Int J CARS manuscript No.
(will be inserted by the editor)

An algorithm for calculi segmentation on ureteroscopic images

Benoît Rosa · Pierre Mozer · Jérôme Szewczyk

Received: 28 August 2009 / Accepted: 5 June 2010

This document is the final draft of the paper, as accepted by IJCARS for publication.
The final publication is available at www.springerlink.com or via http://dx.doi.org/10.1007/s11548-010-0504-x

Abstract

Purpose The purpose of the study is to develop an algorithm for the segmentation of renal calculi on ureteroscopic images. In fact, renal calculi are common source of urological obstruction, and laser lithotripsy during ureteroscopy is a possible therapy. A laser-based system to sweep the calculus surface and vaporize it was developed to automate a very tedious manual task. The distal tip of the ureteroscope is directed using image guidance, and this operation is not possible without an efficient segmentation of renal calculi on the ureteroscopic images.

Method We proposed and developed a region growing algorithm to segment renal calculi on ureteroscopic images. Using real video images to compute ground truth and compare our segmentation with a reference segmentation, we computed statistics on different image metrics, such as Precision, Recall, and Yasnoff Measure, for comparison with ground truth.

Results The algorithm and its parameters were established for the most likely clinical scenarii. The segmentation results are encouraging: the developed algorithm was able to correctly detect more than 90% of the surface of the calculi, according to an expert observer.

Conclusion Implementation of an algorithm for the segmentation of calculi on ureteroscopic images is feasible. The next step is the integration of our algorithm in the command scheme of a motorized system to build a complete operating prototype.

Keywords Ureteroscopy · Medical Imaging · Laser Lithotripsy · Image Segmentation

Benoit Rosa
UPMC Univ Paris 06, UMR 7222, ISIR, F-75005, Paris, France
E-mail: benoit.rosa@centraliens.net

Pierre Mozer
UPMC Univ Paris 06, UMR 7222, ISIR, F-75005, Paris, France, and AP-HP, Pitié-Salpétrière Hospital, Department of Urology, F-75013, Paris, France
E-mail: pierre.mozer@psl.aphp.fr

Jerome Szewczyk
UPMC Univ Paris 06, UMR 7222, ISIR, F-75005, Paris, France
E-mail: jerome.szewczyk@isir.fr

1 Introduction

Urolithiasis is a frequent pathology, affecting about 10% of people over 40 years of age in industrialized countries. The recurrence rates are estimated to be 53% [1]. Until the end of the 1970s, open surgery was the only effective therapy. However, the frequency of stone disease, its evolution into frequent recurrences and the problems caused by open surgery led to the development of new therapeutic approaches. As a result, during the last 30 years, technical progress allowed for the development of less-invasive techniques.

The European Association of Urology recommends three therapies [2]: Percutaneous Nephrolithotomy (PCNL), Extracorporal Lithitripsy (ECL) and ureteroscopy. PCNL is the most invasive technique, which is used for extreme cases, such as very large calculi or failure of previous surgeries, while ECL is very effective for small calculi. Furthermore, endoscopic techniques via natural ways are recommended for 10-20 mm calculi.

This paper focuses on these last techniques, especially on flexible ureteroscopy. The operation introduces the device into the kidney through the urethra, bladder, and ureter. The device is flexible and can be bent at its distal tip. This command is made in a plane by moving the cables at the proximal tip. The whole kidney can be explored by combining the movements of the distal tip with rotations of the whole device. Though different ureteroscopes exist [3], their external diameter is often 12 Fr (1 Fr = 1/3 mm). They have an optical fiber matrix which returns an image and an operating channel of 3.6 Fr in most cases, which allows the surgeon to pass surgical instruments.

