Real-time segmentation of surgical instruments inside the abdominal cavity using a joint hue saturation color feature

C. Doignon*, P. Graebling, M. de Mathelin

Laboratoire des Sciences de l'Image, de l'Informatique et de la Télédétection (UMR ULP-CNRS 7005), Control, Vision and Robotics Group, University of Strasbourg, Bd. Brant, 67400 Illkirch, France

Available online 18 August 2005

Abstract

In this paper, the real-time segmentation of surgical instruments with color images used in minimally invasive surgery is addressed. This work has been developed in the scope of the robotized laparoscopic surgery, specifically for the detection and tracking of gray regions and accounting for images of metallic instruments inside the abdominal cavity. With this environment, the moving background due to the breathing motion, the non-uniform and time-varying lighting conditions and the presence of specularities are the main difficulties to overcome. Then, to achieve an automatic color segmentation suitable for robot control, we developed a technique based on a discriminant color feature with robustness capabilities with respect to intensity variations and specularities. We also designed an adaptive region growing with automatic region seed detection and a model-based region classification, both dedicated to laparoscopy. The foreseen application is a good training ground to evaluate the proposed technique and the effectiveness of this work has been demonstrated through experimental results with endoscopic image sequences to efficiently locate boundaries of a landmark-free needle-holder at half the video-rate.

1. Background

One can observe in the last few years the increase of computer vision applications in assisted surgery, in particular to intra-operative guidance procedures. One of these techniques is the minimally invasive surgery like the laparoscopy. In laparoscopic surgery, small incisions are made in the human abdomen to introduce surgical instruments as well as a laparoscopic optical lens (a laparo-scope) connected to a camera, through trocars. Looking at a monitor device, the surgeon moves the instruments in order to perform the desired surgical task [1]. Minimally invasive surgery is a very attractive technique since it avoids surgical opening and it reduces the recovery time for the patient but, in counterpart, a large number of repetitive gestures such as the cleaning-suction process, clamping, cauterization, needle manipulation are needed and requires much training from the surgeon. Moreover, the main drawback of this surgical technique is the posture of the surgeon which is very tiring. In the last few years (less than a decade), robotized laparoscopic systems have appeared [2] and are designed to reduce the surgeon's tiredness and to increase the accuracy (like the Zeus medical robot from Computer Motion Inc. [3] or the Da Vinci system from the Intuitive Surgical Inc. [4]). With these systems, robot arms are used to manipulate the surgical instruments as well as the laparoscope. However, motions of instruments are constrained to by the insertion point in the abdominal wall which reduces the mobility since only four degrees of freedom are available.[1] Our research in this field aims at expanding the potentialities of such

*Corresponding author. Tel.: +33 390244471; fax: +33 390244480.

E-mail addresses: christophe.doignon@ensps.u-strasbg.fr (C. Doignon), pierre.graebling@ensps.u-strasbg.fr (P. Graebling), demath@eavr.u-strasbg.fr (M. de Mathelin).

[1]In minimally invasive surgery, extensive range of articulations and referred to as endowrist have been recently designed at the tip of some laparoscopic instruments.

systems by means of computer vision techniques to realize semi-autonomous tasks.

With the help of a monocular vision system, there exist several obstacles to achieve an accurate positioning of laparoscopic instruments in the abdominal cavity. The first one is the unknown relative position and orientation of the robot arm holding the instrument with respect to the camera frame [5]. Other hard problems are mostly emanating from the image segmentation and data interpretation with monochrome camera. The aim of image segmentation is to divide an image into parts that have a strong correlation with objects or areas of the real world contained in the image. In laparoscopy, difficulties are mainly due to the complexity of the observed scene, the time-varying lighting conditions, the moving background (due to breathing and heart beating) and the presence of many specularities and bloodstained parts onto the instruments surface. To overcome some of these harmful effects, prior researches involving color images to this field have been conducted and visual servoing techniques have been already applied to the laparoscopic surgery. Casals et al. [6] employed patterned marks on the instrument mounted onto an industrial robot to realize an instrument tracking task. Projections of marks in the image plane were approximated by straight lines in the image segmentation process. This guidance system worked at a sampling rate of 5 Hz with the aid of an assistant. Hirzinger et al. [7] used a color stereovision system to realize a tracking task with the endoscope mounted on a robotic arm. With a judicious and simple histogram analysis, they selected the color that did not appear to artificially mark the instrument with that color. This spectral mark was then used to control the robot motion at a sampling rate of 17 Hz. An interesting feature in this technique is the choice of the HSI color space representation to bring to the color segmentation, leading to a good robustness with respect to lighting variations.

In order to assist the surgeon, we have conceived two years ago an autonomous system that automatically brings the instrument at the center of the endo-scopic image. It included a video-rate visual servoing system and the design of a special device to hold the surgical instrument with tiny laser pointers and optical markers. The laser pointer projected four spots onto the organ surface which in turn were captured by a camera. Optical markers (composed of three circular LEDs) placed at the tip of the instrument were directly projected onto the image and in conjunction with images of the laser pattern, they were used to recover the depth between the pointed organ and the instrument [5].

One can notice that all previous related works require special markers, color images and are dedicated to navigation of laparoscopic instruments inside the abdominal cavity. To help the surgeon, more ambitious tasks must be investigated. New proposed tasks may require interactions with tissues and more autonomy. For example, autonomous needle manipulation is one such hard problem for which we wish to contribute through some subtasks like fast instrument recognition and localization so as to enhance capabilities of vision for medical robots. Furthermore, since surgical instruments must be autoclavable for the surgical operation and since several instruments may alternatively be used through a trocar (depending on the subtask addressed), it seems irrealistic to always keep some markers placed on the surgical instrument. The objective of this paper is to provide a robust segmentation of laparoscopic instruments without additional landmarks and as fast as possible in order to be integrated as a module of an image guidance system. The development of a reliable segmentation for color images as part of a vision-driven endoscopic system is still a challenging task as mentioned by Ascari et al. in [8] which have presented an autonomous color segmentation for subarachnoid spinal images through endoscopes. It was part of an obstacle-avoidance module for a vision-driven spinal endoscopic system. Based on the luminance signal for the lumen detection and chromaticity components (hue and saturation) for the nerves and vessels recognition inside the lumen, image segmentation was performed by means of a multi-step approach with adaptive processings and histogram analysis. In a very recent work, Burschka et al. [9] have presented several useful image processings (from segmentation to registration) fully dedicated to laparoscopy. In their approach they developed a set of techniques (some of them are adaptive algorithms) for tracking deforming surfaces from stereo streams whereas we use image sequences of a single-color camera. To conclude with related works, it emerges that the ability of the whole segmentation to be adaptive is a key feature for success. Adaptive algorithms may be computationally expensive, thus one must deal as well with low and high resolutions but also with efficient implementations. In laparoscopic surgery, most of surgical instruments are metallic leading to the presence of gray regions but also high surface reflectance in the image. To deal with these phenomena, we propose a segmentation scheme which first consists in some improvements of robust existing algorithms applied to a color-based region segmentation, second consists in the design of an automatic region seed detection followed by a region growing with adaptive-threshold, and third we developed a model-based regions classification dedicated to images through laparoscopes.

The remainder of the paper is organized as follows. In Section 2, low-level processings emphasizing the color purity component are detained. This section also includes an efficient region smoothing algorithm and introduces the color feature used. The region-based

color segmentation is developed in Section 3 and we describe in this section the automatic region seed detection together with the region growing. Throughout this paper, experimental results are presented with color images of laparoscopic surgical instruments in presence of living tissues. Section 4 presents the region classification based on region-boundary pixels. In the same section, results concerning the region shape description with moments are applied to the on-line estimation of the insertion point localization. In Section 5, we give computation times of each stage of this fast color segmentation and we conclude the paper in Section 6.

