segments. Moreover, DPCA-based method[36] and Cavallo's method[45] shift the boundary of segment S_{i+1} right. In another word, these two methods assign a few first members of the segment S_{i+2} to the segment S_{i+1} mistakenly. The reason is that they are highly sensitive to the threshold selected for identifying the segment boundaries. However, as illustrated by Figure 9, MIVS can detect the boundaries of all mentioned segments accurately.

There is no benchmark data set in the previous papers can be used for validating MIVS. For this reason, all the compared methods are examined on each subset of all data sets extracted in this paper and the best solution obtained by each method is considered. The performances of MIVS and the compared methods in temporal segmentation of laparoscopic videos are listed in Table 3.

Table 2: Comparing the performance of GA or MOGA segmentation methods for different combinations of data sets and cost functions

Data	Cost function	Detection rate	Recognition rate	Accuracy
DS_1	CF_1	67.11	60.43	63.59
DS_2	CF_1	66.46	62.39	64.36
	CF_2	65.23	62.11	63.63
	CF_3	73.61	71.25	72.41
	MOGA_CF	86.19	83.12	84.63
DS_3	CF_1	64.30	61.48	62.86
DS_4	CF_1	68.35	64.03	66.12
	CF_2	66.55	63.61	65.05
	CF_3	75.19	72.82	73.99
	MOGA_CF	88.19	83.04	85.54
DS_5	CF_1	64.20	61.73	62.94
DS_1, DS_3	CF_1	67.32	63.58	65.40
	CF_2	67.01	62.92	64.90
	CF_3	75.14	71.58	73.32
	MOGA_CF	86.17	83.22	84.67
DS_3, DS_5	CF_1	68.04	63.60	65.74
	CF_2	67.81	63.25	65.45
	CF_3	78.62	73.44	75.94
	MOGA_CF	88.36	83.01	85.60
DS_1, DS_3, DS_5	CF_1	69.00	63.91	66.36
	CF_2	68.33	63.38	65.76
	CF_3	80.20	75.51	77.78
	MOGA_CF	90.12	85.57	87.79
MOGA data	CF_1	75.27	72.12	73.66
	CF_2	75.05	71.84	73.41
	CF_3	89.69	85.22	87.40
	MOGA_CF	95.59	94.20	94.89

GA – Genetic algorithm; MOGA – Multi-objective genetic algorithm

Table 3: Comparing performances of MIVS and the compared methods in temporal segmentation of laparoscopic videos

Segmentation algorithm	Detection rate	Recognition rate	Accuracy
Change detection method[34]	88.04	78.73	83.12
DPCA-based segmentation[36]	89.4	83.66	86.43
SW with cost function of σ_2/σ_1[45]	88.21	79.52	83.64
MIVS	95.59	94.20	94.89

MIVS – Minimally-invasive videos; DPCA – Dynamic principal component analysis; SW – Sliding windows

As listed in Table 3, MIVS outperforms compared methods in terms of detection rate, recognition rate and accuracy of segmentation.

Finally, the average execution time of MIVS and the compared methods are computed and listed in Table 4.

As listed in Table 4, MIVS is slower than the compared methods in temporal segmentation of laparoscopic video. However, all methods are applied to the same data set extracted by the proposed method. As listed in Table 4, the data extraction from video is much slower than segmenting the extracted data.