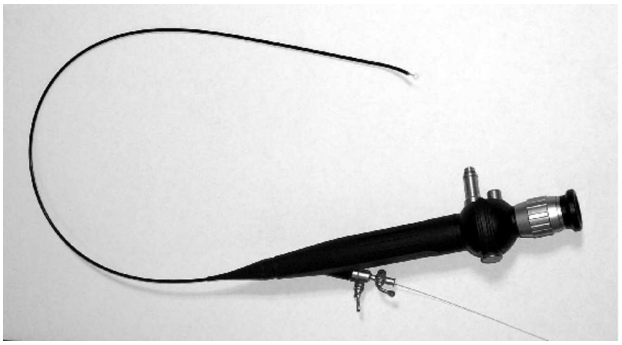

Fig. 1: A Storz 11274AA1 ureteroscope, with a laser fiber passing through the operating channel

After the introduction of the ureteroscope in the kidney, the calculus can be treated with an optical fiber passing through the operating channel (see Figure 1). In most of the cases, a Ho:YAG laser is used [4]. The laser has two distinct states: an inactive state in which a red laser diode projects a red spot on the calculus indicating where the fiber points to, and an active state, activated by a pedal, in which the Ho:YAG power laser passes through the fiber at a rate of 0.8 Hz.

Ideally, the surgeon should sweep the lithiasis surface with laser to vaporize it. This operation avoids fragmentation of the lithiasis, which makes the surgeon bring the fragments outside the patient one by one, using a wire basket. However, manual sweeping is very difficult owing to the poor maneuverability of the ureteroscope, and usually, a surgeon may take about an hour to vaporize a 10 mm wide lithiasis. Thus, the main aim of this study is to assist the surgeon in this tedious task. The objectives are to minimize the surgery time and possibly to treat larger calculi of up to 20 mm. An additional constraint is to avoid burdening the surgeon with additional work, because during the operation, he/she must handle the video return and the foot pedal for the laser while manipulating the ureteroscope.

Similar to active catheter, motorization of the distal tip of the instrument using shape memory alloys has been proposed [5,6]. Little movements of the distal tip necessary for calculus vaporization can be carried out automatically, while larger movements necessary to bring the device in the calculus zone should be handled by the surgeon. The automatic sweeping task needs careful planning of the movement to avoid shooting outside the lithiasis. Hence, ureteroscopic images have to be segmented to extract a safe region inside the lithiasis (Figure 2).

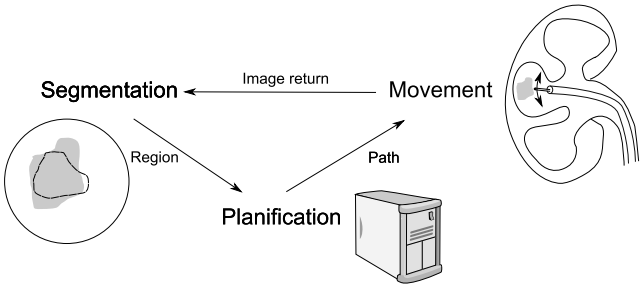

Fig. 2: The proposed command scheme

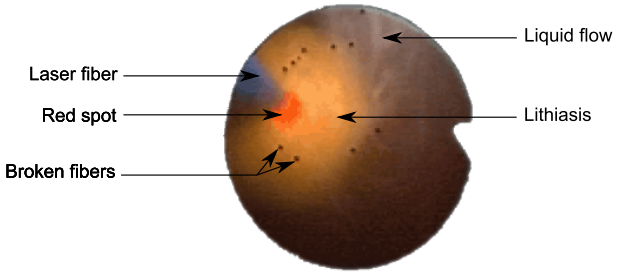

Fig. 3: Video image from a Storz Flex-X2 ureteroscope

This study has focused on the lithiasis segmentation algorithm. There are numerous difficulties in carrying out this task. First, the kidney moves within the patient owing to breathing, at several centimeters of amplitude [7], and irrigation liquids are used during the operation. These two constraints have a direct effect on image quality. Moreover, it is not uncommon to break some optical fibers, resulting in black dots on the image (Figure 3). Second, lithiasis has demonstrated different chemical compositions, resulting in different shapes, colors and textures [8]. Finally, the system must be fast enough to work over the laser shooting rate. Though medical image segmentation is a very active domain [9,10], few studies have examined this particular subject. Most of them have focused on MRI [11,12], CT scan or ultrasound images [13–15]. Except for laparoscopic images [16], only a few researches have directly examined

video images.

To our knowledge, no other study has been carried out on lithiasis segmentation on ureteroscopic images. In this study, an algorithm that performs this function semi automatically, without the need for an interaction from the user, except at the initiation stage has been presented, and its precision, robustness, and speed have been examined.