2. Segmentation using color features

2.1. Introduction

A fundamental requirement of reliable vision systems is the ability to extract from digital images visual cues relevant to the imaged scene. To handle the big amount of data contained in a video information, segmentation as well as classification/recognition and parameterization are some important steps which produce a data representation in structural form. Most of earlier works in the field of data classification involved techniques mainly based on Markov Random Fields (MRF-based energy minimization), multiresolution scheme, statistics or Genetic Algorithms (GA) which are not yet compatible with a real-time video control [10]. For applications involving robots, image segmentation as well as classification and recognition must also be fully automated and when one has to deal with color images, it is suitable to take advantage of the multispectral contents of the video information. With color images, a pixel is a mixture of the three fundamental colors: red, green and blue (RGB) and there exist many representations to encode this information. The RGB space [11] may be the most well-known color representation since it is useful for data storage. However, some color image processings such as enhancement and restoration require that only the luminance component (as the amount of visible light) to be processed whereas some other applications require color (hue and saturation) components to be preserved or modified [12,13]. It is known that the human eye can detect only in the neighborhood of one or two dozen intensity levels at any point in a complex image due to brightness adaptation, but it can differentiate thousands of color shades and intensities [14,15]. Following this purpose, the color saturation seems to be a discriminant attribute for gray regions segmentation since it is a measure of the amount of white within the color despite that it may be affected by surface reflectance [16]. A low saturation value indicates a low colored pixel and a high value corresponds to a purely colored pixel. Coordinate
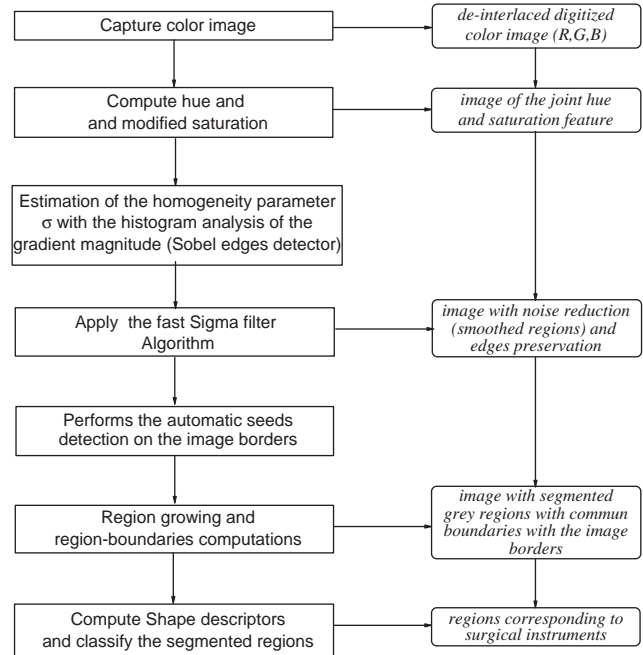

Fig. 1. Flowchart of the proposed segmentation.

systems related to the psychological perceptual attributes (Hue, Saturation and Intensity—HSI for short) are more adequat for analyzing colors distribution in the image than RGB since the chromaticity plane (H and S) is perpendicular to the intensity axis and furthermore, RGB space brings a non-uniform chromaticity scale [17]. This is inappropriate when most segmentation techniques need a similarity measure to discriminate two colors [13]. In the chromaticity plane, hue and saturation are, respectively, related to the orientation and magnitude. Transformations to other perceptually based spaces such as CIE Lab and CIE Luv need much computation time and do not provide discriminating cues more significant. In [18], the color saturation is related with RGB by

$$S = 1 - 3\,\frac{\min\{R, G, B\}}{R + G + B}. \qquad (1)$$

This definition clearly shows that pixels may have the same saturation whatever are their brightness or color hue values (excepted for the singularity located at $R = G = B = 0$). There exist other definitions for the saturation signal like the radius of the chromaticity circle perpendicular to the luminance axis, as it is with YIQ color coordinates system (Fig. 1).

2.2. Saturation enhancement

With the objective of improving the detection of gray regions in the image, we intend to look for a little bit more discriminant visual cue in this subsection in order

to better classify chromatic pixels from achromatic ones. With the purpose of highlighting this aspect, we propose to define the color purity with a slightly modification of the above definition for saturation as follows:

$$S' = 1 - \frac{\min\{R, G, B\}}{\max\{R, G, B\}}. \qquad (2)$$

This can be thought as a color purity stretching with a simple non-linear point operation. Compared to the saturation attribute defined in (1), this new color purity attribute is a little bit more sensitive to brightness changes but mainly for chromatic pixels. In counterpart, S' rather affects more high values than low values which tends to separate more chromatic pixels from achromatic ones. For instance, with $(R, G, B) = (3, 6, 3)$, the saturation $S = \frac{1}{4}$ whereas $S' = \frac{1}{2}$ for this rather green pixel. The original saturation value is identical to the one computed with some much more gray pixels, such as with $(R, G, B) = (3, 4, 5)$ but not S' which has a lower value ($S' = \frac{2}{5}$). The effect of this new attribute is illustrated in Fig. 2 in which one can see that the absolute difference $|S' - S|$ is more significant for high values than for low values. To display the bottom left image in Fig. 2, a color transformation from RGB to HSI coordinate system is carried out and S' is computed with Eq. (2). Then, the transformation from HSI to RGB is computed with S' instead of S.

In addition to a low saturation value, pixels can equally be categorized as achromatic if the brightness is very low or very high (for intensity values lower than 10% or greater than 90% of full scale as suggested by Ikonomakis [15]). So, pixels that fall into this category

are labeled as "meaningless". The region of interest is assumed to be gray and will be referred to as the foreground. It could not be assumed to be the only gray region in the whole image since other parts of image can also contained many pixels with a gray distribution. This is mainly due to the presence of specular reflections (near the image center in Fig. 2(a)) or hue discontinuities, both are accounting for local disturbances in the detection of regions of interest (as it is for color transitions occurring at the needle boundaries) leading to high brightness values and low color purity values (see Fig. 5). At specular areas, pixels can be "repaired" by means of an iterative extraction/interpolation process of the local image structure [19]. However, the computational cost of existing algorithms are too high to look to a real-time implementation. Moreover, specularities may appear in the background as well as in the imaged part of a surgical instrument. So, we chose another way to deal with this problem as it will be explained in the next section.

2.3. Fast and shape-preserved edges filtering

Noise filtering is commonly used as one of the first operations applied to digitized images. Non-linear filtering allows to detect lack of spatial coherence and either replace an inconsistent pixel value by using some or all pixels in a neighborhood. Such low-level processing is crucial to keep away an over-segmentation result which is very awkward for pixels classification. Some of non-linear filters have the capabilities to smooth intensity values of pixels in a given region and to
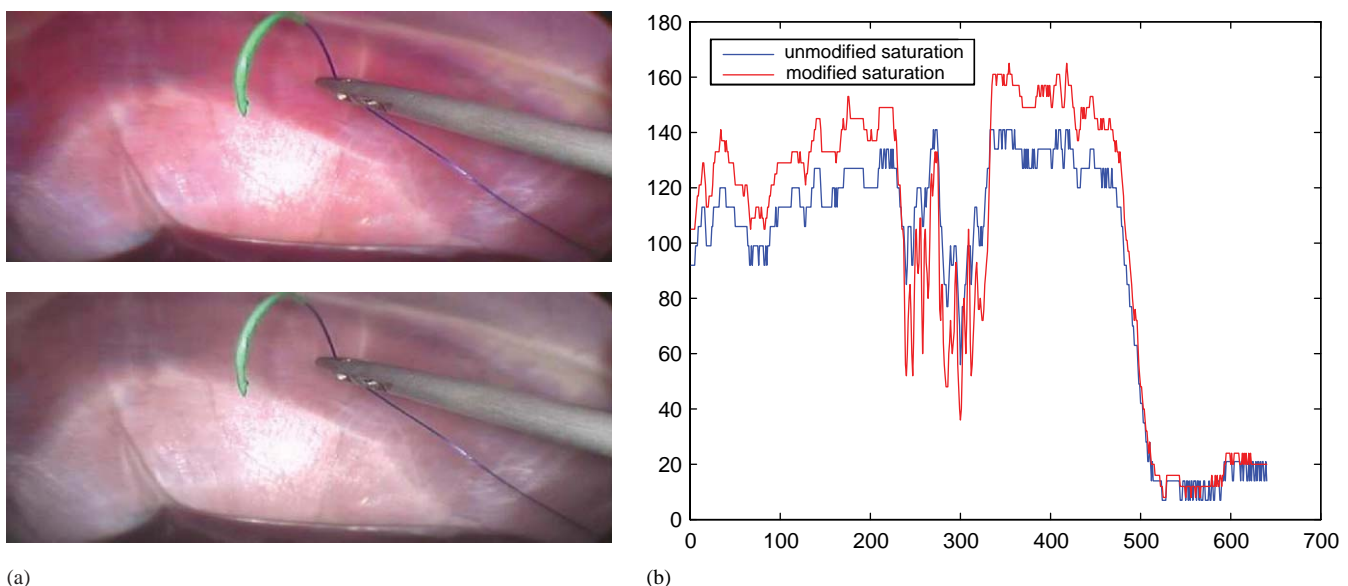

(a) (b)

Fig. 2. Images of laparoscopic instruments (needle-holder and the needle). (a—top) original color image. (a—bottom) color image with the modified saturation attribute S'. (b) Comparison between saturation S (blue) and modified saturation S' (red) for pixels on the horizontal line at middle height of the top left image. Color purity value is enhanced for chromatic pixels whereas it is quasi-unchanged for nearly achromatic ones.