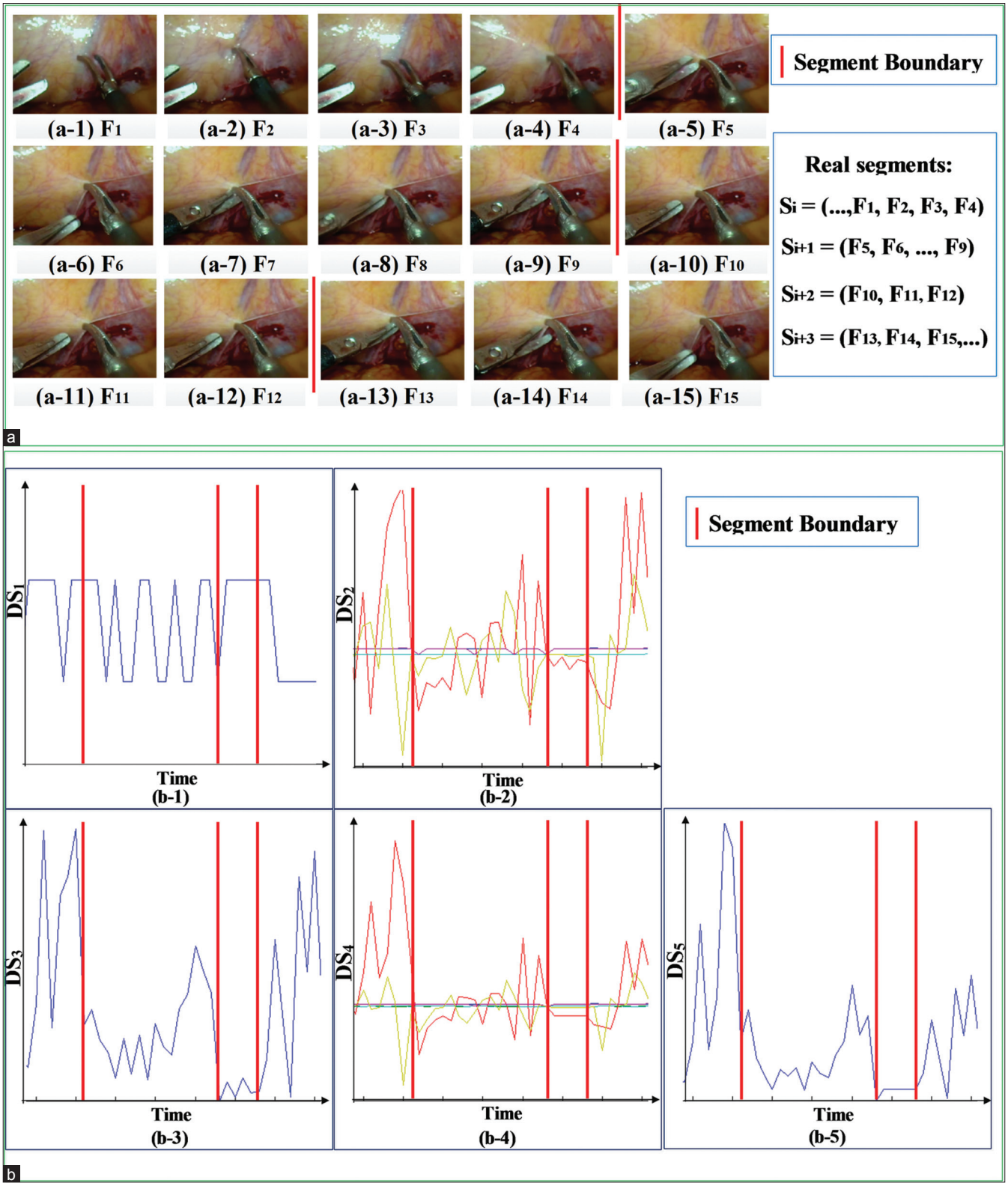

Figure 8: A sample of temporal segments of a real laparoscopic video: (a) Sample frames of different temporal segment. (b) Data sets extracted from the sample segments

Parallel processing methods, such as pipeline computing, can be used for reducing total run time of MIVS. Therefore, only data extraction method has the most significant on the execution time. For this reason, MIVS is not much worse than the compared methods in terms of the execution time.