2 Materials and methods

The study was carried out on 10 videos taken during interventions. These videos were taken with the Storz Flex-X2 ureteroscope and the Storz software suite (KARL STORZ, Germany). Their resolution was 720x576, and they were encoded with Radius Cinepak Codec. They had a typical framerate of 25 frames/s.
The videos presented four different kinds of calculi, whose size varied from 9 to 18 mm (Table 1).

With regard to image processing, all the algorithms were implemented using the C and C++ languages with the OpenCV library (Intel Corporation, USA).

The computer was a PC with an Athlon 64 X2 4200+ processor (Advanced Micro Devices, USA) and a 2 GB RAM.

2.1 Segmentation algorithm

2.1.1 Region growing algorithm

As stated in the Introduction section, the segmentation of the calculi in the ureteroscopic images is made difficult by:

- The perturbated environment: Liquids flowing make the limits of the shapes blurry, and patients breathing causes movements of the image. Moreover, the luminosity can change drastically, depending on the location in the kidney.
- The biological variety of the calculi's chemical composition. As the calculi are a concretion of crystalline materials, their shape and size can greatly vary. Moreover, their color and texture are greatly affected by their chemical composition. Colors can vary from yellow to black, and textures can differ from a very smooth surface to a texture composed of spots with variable luminosity, mainly owing to specular reflections.
- The fragility of the optic fibers produces black dots of a non negligible size, which appear on the image.

The region growing algorithm [17] was used. This algorithm is simple to implement for a first approach of the problem. It requires a seed, and the definition of a similarity criterion along with a stopping criterion.

The seed is the starting point of the region growth. The position of the seed must be inside the calculus in the image. The segmented image is then a binary image representing the inside and the outside of the calculus on the image.

The definition of the similarity criterion is the key point of the region growing algorithm: it must allow to differentiate the calculus from the rest of the the image. Hence, it must be able to work with calculi of different colors and textures to ensure a proper functioning in most of the clinical scenario.

The similarity criterion gives a score representing the similarity between the region already found and the pixel or group of pixels examined. The stopping criterion is defined as a threshold on the value of the similarity criterion. Its value must be optimized to give the best segmentation possible.

The region growing algorithm is able to answer the main problems raised previously. In fact, the use of a similarity criterion directly answers to the problem of the biological variety of the chemical composition of calculi: it is the similarity between a pixel or group of pixels and the already found region which is taken into account. Thus, given the fact that the seed is inside the lithiasis, the optimization of the stopping criterion value can answer the problem efficiently. Similarly, the definition of the size of the group of pixels considered by the similarity criterion can answer the problem of the texture variability of the pixels and the problem of the black dots appearing on the image.

2.1.2 Implementation

The main concept is simple. By knowing a point inside the region, the algorithm looks for neighbors and sees if they belong to the region or not, using the similarity criterion and the stopping criterion. It loops with these new points and their neighbors until no more point is added. Owing to the fact that the aim is to sweep the surface of a lithiasis using precise little movements of the distal tip of the ureteroscope, the analysis of the whole image is not required. Hence, a square of 80x80 pixels is set as a limit to the region growth.

The algorithm consists of three steps (Figure 4):

In the first step, the position of the seed in the image is determined. The detection uses the fact that the surgeon pushes the laser pedal when the fiber points toward the calculus. At this point, the laser fiber projects a red spot on the calculus surface. This spot can easily be detected because

Video	Used images	Lithiasis size (mm)	Lithiasis composition
1	353	18	struvite
2	248	13	calcium oxalate dihydrate
3	59	18	calcium oxalate dihydrate
4	10	12	struvite
5	12	15	calcium oxalate monohydrate
6	11	12	calcium oxalate dihydrate
7	55	11	carbonate apatite
8	78	15	struvite
9	17	9	calcium oxalate monohydrate
10	53	15	calcium oxalate monohydrate