equally preserve the topological properties of edges. The homogeneity plays a significant role in separating the objects from each other, usually in separating the region of interest from the background and a very attractive color segmentation based on homogeneity histogram has been presented by Cheng et al. [20] and a comparative study of performances for the mostly used non-linear filters is given by Seeman [21] or Pitas [22] for instance. For instance, the apply of anisotropic diffusion to computer vision received a great attention, since this technique encourages smoothing within a region whereas region boundaries remain sharp [23]. However, although a geometry-driven approach is elegant and powerful, it suffers from a cumbersome computing time, and in practice it is not (yet) suitable as part of a video-rate tracking scheme. The so-called sigma filter proposed by Lee [24] is a good computational tradeoff. With the sigma filter, a pixel can be averaged with its neighbors that are close in value. Lee suggests looking to all values in the neighborhood of a given pixel (u,v) and averaging $f(u,v)$ only with those values that are within the two-sigma (σ) interval of f. If $N(u,v)$ is the $(2w+1) \times (2w+1)$ neighborhood size around the pixel (u,v), then the estimate $\tilde{f}$ is computed with

$$\tilde{f}(u,v) = \begin{cases} \dfrac{1}{n_c} \sum_{(m,n) \in N_c} f(u,v) & \text{if } n_c \geqslant \alpha \text{ card}\{N\}, \\ f(u,v) & \text{otherwise,} \end{cases} \qquad (3)$$

where $N_c = \{(m,n) \in N(u,v) : |f(u,v) - f(m,n)| \leqslant 2\sigma\}$ and $n_c = \text{card}\{N_c\}$. If less than $\alpha = \frac{9}{100}$ of the pixels (also suggested by Lee) in $N(u,v)$ are close in value to the pixel (u,v), the pixel value is left unchanged; such pixel is presumably a region-boundary pixel. This is illustrated

in Fig. 3 with endoscopic images of Fig. 2(a). Compared to the well-known median filter, the sigma filter is more efficient for smoothing regions since the median filter is much more dedicated to peak noise cleaning rather than gaussian noise cleaning.

The sigma filter is a much more computationally intensive algorithm than the well-known implementation of the median filter, based on a local histogram, due to Huang et al. [25], since the latter one is as fast as separable linear filters to compute. However, to obtain similar smoothing effect as with the sigma filter, the median filter must be used iteratively. One may deal directly with the color image to perform the sigma filtering since various multichannel modifications of the sigma filtering concept have been proposed for color image processing [26,27]. However, it requires too many computations in regard to real-time capabilities with PCs. Here, we propose an approximation of the sigma filter which may be close in performances to the iterative median filter thanks to an efficient implementation of the filtering. Since we have no idea about appropriate value for the contrast parameter σ, a method for estimating its value should be useful to get a noise reduction and segmentation with low loss of details [28]. In a similar manner to Perona and Malik in a classical scale-space analysis of the diffusion process [23], we used the so-called noise estimator described by Canny [29] but also revisited and compared with a morphological method in [30]. According to this method, the histogram of the absolute values of the gradient throughout the image is computed. Then the 90% of value of its cumulative sum at each integration step is taken as a noise estimator. This is illustrated in Fig. 4(a and b).
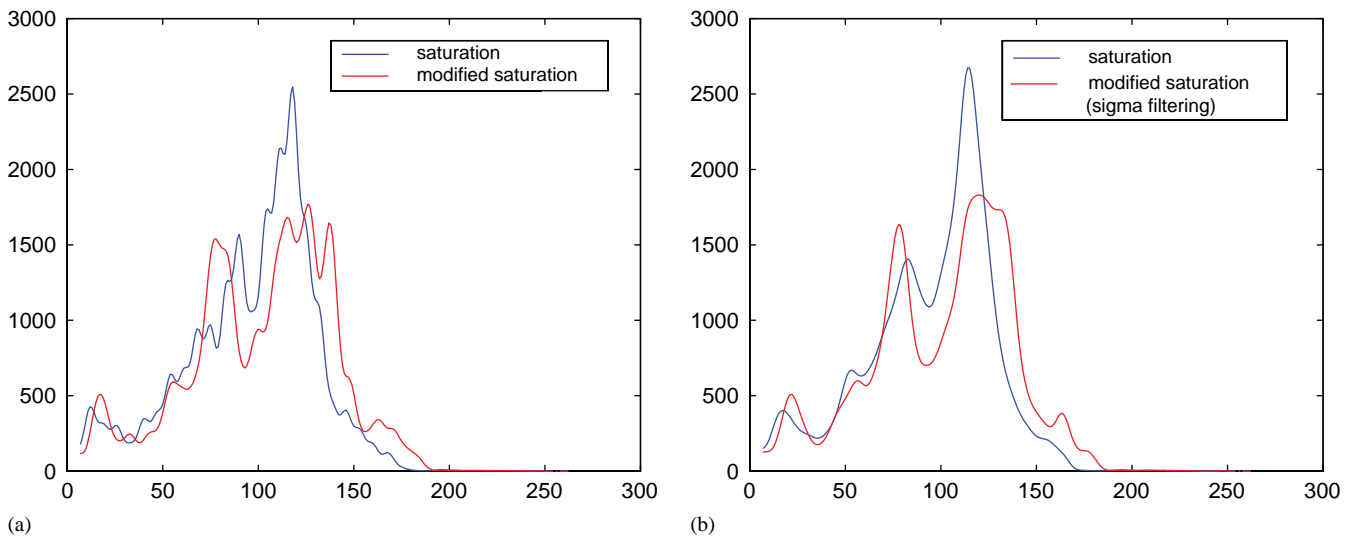

Fig. 3. (a) Histograms of the color saturation feature S (blue) and the modified saturation feature S' (red) all over the input image. (b) idem but once the sigma filtering has been proceed. The homogenization effect within a class of pixels and the stretch of the histogram towards the high saturation values are both significant.
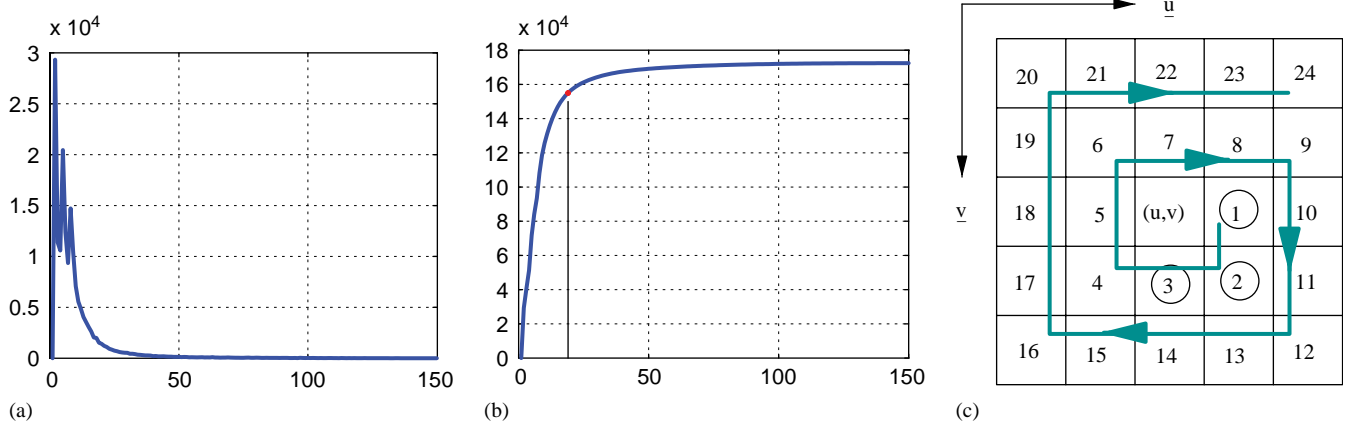
Fig. 4. (a) Histogram of the gradient magnitude of saturation image for image in Fig. 2(a). (b) The histogram of cumulative sum. The value chosen for σ is the abscissa coordinate corresponding to 90% of the maximum. (c) The "evolute-shape" run coding to speed-up the sigma filtering. Pixels denoted with ①, ② and ③ are also set to $\tilde{f}(u,v)$ if all the pixels labeled with "1" up to "15" in the neighborhood have values within the two-sigma interval.