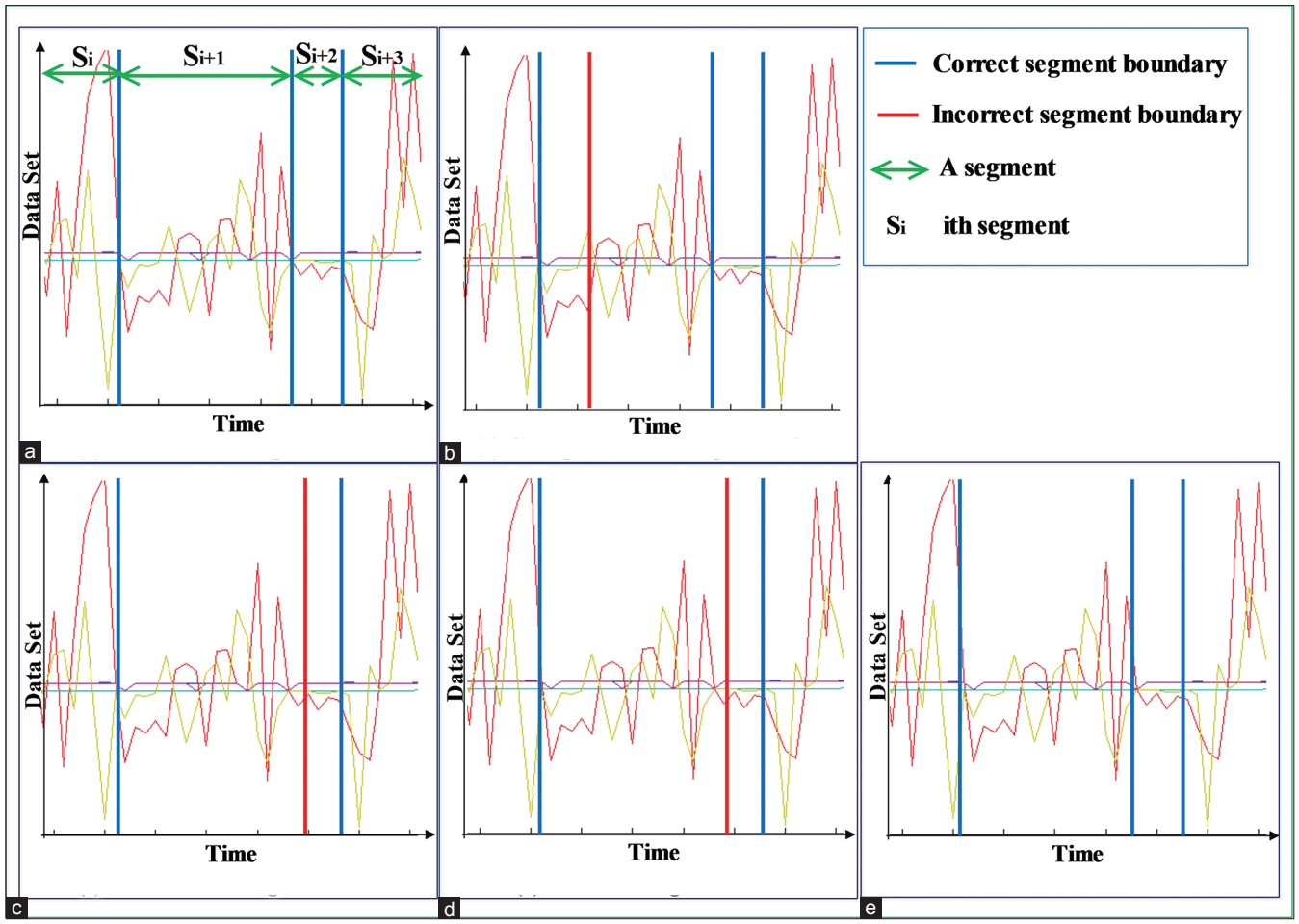

Figure 9: Results of minimally-invasive videos (MIVS) and other compared methods applied to a real sample: (a) Ground-truth segmented sequence. (b) Sequence segmented by change-detection method. (c) Dynamic principal component analysis-based segmented sequence. (d) Sequence segmented by Cavallo's method. (e) MIVS-based segmented sequence

Table 4: Comparing the average execution time of MIVS and the compared methods in temporal segmentation of laparoscopic videos

Segmentation algorithm	CPU time (s) for 1 min of laparoscopic video		
	Data extraction	Temporal segmentation	Total time
Change detection method[34]	156.9	9.6	166.5
DPCA-based segmentation[36]	156.9	38.3	195.2
SW with cost function of σ_2/σ_1[45]	156.9	7.1	164.0
MIVS	156.9	52.4	209.3

MIVS – Minimally-invasive videos; CPU – Central processing unit; DPCA – Dynamic principal component analysis; SW – Sliding windows

CONCLUSION

In this paper, a novel method for temporal segmentation of MIVS is introduced. The proposed method has the potential to be applied for anomaly detection, disruption and interrupt identification. Temporal segmentation of video is a prerequisite step for laparoscopic action recognition. Instead of extracting only one data set and segmenting video based on this data set, several data sets are extracted from video in this paper. Experimental results show that considering several data sets can improve the accuracy of video temporal segmentation.