Table 1: Characteristics of the 10 used videos

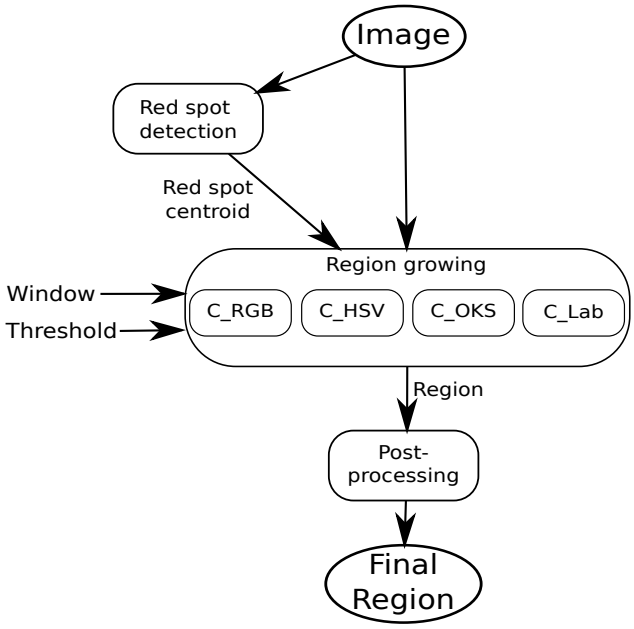

Fig. 4: The different steps and parameters of the algorithm

the laser light has a well-defined color and a high luminosity. The image color space is transformed into HSV color space, and the red spot is detected, using its color and luminosity characteristics. Its centroid is then used as a seed for the region growth.

In the second step, the region growing itself is applied. The key point of this step is the definition of the homogeneity criterion. In [18], Tremeau and Borel defined an *Average Homogeneity Criterion* in RGB color space. This criterion computes the Euclidean color distance between the color value of the considered pixel and the mean value of the already found region. This criterion was modified, using a *window* parameter to reduce the effects of texture variability. Instead of comparing a unique considered pixel with the already found region, the whole window is considered. If the result is inferior to the stopping criterion, the whole window is integrated to the region. Otherwise, no pixel is integrated to it. The size of the window was limited to the [3x3, 5x5, 7x7, 9x9] range because of the size of the search region (80x80).
Four different criteria have been tested:

- The color distance in RGB color space, which is the most common color space. Because of the red spot around the seed, the Green and Blue layers are weighted more than the Red layer ($\alpha = 1/4$. This value was determined experimentally by an expert by considering visual results obtained with different values of α):

$$C_{RGB} = \sqrt{\alpha(R_{win} - R_{reg})^2 + (G_{win} - G_{reg})^2 + (B_{win} - B_{reg})^2} \quad (1)$$

- The color distance in HSV color space, which is commonly recognized to be closer to the human perception of color than the RGB color space. If the window belongs to red spot, $C_{HSV} = 0$. Else,

$$C_{HSV} = \sqrt{(H_{win} - H_{reg})^2 + (S_{win} - S_{reg})^2 + (V_{win} - V_{reg})^2} \quad (2)$$

- The color distance in (I1,I2',I3') color space. (I1,I2',I3') has been proven by Ohta, Kanade and Sakai to be effective in color region segmentation [19,20]. Compared to RGB color space, I1 is the mean value of the three RGB layers, I2' is the difference between Red and Blue layers and $I3' = \frac{2G-R-B}{2}$:

$$C_{OKS} = \sqrt{(I^1_{win} - I^1_{reg})^2 + (I^{2'}_{win} - I^{2'}_{reg})^2 + (I^{3'}_{win} - I^{3'}_{reg})^2} \quad (3)$$

- The color distance in CIE-Lab color space, in which the Euclidean color distance is strongly related to the human perception of color distance [21]:

$$C_{Lab} = \sqrt{(L_{win} - L_{reg})^2 + (a_{win} - a_{reg})^2 + (b_{win} - b_{reg})^2} \quad (4)$$

In these equations, the subscripts *win* and *reg* mean window and region, and indicate that the mean value is used. For instance, R_{reg} is the mean value of the pixels of the Red layer of the already found region.

The last step in the algorithm is post-processing, which allows filling eventual holes in the found region. It consists of morphological closing with a square structuring element of the same size as the window.