To reduce the computations with the sigma filter, we have assumed that there are many more pixels inside a region than region-boundary pixels. This means that there are a majority of pixels (u,v) for which the quantity $n_c(u,v)$ is greater than $(\alpha \, \text{card}\{N\})$, thus there are many more pixels to smooth than unchanged pixels. This reasonable assumption guides us to present the following scheme which is an approximation of the sigma filter. The neighborhood $N(u,v)$ with $\text{card}\{N\} = (2w + 1)^2$ pixels is not entirely looked into. The counter n_c is incremented if a pixel value is within the two-sigma interval. Once n_c is greater than a fraction $\beta(\beta > \alpha)$ of $\text{card}\{N\}$, the pixel value $f(u,v)$ is averaged with these neighbors whatever are the others $((1 - \beta) \times \text{card}\{N\})$ pixels inside the neighborhood. Then, the computations cost strongly depends on the value for β. To get an accurate smoothing, it must be chosen greater than α, and the smoothing must be done with the closest neighbors preferably to pixels near the neighborhood boundaries (typically, β is 45% (if $w > 2$ otherwise $\beta = \frac{15}{25}$ for $w = 2$) and α is 9%). Moreover, the (unvisited) pixels $(u+1,v)$, $(u+1,v+1)$, $(u,v+1)$ are often strongly correlated with the pixel (u,v). To do so, an "evolute-shape" run coding centered to the current pixel (u,v) is automatically built to carry out the algorithm either to smooth the central pixel with the (nearly) closest pixels and either to set the three pixels $(u+1,v)$, $(u+1,v+l)$, $(u,v+1)$ to the value $\tilde{f}(u,v)$ if the 15 first neighbors of the run are all in the two-sigma interval. This very simple algorithm is illustrated in Fig. 4(c) for $w = 2$ and applied to the image in Fig. 2(a).

Quantitatively, the performance in noise reduction and edge preservation may be measured as the ratio of the number of region-pixel Nr and the total number of pixels NP. Depending on the value for parameter β and

the size w of the neighborhood, this ratio can be seen as a degree of over-segmentation (see Fig. 5(a)). A comparison of the apply of anisotropic diffusion and the sigma filter to the image of color purity is shown in Fig. 5(b and c). One can observe that homogenization and edges structure preservation are very similar whereas the computing time is greatly reduced with the proposed implementation of the sigma filter. However, it is still too much time consuming to achieve the video-rate. Then, an image pyramid is built [31] to reduce the computations.

2.4. A quasi-invariant photometric color feature

A significant improvement in color understanding is the dichromatic reflection model due to Shafer 20 years ago. This model separates the reflected light into body reflection (object color) and surface reflection (specularities). The connection between differential-based features and photometric invariance theory has been proposed by Geusebroek et al., and recently, van de Weijer et al. [32] have presented a new class of features also derived from the dichromatic reflection model which do not have the inherent instabilities of the previous ones. A very important property shared by all the proposed features (also called shadow-shading specular quasi-invariants) is their perpendicularity to the shadow-shading direction and the specular direction. This kind of features is very suitable for an edge-based color segmentation since the authors have shown that it introduces a less edge displacement than many other features. Keeping in mind a region-based segmentation for the current work, we have considered the following quantity:

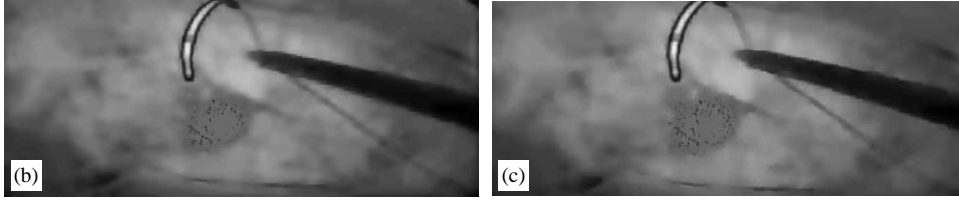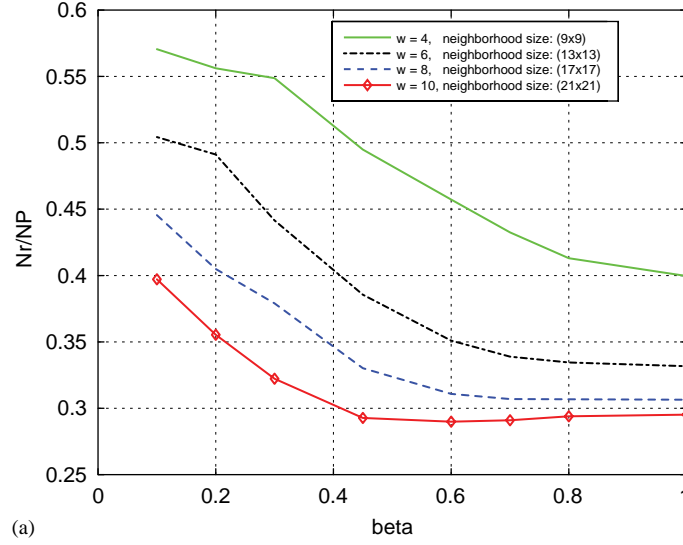$$\mathcal{H} = S'H, \tag{4}$$

(a)

(b)

(c)

Fig. 5. (a) Influence of the neighborhood size w and parameter β on the over-segmentation. (b) Results with the anisotropic diffusion process (an implementation due to Perona et al.) applied to the color purity feature S' for image in Fig. 2(a) (30 iterations, $K = 10$, $\lambda = 0.25$, computational time is about 2 s). (c) Results with the efficient implementation of (an approximation of) the sigma filtering ($\sigma = 8$, $w = 8$, $\beta = 45\%$, computational time is 60 ms for the (640×240) image).

where H is the color hue feature

$$H = \arccos\left(\frac{\frac{1}{2}[(R - G) + (R - B)]}{\sqrt{(R - G)^2 + (R - B)(G - B)}}\right). \quad (5)$$

and from which the first derivative is closely related to the shadow-shading-specular quasi-invariant $|H_x^c| = S \cdot H_x$. H_x denotes the spatial differentiation of the hue H and a change of $\mathcal{H}$ may also occur with a change of the color purity S' [32]. A well-known drawback of hue is its undefinedness for achromatic pixels, i.e., for small S' and small changes round the gray axis result in large changes of the direction of that quasi-invariant and therefore the derivative of hue is unbounded. However, van de Weijer et al. have shown that the norm of H_x^c remains bounded. It follows that its integral is also bounded, and hence, $\mathcal{H}$ is bounded (Fig. 6).

3. Region-based segmentation

With the resulting smoothed image of the color hue saturation feature $\mathcal{H}$, a region-based segmentation technique is carried out to provide closed region boundaries. We select one of the existing seeded region growing techniques [33] and taken advantages of the current context, initial seeds are automatically detected and located with local statistical characteristics. Hence, an histogram-based adaptive thresholding for region growing is presented.

3.1. Region seed detection

In this part, we present the automatic detection of seeds to initiate the region growing process. It is fully dedicated to endoscopic images in laparoscopy, since projections of laparoscopes in the image are known to always intersect image borders (any laparoscope is constrained to pass through the incision point). As a consequence, the signal $\mathcal{H}$ is analyzed only at image borders, that is inside a crown of few pixels of width. If a pixel with a $\mathcal{H}$ value smaller than a threshold $\tau_{\mathcal{H}}$ is found, the process is looking for consecutive pixels so long as the $\mathcal{H}$ value is below this threshold and thus detecting the local minimum. Its position is chosen as the seed location and the mean value computed with the set of pixels below the threshold is taken as the seed value (if the number of such achromatic pixels is significant with respect to the image size). The process is repeated all along the image borders and provides several putative seeds for region growing. Two problems
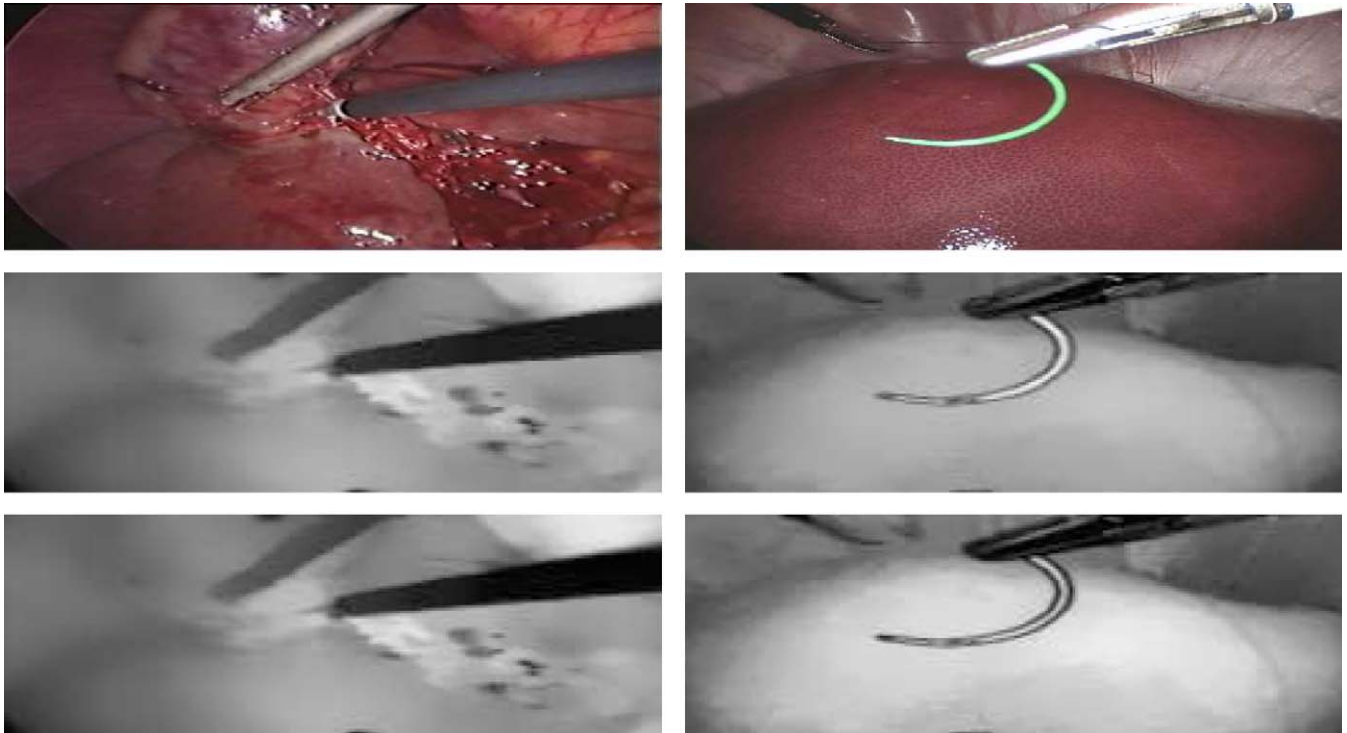
Fig. 6. The discriminance of the color feature $\mathscr{H}$ and its robustness with respect to specularities and changes in intensity. (top) Images picked up from a video sequence. (middle) The apply of the sigma filter to the image of the modified saturation S'. (bottom) The apply of the sigma filter to the color feature $\mathscr{H}$. Differences between the left middle and left bottom images are not very significant for the strongly red-colored background of the top image with few specularities whereas differences are noticeable on right images with many specularities onto the instrument surface and with a less colored background.