Moreover, temporal segmentation is an optimization problem on which choosing the appropriate objective function is very important. Therefore, various cost functions are examined. The cost functions having negative correlation with positive performance measures are considered as selected objective functions. A MOGA is designed and implemented for temporal segmentation of the extracted data sets. The proposed method has reached to the best solution with detection rate of 95.59%, recognition rate of 94.20% and accuracy of 94.89%, which is so promising.

In this paper, three different cost functions are examined. One possible extension of this work can be examining more complex cost functions for MOGA. Moreover, the examined cost function for univariate data doesn't have negative correlation with the mentioned performance measures.

Therefore, we have to combine univariate data sets to create multivariate ones. So, finding the appropriate cost functions for univariate data is another possible extension of this work. Final proposed future direction is to recognize the surgical actions occurring in the identified segments of laparoscopic video.

ACKNOWLEDGMENTS

The authors wish to thank Hasheminejad Kidney Center which gives their laparoscopic surgical videos to authors and allows authors to arrive in the operating rooms.

REFERENCES

1. Bartoli A, Collins T, Bourdel N, Canis M. Computer assisted minimally invasive surgery: Is medical computer vision the answer to improving laparosurgery? Med Hypotheses 2012;79:858-63.
2. Velanovich V. Laparoscopic vs open surgery: A preliminary comparison of quality-of-life outcomes. Surg Endosc 2000;14:16-21.
3. Blum T, Feussner H, Navab N. Modeling and segmentation of surgical workflow from laparoscopic video. Med Image Comput Comput Assist Interv 2010;13:400-7.
4. Chen Y, Lee J. A review of machine-vision-based analysis of wireless capsule endoscopy video. Diagn Ther Endosc 2012;2012:418037.
5. Mackiewicz M, Berens J, Fisher M. Wireless capsule endoscopy video segmentation using support vector classifiers and hidden markov models. Medical Image Understanding and Analyses Conference; 2006.
6. Padoy N, Blum T, Ahmadi SA, Feussner H, Berger MO, Navab N. Statistical modeling and recognition of surgical workflow. Med Image Anal 2012;16:632-41.
7. Oropesa I, Sánchez-González P, Chmarra MK, Lamata P, Fernández A, Sánchez-Margallo JA, *et al*. EVA: Laparoscopic instrument tracking based on endoscopic video analysis for psychomotor skills assessment. Surg Endosc 2013;27:1029-39.
8. Climent J, Hexsel RA. Particle filtering in the Hough space for instrument tracking. Comput Biol Med 2012;42:614-23.
9. Doignon C, Graebling P, Mathelin D. Real-time segmentation of surgical instruments inside the abdominal cavity using a joint hue saturation color feature. Real Time Imaging 2005;11:429-42.
10. Voros S, Long JA, Cinquin PH. Detection of instruments in laparoscopic images: A first step towards high-level command of robotic endoscopic holders. Robotic Res 2007;26:1173-90.
11. Brox T, Bruhn A, Papenberg N, Weickert J. High accuracy optical flow estimation based on a theory for warping. 8th European Conference on Computer Vision; 2004.
12. Danafar S, Gheissari N. Action recognition for surveillance applications using optic flow and SVM. Asian Conference on Computer Vision; 2007. p. 457-66.
13. Dollar P, Rabaud V, Cottrell G, Belongie S. Behavior recognition via sparse spatio-temporal features. International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance; 2005. p. 65-72.
14. Holte MB, Moeslund TB, Fihl P. View-invariant gesture recognition using 3D optical flow and harmonic motion context. Comput Vis Image Underst 2010;114:1353-61.
15. Bobick AF, Davis JW. The recognition of human movement using temporal templates. IEEE Trans Pattern Anal Mach Intell 2001;23:257-67.
16. Meng H, Pears N. Descriptive temporal template features for visual motion recognition. Pattern Recognit Lett 2009;30:1049-58.
17. Ahad MA, Ogata T, Tan JK, Kim H, Ishikawa S. Motion recognition approach to solve overwriting in complex actions. International Conference on Automatic Face and Gesture Recognition; 2008. p. 1-6.