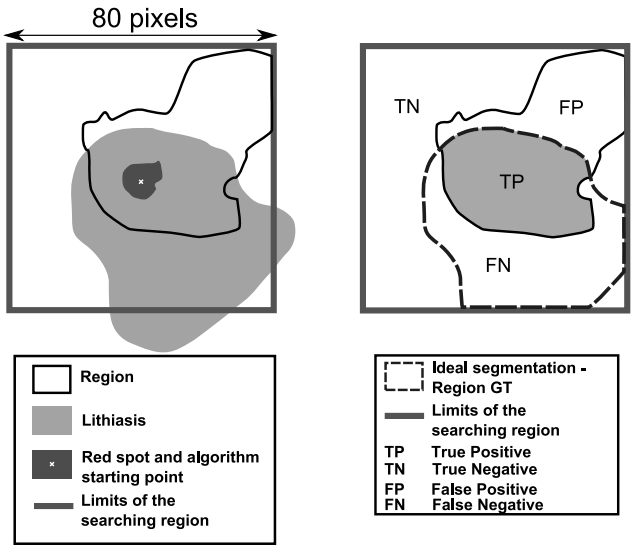

Fig. 5: Diagram representing the different considered regions

2.2 Validation method

To validate the algorithm and optimize its parameters, the images were manually segmented by an expert to constitute a ground truth. Only the images in which the laser red spot is inside the calculus were kept. A total of 923 images from the 10 videos were segmented.

Several indicators were defined to evaluate the performance of segmentation by the algorithm, when compared with an ideal segmentation from the ground truth. These indicators make use of the different regions defined in Figure 5.
Region denotes the region found by the algorithm; region GT stands for the ground truth; and TP, TN, FP, and FN are defined as:

$$TP = region \cap GT \tag{5}$$
$$FP = region \cap \overline{GT} \tag{6}$$
$$TN = \overline{region} \cap \overline{GT} \tag{7}$$
$$FN = \overline{region} \cap GT \tag{8}$$

For simplicity, the same notation is used for a region and its pixel number. For instance, TP refers to the True Positive region or to its pixel number, depending on the context.

2.2.1 Quantitative error measurement

Several indicators were computed

Precision, Recall and F-score The first and simplest idea is to evaluate the performance of the algorithm using the surface of correctly and incorrectly segmented pixels, by employing common image metrics such as Precision, Recall and F-score. These indicators come from semantic researches and the separation of correct and incorrect results [22]. They are defined as:

$$Recall = \frac{TP}{TP+FN} \tag{9}$$
$$Precision = \frac{TP}{TP+FP} \tag{10}$$
$$F = \frac{2 * Recall * Precision}{Recall + Precision} \tag{11}$$

Precision informs about the proportion of correctly segmented pixels in the found region (TP + FP), whereas Recall informs about the proportion of correctly segmented pixels compared to the ground truth (TP + FN). These two metrics are complementary and must be used together. Consequently, the Precision/Recall (PR) curve is used. In PR space, the better the algorithm is, the closer its curve gets to the upper right-hand corner of the graph. It corresponds to a maximization of Precision and Recall altogether. F-score is the harmonic mean of Precision and Recall and is used to summarize the information given by Precision and Recall in a single metric.

For a total match, $Precision = Recall = F = 1$. However, a total mismatch is impossible, because the seed hypothetically belongs to TP. The limit values are:

$$\lim_{\substack{TP \to 0 \\ TN=0}} Recall = 0 \tag{12}$$
$$\lim_{\substack{TP \to 0 \\ TN=0}} Precision = 0 \tag{13}$$
$$\lim_{\substack{TP \to 0 \\ TN=0}} F = 0 \tag{14}$$

Relative compactness Compactness represents the dispersion of the contour of a region when compared with a circle.

$$c = \frac{4\pi A}{P^2} \tag{15}$$

where P is the perimeter of the considered region and A its number of pixels.

This indicator is not used as such, because the region does not necessarily need to be compact. It is used as an indicator for texture sensitivity of the algorithm. Claridge et al. [23] reported two types of border irregularities – texture and structure irregularities. Texture irregularities are the fine variations along the border, and structure irregularities are the general undulations of the shape. Compactness may

be used to measure structure irregularities. However it is very sensitive to noise along the border, and therefore to texture irregularities [24]. The Compactness indicator was then adapted to fit the case studied here: the measure of texture irregularities without taking into account structure irregularities. Hence, a *Relative Compactness* indicator was created. It consists of dividing region Compactness by ground truth Compactness:

$$Rc = \frac{c(region)}{c(GT)} \tag{16}$$

With the Relative Compactness reaching closer to 1, the algorithm's sensitivity to texture becomes lesser. To limit the influence of structure irregularities, this indicator is used only when the F-score is near to its maximum value and Yasnoff values are low, which means that the overall shapes of the region and the ground truth are very close.