are encountered with this method. First, an unexpected low value for $\mathscr{H}$ may occur at the vicinity of image borders in presence of high reflectance or low illumination and leads to the detection of a "false" seed. This problem will be solved thanks to the region classification. The second problem is the determination of the threshold $\tau_{\mathscr{H}}$. With a too low value, may be no seed is found and the process fails in the detection of instruments. With a too high value, many seeds are detected, leading to expand many regions. To tackle this, the mean $\mu_{\mathscr{H}}$ and the standard deviation $\sigma_{\mathscr{H}}$ of the set of $\mathscr{H}$ values are computed all along the crown and $(\mu_{\mathscr{H}} - \sigma_{\mathscr{H}})$ is taken as the threshold $\tau_{\mathscr{H}}$ and provides satisfactory results (see Fig. 9 (right)).

3.2. Region growing and boundaries computation

Given the set of detected seeds $S_1, \ldots, S_q$, each step of a recursive region growing incorporates one additional pixel with a four-connected neighborhood operator. At each step, the position of the unallocated pixel (u,v) replaces the initial seed and preceding allocated pixel (x,y) to generate homogeneous regions $R_1, \ldots, R_q$. To do so, a criterion of homogeneity must be satisfied and the pixel (u,v) is said to belong to the same class as one of its neighbors. The choice of the homogeneity criterion

is critical for even moderate success of region growing. Here, a threshold σ for the fast and shape-preserved edges filtering has been previously estimated. The parameter σ can be thought as a homogeneity threshold and the criterion used is the absolute difference in the joint hue saturation values

$$|\mathscr{H}(u,v) - \mathscr{H}(x,y)| < \sigma. \qquad (6)$$

Moreover, given the seed location, an histogram of a subimage with the color hue saturation attribute $\mathscr{H}$ is computed. It is a means of estimating the most appropriate threshold t for the region growing starting with that seed. The size of the subimage is a quarter of the image size and is defined as illustrated in Fig. 7(a). This way, the produced histogram shows two dominant modes, once it is smoothed with a gaussian (see Fig. 7(b and c)). Then, among the numerous existing histogram-based methods for thresholding, we employ the Otsu's approach which minimizes the within-group variance of this two-class clustering problem [34]. To be more precise, we used the efficient implementation due to Reddi et al. [35] and in addition to the homogeneity criterion defined by Eq. (6), the unallocated pixel (u,v) must also satisfy

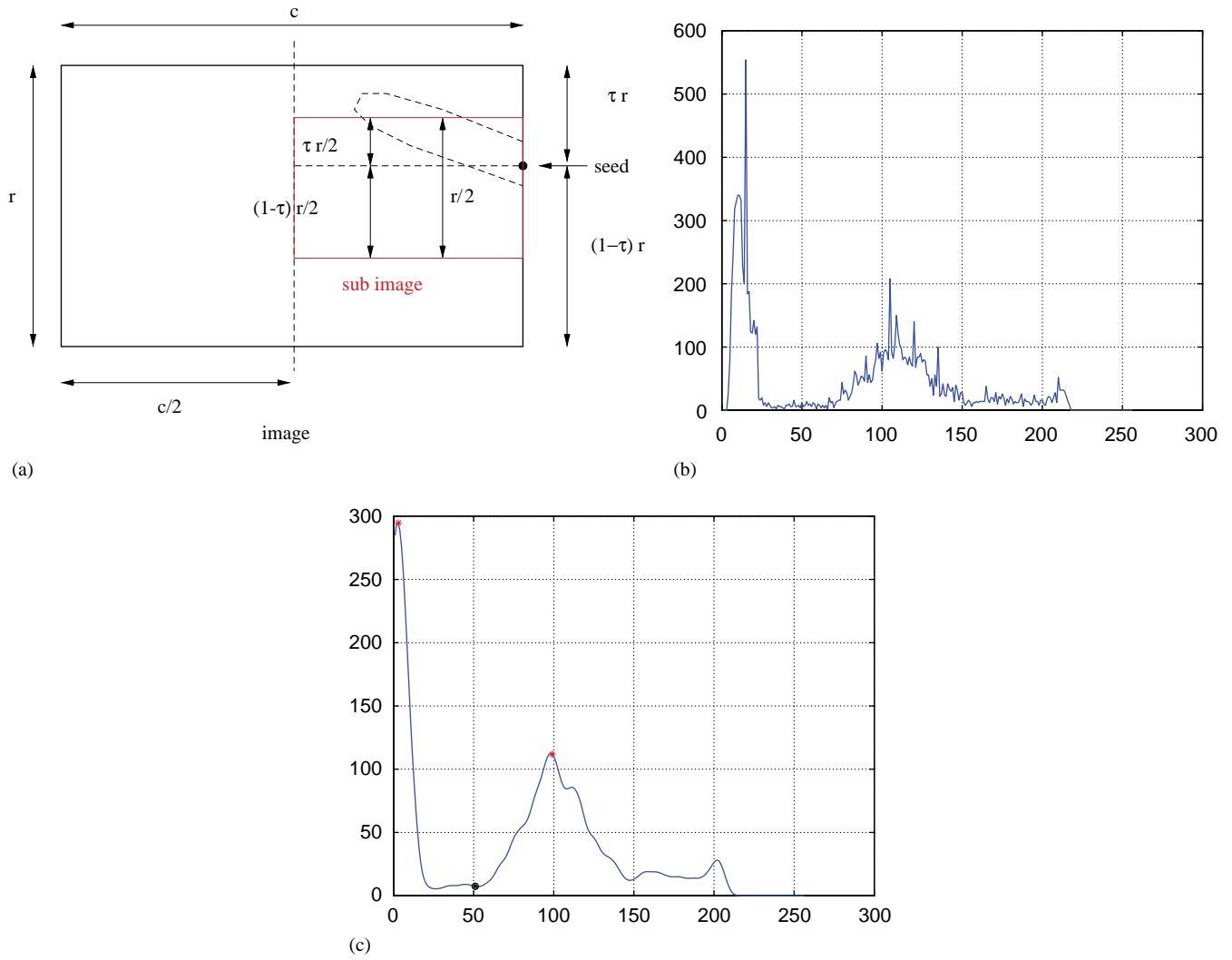$$|\mathscr{H}(u,v)| < t. \qquad (7)$$

(a)

(b)

(c)