18. Pers J, Sulic V, Kristan M, Perse M, Polanec K, Kovacic S. Histograms of optical flow for efficient representation of body motion. Pattern Recognit Lett 2010;31:1369-76.
19. Wang H, Klaser A, Schmidt C, Liu CL. Action recognition by dense trajectories. IEEE Conference on Computer Vision and Pattern Recognition; 2011.
20. Lowe DG. Object recognition from local scale-invariant features. Int Conf Comput Vis 1999;2:1150-7.
21. Bay H, Tuytelaars T, Gool V. SURF: Speeded up robust features. The Ninth European Conference on Computer Vision; 2006.
22. Harris C, Stephens M. A combined corner and edge detector. Paper Presented at: The 4th Alvey Vision Conference; 1988.
23. Rosten E, Drummond T. Fusing points and lines for high performance tracking. Paper Presented at: the IEEE International Conference on Computer Vision; 2005.
24. Shi J, Tomasi C. Good features to track. Paper Presented at: The IEEE Conference on Computer Vision and Pattern Recognition; 1994.
25. Hartley R, Ziserman A. Multiple View Geometry in Computer Vision. 2nd ed. UK: Cambridge university press 2004.
26. Yang F, Wei L, Zhang Z, Tang H. Image mosaic based on phase correlation and Harris operator. J Comput Inf Syst 2012;8: 2647-55.
27. Feil B, Abonyi J, Németh S, Árva P. Monitoring process transitions by Kalman filtering and time-series segmentation. Comput Chem Eng 2005;29:1423-31.
28. Albu AB, Bergevin R, Quirion S. Generic temporal segmentation of cyclic human motion. Pattern Recognit 2008;41:6-21.
29. Chung A. An evolutionary approach to pattern-based time series segmentation. IEEE Trans Evol Comput 2004;8:471-89.
30. Tseng VS, Chen C, Huang PC, Hong TP. Cluster-based genetic segmentation of time series with DWT. Pattern Recognit Lett 2009;30:1190-7.
31. Weng S, Liu YH. Mining time series data for segmentation by using ant colony optimization. Eur J Oper Res 2006;173:921-37.
32. Weinland D, Ronfard R, Boyer E. A survey of vision-based methods for action representation, segmentation and recognition. Comput Vis Image Underst 2011;115:224-41.
33. Gould K, Rangarajan K, Shah M. Detection and representation of events in motion trajectories. In Gonzalez and Mahdavieh, editors, Advances in Image Processing and Analysis. SPIE Optical Engineering Press; 1992.
34. Briassouli A, Tsiminaki V, Kompatsiaris I. Human motion analysis via statistical motion processing and sequential change detection. Journal on Image and Video Processing 2009; DOI is: 10.1155/2009/652050.
35. Yu J, Yin J, Zhang J. Comparison of distance measures in evolutionary time series segmentation. Third International Conference on Natural Computation (ICNC); 2007.
36. Dobos L, Abonyi J. On-line detection of homogeneous operation ranges by dynamic principal component analysis based time series segmentation. Chem Eng Sci 2012;72:96-105.
37. Himberg J, Korpiaho K, Mannila H, Tikanmaki J, Toivonen HT. Time-series segmentation for context recognition in mobile devices. IEEE International Conference on Data Mining (ICDM); 2001. p. 203-10.
38. Vasko K, Toivonen HT. Estimating the number of segments in time series data using permutation tests. IEEE International Conference on Data Mining; 2002. p. 466-73.
39. Laptev I, Caputo B, Schüldt C, Lindeberg T. Local velocity-adapted motion events for spatio-temporal recognition. Comput Vis Image

Underst 2007;108:207-29.
40. Han J, Kamber M, Pei J. Data Mining: Concepts and Techniques. United States of America: Morgan Kauffman; 2012.
41. Hodge VJ, Austin J. A survey of outlier detection methodologies. Artif Intell Rev 2004;22:85-126.
42. Kriegel HP, KP, Zimek A. Outlier detection techniques. Tutorial at 16th ACM SIGKDD Conference on Knowledge Discovery and Data Mining; 2010.
43. Wang Q, Zheng M. An improved KNN based outlier detection algorithm for large datasets. Advanced Data Mining and Applications 2010; 6440: 585-592.
44. Shao L JL, Liu Y, Zhang J. Human action segmentation and recognition via motion and shape analysis. Pattern Recognit Lett 2012;2:438-45.
45. Cavallo A. Primitive actions extraction for a human hand by using SVD. IEEE 9th International Symposium on Intelligent Systems and Informatics; 2011.