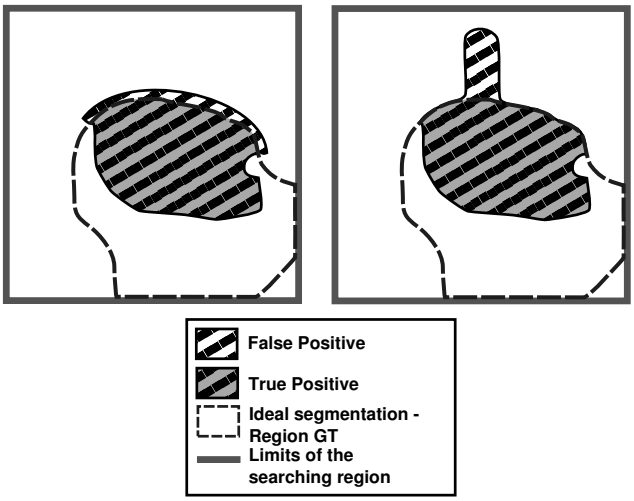

Fig. 6: Two segmentation examples to illustrate the need for Yasnoff measure

Yasnoff measure Yasnoff measure [25] gives an indication on the dispersion of the pixels of FP.

With regard to the two examples given in Figure 6, as the numbers of pixels in regions TP, FP, TN, FN, and GT are the same, they have the same Precision, Recall, and F-score. However, the right segmentation is much worse than the left, because it could lead to a laser shot significantly outside the calculus. Yasnoff measure gives a piece of information about the dispersion of the incorrectly segmented pixels. It is defined as:

$$\varepsilon = \frac{100}{A} * \sqrt{\sum_{i=1}^{k} d_i^2} \tag{17}$$

where A is the number of pixels of the searching region, k is the number of false-positive pixels, and d_i is the distance between the ith pixel of FP and its nearest neighbor in GT. For the images given in Figure 6, $\varepsilon = 7.8$ for the left segmentation and $\varepsilon = 57.8$ for the right one.

This definition of the Yasnoff measure only takes FP into account, and not FN. This was chosen for clinical reasons. In fact, the system must ensure that it will not shoot outside the calculus. Hence, the most important thing is that the segmentation algorithm has low FP. If there is no FP, the Yasnoff measure will indicate 0. This will not mean that contours match totally, but it will mean that no laser shooting could be made outside the calculus, using the algorithm found region. Moreover, if there was few FP and a lot of FN, Precision would be near to 1 and Recall near to 0, so the bad segmentation would be highlighted by a PR-curve near the Precision axis.

2.2.2 Limit values for indicators

Indicators for quantitative error measurement have been defined in the previous section. The operating process consists in computing these indicators on each image, with regard to the ground truth and the found region, using several combinations of the parameters defined in Section 2.1.2. However, figures without interpretation are not very informative. Hence, limit values were defined for the indicators to separate good and bad segmentations.

PR curves are essentially used for comparison of algorithms: if an algorithm curve is above another, this algorithm is superior in terms of Precision and Recall. A particular emphasis was put on the 0.9 value for Precision and Recall: if Precision and Recall are greater than 0.9, it means that the algorithm returns a region which covers 90% of the ground truth with less than 10% of error. On the PR-curve, this corresponds to the 0.9 upper right-hand square.

The limit value for Yasnoff measure was set experimentally. In fact, Yasnoff measure is used to minimize the number of FP in the region. This particular dissymmetry was introduced to limit the risk of shooting outside the calculus (see sect. 2.2.1). Hence, the limit value of the Yasnoff measure was set in comparison with the diameter of the laser fiber: if the size of the FP region is greater than half the diameter of the laser fiber, the risk of shooting oustide the calculus is important. The average size of the tip of the laser fiber in the images was computed by analyzing images from most common clinical situations and was found to be 40pixels. Then, Yasnoff value was computed for a FP region composed of a circle of 20pixels diameter. A value of $\varepsilon = 13.95$ was found. The maximum Yasnoff value was then

set to $\varepsilon = 14$.

With the Relative Compactness reaching closer to 1, the algorithm's sensitivity to texture becomes lesser. A typical security range of 5% was used, setting limit Rc values to $0.95 < Rc < 1.05$.

3 Results

The parameters of the algorithm were set up on 660 images taken from the first three videos, and validated and refined on 263 images taken from the rest of the seven.