Fig. 7. Estimation of the threshold t for any region to grow. (a) The definition of the subimage used for computing the histogram, (b) its histogram and (c) smoothed *bi-modal* histogram and the threshold found ($t = 51$) with the Otsu's method.
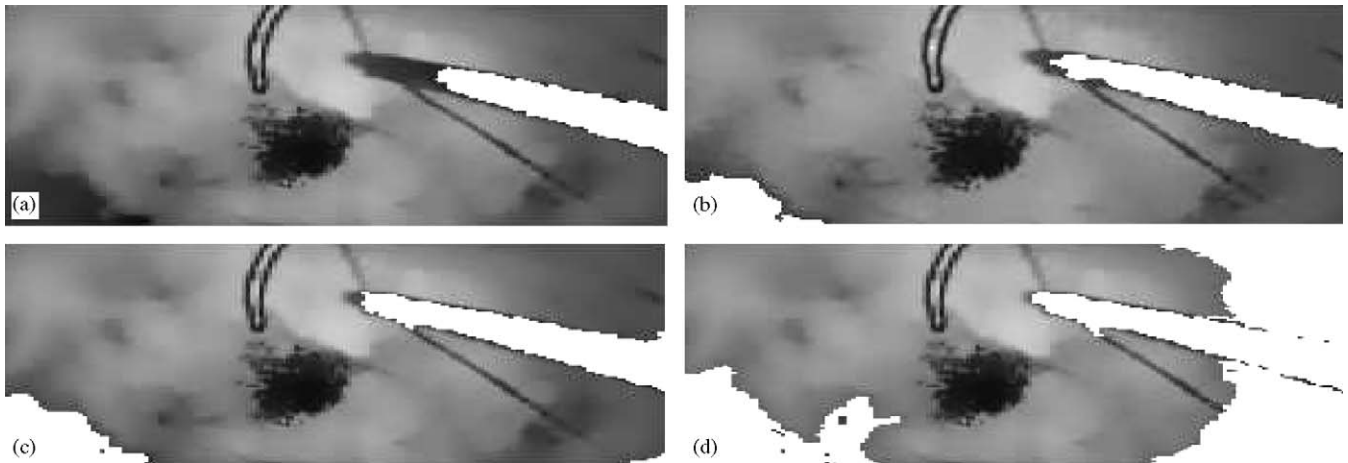

Fig. 8. Influence of the threshold value t on the region growing results with image of the joint hue saturation (filtered) issued from the left bottom image in Fig. 2(a). (a) $t = 35$. (b) $t = 51$ (Otsu's threshold). (c) $t = 72$ and (d) $t = 120$.

Fig. 8 shows the influence of that threshold value on the region growing results. With a too low threshold value, many region pixels are missing, whereas with a too high threshold value, several regions have been merged, yielding to fail the forthcoming region recognition stage. Pixels in the same region are labeled with the same
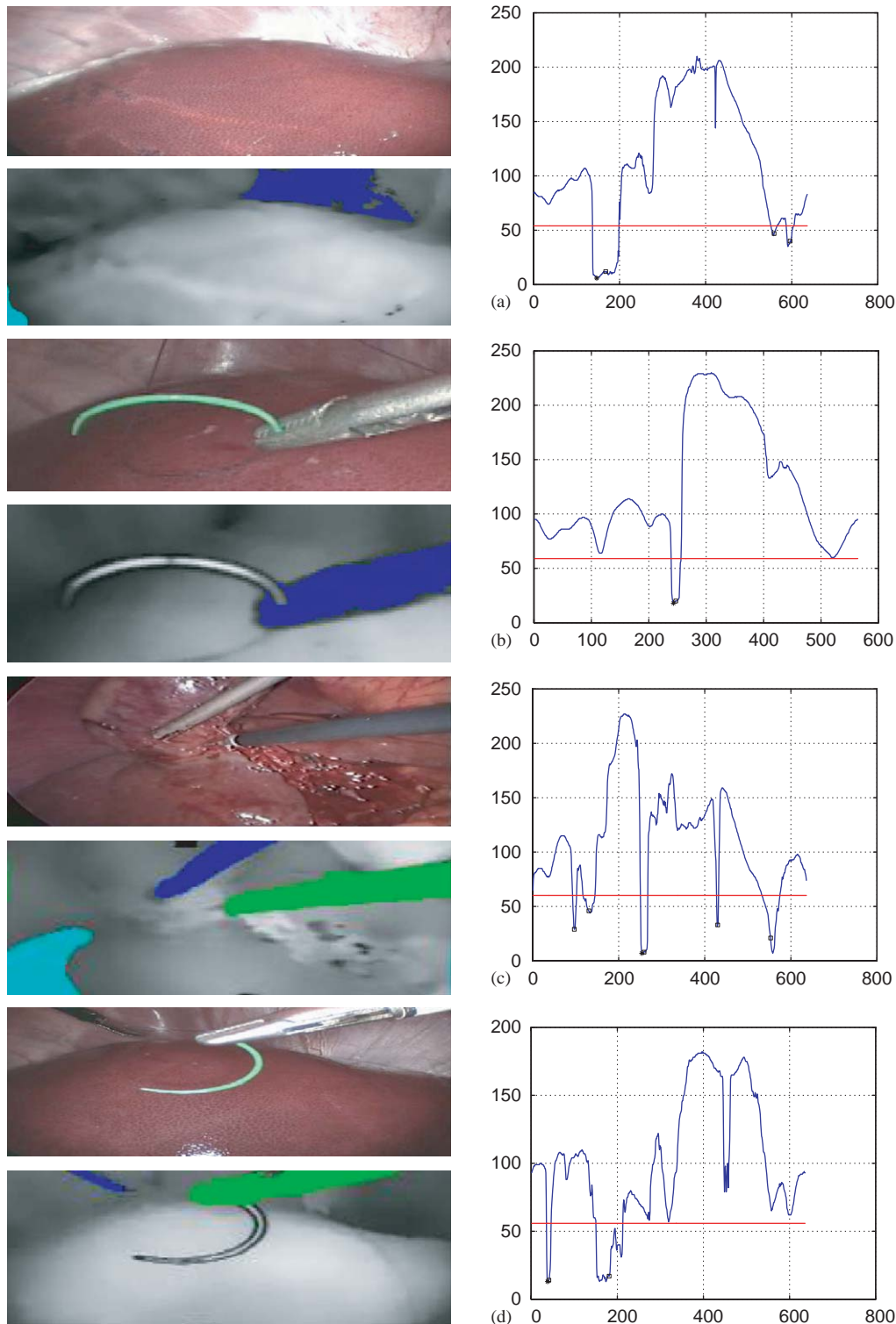

Fig. 9. Four examples ((a)–(d)) of the region-based segmentation. (left) Regions have been grown (some have been merged or removed) and those corresponding to the image of surgical instruments are always detected. (right) Color feature $\mathcal{H}$ all along the image borders (the top-left image corner is taken as the origin and the run is clockwise). The horizontal straight line is corresponding to the threshold $\tau_{\mathcal{H}}$.
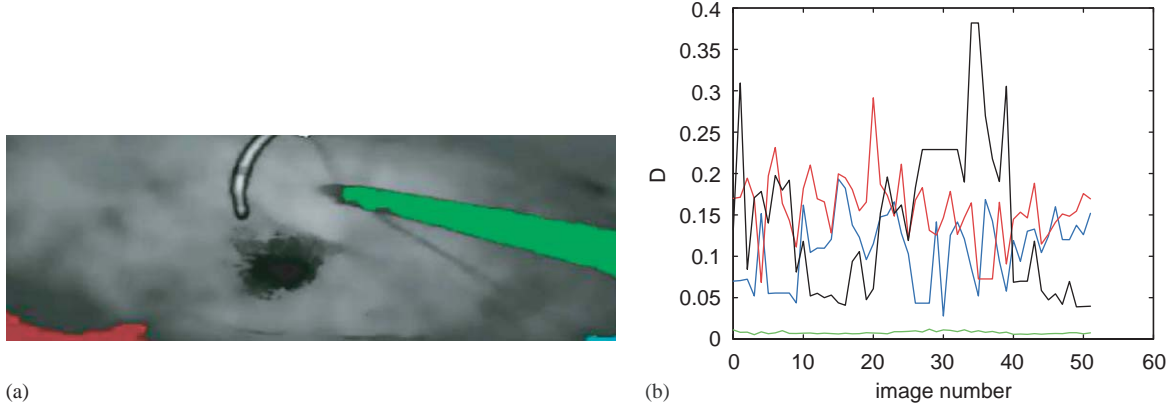
Fig. 10. Region classification with the similarity measure D_f based on 30 normalized Fourier descriptors for a video sequence of 52 images. (a) Segmentation of a sample image extracted from the video sequence. The green region is that of corresponding to the instrument. (b) The most similar region to the model of laparoscopic instrument is that of corresponding to the smallest D_f value (green curve).

symbol whereas pixels in distinct regions are labeled by different symbols. Hence, the classical Papert's "turtle" algorithm [18] is performed to locate the boundaries using a dynamic thresholding. Fig. 9 (left) illustrates such segmentation with four examples: (a) no instrument in the image, (b) a needle-holder holding a needle, (c) two instruments with different mean gray values, (d) two instruments with specularities. Among all segmented regions, those corresponding to the surgical instruments have always been selected.