How to cite this article: Khatibi T, Sepehri MM, Shadpour P. A Novel Unsupervised Approach for Minimally-invasive Video Segmentation. J Med Sign Sens 2014;4:53-71

Source of Support: Nil, **Conflict of Interest:** None declared

APPENDIX A: EVALUATING SISA METHOD FOR SURGICAL INSTRUMENT IDENTIFICATION

SISA is a surgical instrument segmentation method proposed in this paper. We believe that SISA can alleviate many of the mentioned challenges and complexities existing in laparoscopic video frames. For evaluating the performance of SISA, it is compared against some other instrument segmentation methods.[1-3] For this purpose, the precision of SISA and three compared methods is computed for surgical instrument detection. The precision is calculated as:

$$\text{Precision} = \frac{\text{TP}}{N} \quad (1)$$

Where TP indicates number of correctly identified surgical instruments and N indicates total number of visible surgical instruments in the examined laparoscopic video frames.

The performance of SISA is compared with other proposed surgical instrument detection methods in Table 1.

Table 1: Comparing the performance of SISA algorithm with other surgical instrument detection methods

Algorithm	Precision of laparoscopic instrument detection (%)
Climent and Hexsel algorithm[1]	88.1
Doignon *et al*. algorithm[2]	83.8
Voros *et al*. segmentation algorithm[3]	86.7
SISA algorithm	92.5

SISA – Surgical instrument segmentation algorithm

As illustrated by Table A, SISA outperforms compared methods for surgical instrument detection.

REFERENCES

1. Climent J, Hexsel RA. Particle filtering in the Hough space for instrument tracking. Comput Biol Med 2012;42:614-23.
2. Doignon C, Graebling P, Mathelin D. Real-time segmentation of surgical instruments inside the abdominal cavity using a joint hue saturation color feature. Real Time Imaging 2005;11:429-42.
3. Voros S, Long JA, Cinquin PH. Detection of instruments in laparoscopic images: A first step towards high-level command of robotic endoscopic holders. Robotic Res 2007;26:1173-90.

BIOGRAPHIES

Toktam Khatibi, Ph.D. Candidate in Department of Industrial Engineering of Tarbiat Modares University, Tehran, Iran. Ms. Khatibi has received her BSc. degree in Software Engineering from Sharif University of Technology, Tehran, Iran, in 2003, and her MSc. degree in Socio-Economic Systems Engineering from Tarbiat Modares University, Tehran, Iran, in 2008. Her research interests include Healthcare, Image Processing, Pattern Recognition, Data Mining and Artificial Intelligence.

E-mail: toktam.khatibi@modares.ac.ir

Mohammad Mehdi Sepehri, Associate professor, Department of Industrial Engineering, Tarbiat Modares University (TMU), Tehran, Iran; Founder and Head of The Laboratory for Healthcare Systems Optimization, Engineering, and Informatics (HCSE); Head of Group of Industrial Engineering; Editor-in-Chief, International Journal of Hospital Research, the official peer review publication of Hospital Management Research Center (HMRC), Iran University of Medical Sciences. Dr. Sepehri obtained his MSc and PhD in Management Science from the University of Tennessee, Knoxville, Tennessee, USA, in 1987 and 1991. His current research focuses on Operations Research in Medicine and Healthcare, Healthcare Systems Engineering, Mobile Health, Lean Healthcare and Lean Hospital.

E-mail: mehdi.sepehri@modares.ac.ir

Pejman Shadpour, Associate Professor of Urology, Iran University of Medical Science; Deputy for Education and Research, Hasheminejad Kidney Center; Director of Hospital Management Research Center (HMRC), Iran University of Medical Sciences. Dr. Shadpour received M.D. and Ph.D. (Specialty Degree in Urology) from the School of Medicine, Iran University of Medical Science (IUMS), Tehran, Iran, in 1992 and 1998. He has completed his Clinical Fellowship in Pediatric Urology and Laparoscopy, Shahid Beheshti University of Medical Science, Tehran, Iran, in 2003.

E-mail: pshadpour@iums.ac.ir