3.1 Set-up of the parameters

The aim of this study was to find a set of parameters which is suitable for the most likely clinical scenario. The parameters to test were the following:

- The similarity criterion must be chosen among the four developed criteria: C_{RGB}, C_{HSV}, C_{OKS} and C_{Lab}.
- The window size, inside the [3x3,9x9] range.
- The stopping criterion, which is a threshold on the value of the similarity criterion.

The strategy to optimize the parameters was conducted through three experiments:

- First, the best criterion was found using PR-curves. The impact of the window size on the F-score values was also studied. Moreover, an inferior bound for the threshold was found.
- The best window size was found using the Relative Compactness indicator.
- Finally, the upper bound for the threshold was determined using the Yasnoff measure.

To correctly set up the parameters, first, one must determine the best criterion among the four described in Section 2.2.1. Figure 7 shows a comparative study of the results obtained by the four different criteria. It can be noted that except C_{HSV}, all the criteria achieved a Precision and a Recall of more than 0.9 for at least one threshold value. The PR-curve for the RGB criterion was above the others, and got the maximum F-score of 0.954 at a threshold of 90. Therefore, C_{RGB} was kept as the best criterion among those developed. An inferior bound for the threshold value was also set to 70, because it is the first threshold value which lets the PR-curve of C_{RGB} get in the 0.9 upper right-hand square.

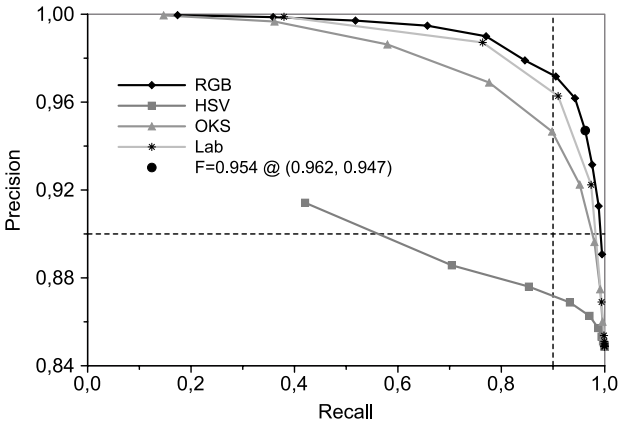

Fig. 7: PR-curves for the four different criteria. The window size is fixed to 9x9. The maximum F-score is observed with RGB criterion for a threshold of 90

The choice of C_{RGB} has been made with a window size fixed to 9x9. Hence, this choice is only valid if the F-score values do not depend on the window size. A Spearman correlation test was carried out to compare the results given by the four criteria with a window size of 9x9 pixels and their results with a window size of 3x3, 5x5 and 7x7 pixels. The test was conducted independently for each criterion. Threshold values were ranked according to the corresponding value of the F-score. Then the ranks were compared for different window sizes and a coefficient ρ was computed. Table 2 shows the results of the test. As there were 12 observations (threshold varying from 10 to 120 with a step of 10) the absence of correlation could be rejected with a confidence of $999^{0}/_{00}$, because every value of ρ is greater than 0.8599. This test meant that for every criterion, F-score values evolved in the same way regardless of the window size. Therefore, it could be accepted that the window size had a minor influence on the F-score values. Hence, the choice of the C_{RGB} criterion was validated, with a minimum threshold value of 70.

Criterion	ρ_{93}	ρ_{95}	ρ_{97}
C_{RGB}	0.993	0.993	0.986
C_{HSV}	0.874	0.993	1
C_{OKS}	0.930	0.951	0.951
C_{Lab}	1	1	1

Table 2: Results of Spearman rank correlation test. The subscripts for ρ indicate the size of the considered windows (ρ_{93} for comparison of 9x9 pixels and 3x3 pixels windows)

Two criteria remained to be set up: the size of the window and the upper bound of the threshold. As the window was introduced to reduce the effects of texture variability,