4. Region classification

To recognize the region corresponding to the laparoscopic instrument, each region has to be characterized by a set of compact descriptors which is invariant to geometrical variations such as scaling, rotation, translation, etc. Then, given a simple model of the instrument, we characterize image regions with a compact representation and we are looking for a similarity measure to match image regions with the object model. For this purpose, two approaches have been investigated, the first one is using the Fourier descriptors and the second one is dealing with moments.

4.1. Robust classification with region boundaries

In order to extract a finite set of numerical features characterizing a region, we can use its boundary and apply a Fourier parameterization [36,37]. For that purpose, we consider the (u,v) coordinates of each point belonging to a boundary as a complex number $u + iv$. The closed contour can be represented with a periodic function of the arc length $l : f(l) = u(l) + iv(l)$ and thereby expanded in a Fourier series. After an appropriate normalization step, we get a set $\{a_k^q\}$ of normalized Fourier descriptors (NFDs) representing

conveniently a compact feature set that characterizes a region R_q. Given the reference set $\{a_k^o\}$ of NFDs accounting for the instrument model (a shape description for a gray rectangle), a similarity measure is evaluated (with a finite number of terms) for each region R_q according to

$$D_f^q = \sum_{k=1}^{n_f} |a_k^q - a_k^o|^2. \tag{8}$$

A minimum value for D_f^q indicates the most similar region shape to the model of the instrument. This technique has been evaluated on endoscopic image sequences. Fig. 10 illustrates the robustness of the proposed technique with $n_f = 30$ terms in the decomposition with the NFD. One can observe that the instrument (green curve) is well detected all along the sequence despite the presence of regions corresponding to highlights or other artefacts.

4.2. $\mathcal{H}$-based region classification with moments

In this part, we are looking for region descriptors based on image moments. Denoting by Ω_{in} the inside region of a shape, moments are statistical characteristics which are able to encode both the shape and the gray-level distribution inside the regions. It is a very well-known description of shape and intensity distribution $G(u,v)$. The moment of order $(i+j)$ is defined by

$$m_{ij} = \int \int_{\Omega_{in}} u^i v^j G(u,v) \, du \, dv. \tag{9}$$

It is used for a long time in shape recognition and visual inspection, since it is very easy to compute and robust w.r.t. noise but it is not reliable in presence of occlusion or geometric transformation. However, there exist seven moments invariant to translation, rotation and scale also known as Hu's moment and there exists a set of
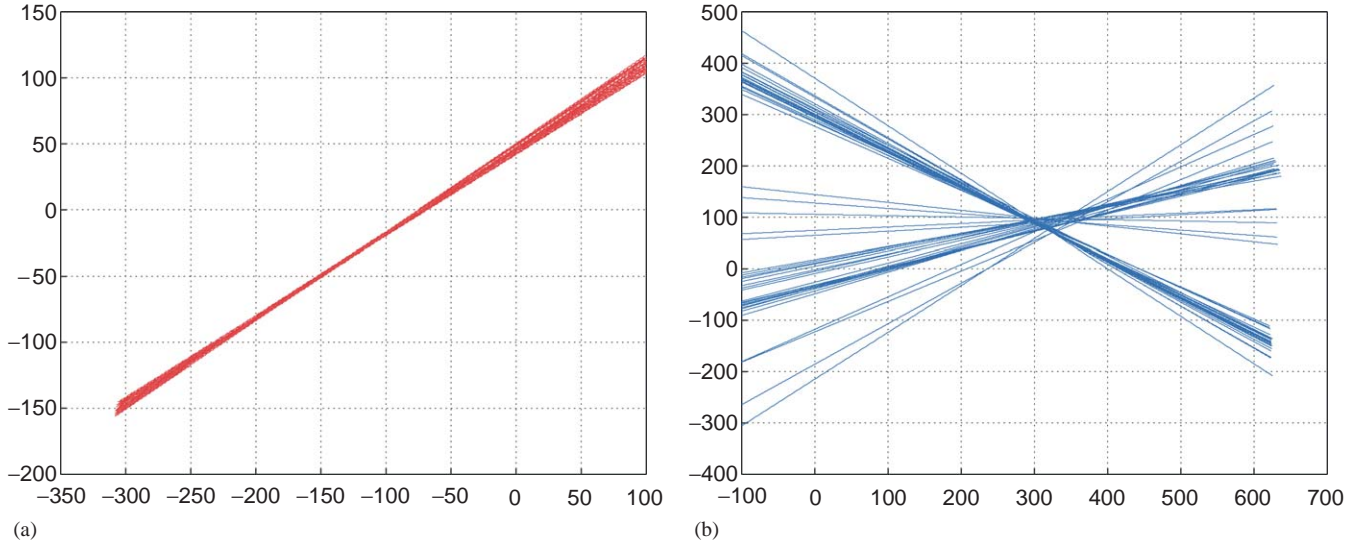
Fig. 11. Localization of the projection of the incision point onto the image plane by means of the central moments of selected image regions corresponding to the two instruments and computed all along an image sequence (a sequence of 52 images, from which images in Fig. 9(d) are picked up). (a) Intersections of red straight lines are corresponding to the trocar localization of the left instrument and is quite motionless, while (b) intersections of blue lines are corresponding to the trocar localization of the instrument, which is moving during the sequence.

four moments, I_k, invariant to affine transformations as those derived from the first-, second- and third-order central moments μ_{ij}, $(i+j<3)$ and proposed by Flusser and Suk [38]. For a $\mathcal{H}$-based classification and with digitized images, μ_{ij} is computed with

$$\mu_{ij} = \sum_u \sum_v (u - \bar{u})^i (v - \bar{v})^j (1 - G_s(\mathcal{H}(u,v))), \qquad (10)$$

where $(\bar{u} = m_{10}/m_{00}, \bar{v} = m_{01}/m_{00})$ is the centroid of the region, and $G_s(\mathcal{H})$ is the filter output for the joint hue saturation color feature as input. Assuming that the object model is a simple gray rectangle $(R = G = B = 127)$, all we only need is to find the most similar region in the image to that rectangle according to these descriptors. To do so, the joint hue saturation color distribution inside the detected image region, R_q, is used to derive the four moments I_k^q, $(k = 1, \ldots, 4)$. It is clear that it is a very simple object model compared to the image of a laparoscopic instrument and the projection in the image is rather closely related with perspective projection than the orthogonal projection. As a consequence, affine invariant moments may not be the most appropriate descriptors for object recognition in the presence of strong perspective effect. Nevertheless, we do not intend to deal with completely projective invariants, we only wish to find the most similar image region to the object model. Thus, like with the Fourier descriptors, a reference set I_k^o of affine moments is built and a similar measure is evaluated for each region R_q according to

$$D_m^q = \sum_{k=1}^4 |I_k^q - I_k^o|^2. \qquad (11)$$

Since for many image regions, it is not possible to transform a rectangular object to approximate the shape with an affinity, the minimum value for D_m^q indicates the image region the most similar to the model.

4.3. Application to the trocar location

In robotized laparoscopy, it is of prime importance to predict the trocar position, since for some difficult tasks such like suturing, it is a crucial parameter in the method to plan the displacement of the instrument so as to constrain the needle to follow an acceptable path during stitching [39]. Since the central moments are computed during the computation of the affine invariant moments, the principal axis of the selected region is determined. It is both used for estimating the tip of the instrument and the incision point over an image sequence. The three second-order moments form the components of the following matrix (called the *inertial tensor*):

$$J = \begin{bmatrix} \mu_{02} & -\mu_{11} \\ -\mu_{11} & \mu_{20} \end{bmatrix} \qquad (12)$$

and the orientation of the instrument axis is given by the eigenvector associated to the smallest eigenvalue of the matrix J. In Fig. 11, the incision point of a needle-holder is computed as intersections of principal axes with 52 images. Instrument on the left of the endoscopic image of Fig. 9(d) does not move very much during the sequence leading to a coarse location of the incision point, whereas the incision point of the moving needle-holder (right) is well localized.

Table 1
Computing times for the overall segmentation (up to 4 regions) with de-interlaced color image frames of size (640 × 240) (with a Pentium IV, 3 GHz)

Video processings	Computing times (ms)
De-interlacing	~1
Image pyramid (2 levels)	5
$\mathscr{H}$ computation	2
σ estimation	5
Sigma filtering ($w = 8$, $\beta = 45\%$)	16 up to 25
Seed detection	~1
Region growing (up to 4 regions)	3 up to 11
Region labeling/boundaries	~1
Region classification (Fourier descriptors)	2 up to 4
Region classification (moments)	2 up to 5
Total	39 up to 61

5. Real-time computation issues

All previous presented image processing techniques have been thought to be integrated in a vision module for an image-guidance system with standard PC computer (3 GHz) running Linux. It is a part of computer-aided robotic assistant dedicated to surgical laparoscopy and specifically for coelioscopy. For that image processing pipeline with real-time segmentation capabilities, we have reported in Table 1, computation times measurements with endoscopic interlaced images of size (640 × 480) from 3 video sequences of 500 color images. De-interlacing is the first operation following the image acquisition. Two levels of an image pyramid decomposition are built with the even field providing an image of size (320 × 120). The join hue saturation is computed with formulae (2) and (4) for this small image together with the sigma filtering, the seed detection algorithm and the region growing process. During this experiment, no more than four regions were found to compete in the two region classification algorithms and the proposed segmentation always requires less than 80 ms, ensuring half the video-rate.

6. Conclusion and future work

We have presented a fast (half the video-rate) segmentation of gray regions in endoscopic color images. This work has been developed with a view to locating surgical instruments as well as endoscopes in a robotized laparoscopic environment. Based on a joint hue saturation color feature, we have presented an automatic region seed detection and a region classification dedicated to images of the abdominal cavity through laparoscopes. We have also improved some fast video processings which work well despite the moving background, the high reflectance of organs and instruments surfaces, the non-uniform and time-varying lighting conditions. An extension of this work would consist in a tracking software to speed up the proposed segmentation over the video sequence. It is also one step toward a computer-aided suturing system currently in progress [39] and which requires, among other things, the 3-D localization of the insertion point to guide of the surgical needle-holder.

References

[1] Sackier J, Wang Y. Robotically assisted laparoscopic surgery: from concept to development. Surgical Endoscopy 1994;8:63–6.

[2] Taylor R, Funda J, LaRose D, Treat M. A telerobotic system for augmentation of endoscopic surgery. In: Proceedings of the IEEE international conference on engineering in medicine and biology, Paris, France, 1992. p. 1054–6.

[3] Reichenspurner H, Damanio R, Mack M, Boehm D, Gulbins H, Detter C, Meiser B, Ellgass R, Reichart B. Use of the voice-controlled and computer-assisted surgical system zeus for endoscopic coronary artery bypass grafting. Journal of Thoracic and Cardiovasc Surgery 1999;118:11–6.

[4] Guthart G, Salisbury J. The intuitive™ telesurgery system: overview and application. In: Proceedings of the IEEE international conference on robotics and automation, San Francisco, USA, 2000. p. 618–21.

[5] Krupa A, Gangloff J, Doignon C, de Mathelin M, Morel G, Leroy J, Soler L, Marescaux J. Autonomous 3-D positioning of surgical instruments in robotized laparoscopic surgery using visual servoing. IEEE Transactions on Robotics and Automation, Special Issue on Medical Robotics 2003;19(5):842–53.

[6] Casals A, Amat J, Prats D, Laporte E. Vision guided robotic system for laparoscopic surgery. In: Proceedings of the IFAC international congress on advanced robotics, Barcelona, Spain, 1995. p. 33–6.

[7] Wei G-Q, Arbter K, Hirzinger G. Automatic tracking of laparoscopic instruments by color-coding. In: Verlag S, editor. Proceedings of the first international joint conference CRVMed-MRCAS'97, Grenoble, France, 1997. p. 357–66.

[8] Ascari L, Bertocchi U, Laschi C, Stefanini C, Starita A, Dario P. A segmentation algorithm for a robotic micro-endoscope for exploration of the spinal cord. In: Proceedings of the IEEE international conference on robotics and automation, New Orleans, LA, 2003. p. 491–6.

[9] Burschka D, Corso JJ, Dewan M, Hager G, Lau W, Li M, Lin H, Marayong P, Ramey N. Navigating inner space: 3-D assistance for minimally invasive surgery. In: Workshop advances in robot vision, in conjunction with the IEEE/RSJ international conference on intelligent robots and systems, Sendai, Japan, 2004. p. 67–78.

[10] Liu J, Yang Y-H. Multiresolution color image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 1994;16(7):689–700.

[11] Wyszecki G, Stiles W. Color science, concepts and methods, quantitative data and formulas. 2nd ed. New York: Wiley; 1982.

[12] Tang B, Sapiro G, Caselles V. Color image enhancement via chromaticity diffusion. IEEE Transactions on Image Processing 2001;10(5):701–7.

[13] Lukac R, Smolka B, Martin K, Plataniotis K, Venetsanopoulos A. Vector filtering for color imaging. IEEE Signal Processing Magazine, Special Issue on Color Image Processing 2005;22(1): 74–86.

[14] Faugeras O. Digital color image processing within the framework of a human visual model. IEEE Transactions on Acoustics, Speech, and Signal Processing 1979;27:380–93.

[15] Ikonomakis N, Plataniotis K, Venetsanopoulos AN. A region-based color image segmentation scheme. In: Proceedings of electrical imaging, SPIE, vol. 3653, San Jose, CA; 1999. p. 1202–9.

[16] Pujas P, Aldon M. Robust colour image segmentation. In: Seventh international conference on advanced robotics (ICAR'95), San Filiu de Guixols, Spain, 1995.

[17] Yang CC, Rodriguez JJ. Efficient luminance and saturation processing techniques for color images. Journal of Visual Communication and Image Representation 1997;3:263–77.

[18] Ballard DH, Brown CM. Computer vision. Englewood Cliffs, NJ: Prentice-Hall; 1982.

[19] Groeger M, Sepp W, Ortmaier T, Hirzinger G. Reconstruction of image structure in presence of specular reflections. In: Proceedings of the DGAM symposium on pattern recognition, Munich, 2001.

[20] Cheng HD, Sun Y. A hierarchical approach to color image segmentation using homogeneity. IEEE Transactions on Image Processing 2000;9(12):2071–82.

[21] Seeman T, Tischer P. Structure preserving noise filtering of images using explicit local segmentation. In: Proceedings of the international conference on pattern recognition, Brisbane, Australia, 1998.

[22] Pitas I, Venetsanopoulos A. Nonlinear digital filters, principles and applications. Dordrecht: Kluwer Academic Publishers; 1990.

[23] Perona P, Shiota T, Malik J. Anisotropic diffusion. In: ter Haar Romency BM, editor. Geometry-driven diffusion in computer vision. Dordrecht: Kluwer Academic Publisher; 1994. p. 73–92.

[24] Lee J. Digital image smoothing and the sigma filter. Computer Vision, Graphics, and Image Processing 1983;24:255–69.

[25] Huang TS, Yang GJ, Tang GY. A fast two-dimensional median filtering algorithm. IEEE Transactions on Acoustics, Speech and Signal Processing 1979;27:13–8.

[26] Lukac R, Smolka B, Plataniotis K, Venetsanopoulos A, Zavarsky P. Angular multichannel sigma filter. In: Proceedings of the IEEE international conference on acoustic, speech and signal processing, vol. 3, Hong-Kong, 2003. p. 745–8.

[27] Lukac R, Plataniotis K, Smolka B, Venetsanopoulos A. Generalized selection weighted vector filters. EURASIP Journal on Applied Signal Processing: Special Issue on Nonlinear Signal and Image Processing—Part I 2004;12:1870–85.

[28] Kuo C-H, Tewfik AH. Multiscale sigma filter and active contour for image segmentation. In: Proceedings of the IEEE international conference on image processing, 1999. p. 353–7.

[29] Canny J. A computational approach to edge detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 1986;8(6):679–98.

[30] Voci F, Eiho S, Sugimoto N, Sekiguchi H. Estimating the gradient threshold in the Perona-Malik equation. IEEE Signal Processing Magazine 2004;21(3):39–46,65.

[31] Burt PJ, Adelson EH. The laplacian pyramid as a compact image code. IEEE Transactions on Computers 1983;COM-31(4):532–40.

[32] de Weijer J, Gevers T, Geusebroek J. Color edge detection by photometric quasi-invariants. In: Proceedings of the international conference on computer vision, Nice, France, 2003. p. 1520–6.

[33] Adams R, Bischof L. Seeded region growing. IEEE Transactions on Pattern Analysis and Machine Intelligence 1994;16(6):641–7.

[34] Otsu N. A threshold selection method from gray-level histograms. IEEE Transactions on Systems, Man and Cybernetics 1979;9(1): 62–6.

[35] Reddi S, Rudin S, Keshavan H. An optimal multiple threshold scheme for image segmentation. IEEE Transactions on System, Man and Cybernetics 1984;14(4):661–5.

[36] Persoon E, Fu K. Shape discrimination using fourier descriptors. IEEE Transactions on Systems, Man and Cybernetics 1977;7(3): 170–9.

[37] Kuhl F, Giordani C. Elliptic fourier features of a closed contour. Computer Vision, Graphics and Image Processing 1982;18:236–58.

[38] Flusser J, Suk T. Pattern recognition by affine moment invariants. Pattern Recognition 1993;16(1):167–74.

[39] Nageotte F, Zanne P, de Mathelin M, Doignon C. A circular needle path planning method for suturing in laparoscopic surgery. In: Proceedings of the IEEE international conference on robotics and automation, Barcelona, Spain, 2005. p. 516–21.