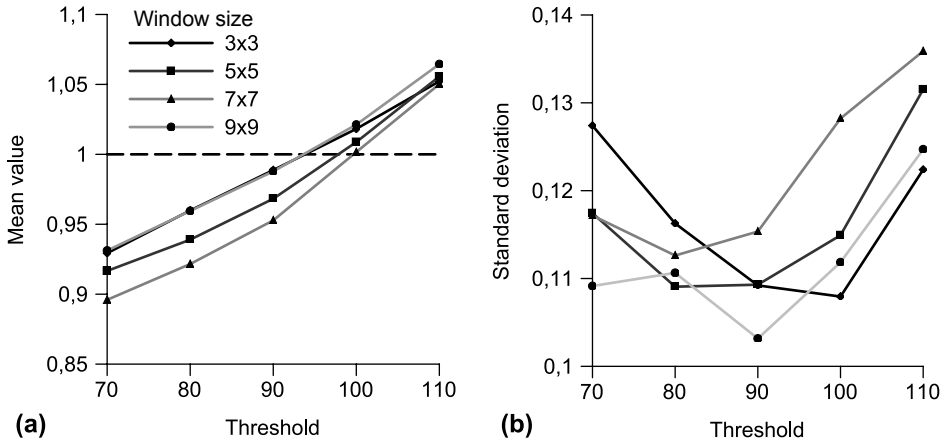

Fig. 8: Rc depending on the threshold for different window sizes. (a) mean value, (b) standard deviation

the Relative Compactness indicator must be used to set its size. Figure 8 shows the mean value and standard deviation of this indicator for different sizes of the window in the acceptable threshold range found previously. Though the mean values were very close, windows of 3x3 and 9x9 pixels distinguished themselves from the other values, especially for a threshold lesser than 100. Moreover, the standard deviation with a 9x9 pixel window was often less than the others, and greatly inferior to the values obtained with a 3x3 window. Thus, the size of the window was fixed to 9x9.

Finally, the upper bound of the threshold was chosen by using the Yasnoff measure. An inferior bound was already defined with the PR-curves, but the highest threshold could not be used. In fact, if the threshold was too high it no longer played its role of stopping criterion and the whole window of 80x80 pixels was included in the region. This could lead to laser shots outside the calculus. Hence, the Yasnoff measure was used.

Thresh.	Mean	Std. dev.	Max. value	9th decile
70	2.008	3.819	29	5
80	2.842	4.353	33	7
90	3.923	4.803	33	9
100	5.111	5.295	33	12
110	6.663	5.700	34	14

Table 3: Statistics of the Yasnoff values

Table 3 shows the statistical values computed on the Yasnoff values. The standard deviation can be used to determine a confidence interval of 95% around the mean value, stating that for a threshold inferior to 100, 95% of the values are inferior to 15.7. This can only be certain if the statistical distribution is gaussian, which is not known here. However, the values of the 9th decile show us that for a threshold of 100, 90% of the values are inferior to 12. The data dit not allow to clearly state that most of the Yasnoff values were inferior to 14 for a threshold value of 110, so this value was eliminated from the acceptable range.

3.2 Validation and refining

The following set of parameters has been found to be acceptable, with regard to the values of different image metrics: a region growing algorithm with C_{RGB} criterion, a window of 9x9 pixels, and a threshold between 70 and 100. However, the set-up of the algorithm was made on an image bank composed of many images coming from only three videos. Hence, the results must be validated on a larger image bank to ensure the robustness of the algorithm regarding biological variability (calculi sizes, chemical compositions, textures). The validation image bank consists in 263 images taken from 7 videos which present calculi of 4 chemical compositions, with sizes varying from 9 to 15 mm.

On the validation image bank, the Yasnoff values were greater than those in the set-up, which meant that a refining of the parameters was needed: Table 4 shows that for threshold values of 100 and 110, the 9th decile of the Yasnoff values was greater than 14, and Relative Compactness values were outside the acceptable range. Hence, the acceptable range for the threshold was reduced to [80,90].

Figure 9 shows the PR-curve of the algorithm on the validation image bank. It can be noted that the general tendency is the same as that observed in Figure 7, although the values of the results are quite low. The algorithm curve gets in the 0.9 upper right-hand square for the threshold value of 90 only. For this value, Precision and Recall values are 0.913

Thresh.	Rc	F	ε		
			mean	std. dev	9th decile
80	0.98	0.90	4.606	4.353	10
90	1.01	0.91	5.983	4.803	13
100	1.05	0.92	7.847	5.295	16
110	1.08	0.91	9.593	5.700	18

Table 4: Results of the validation image bank

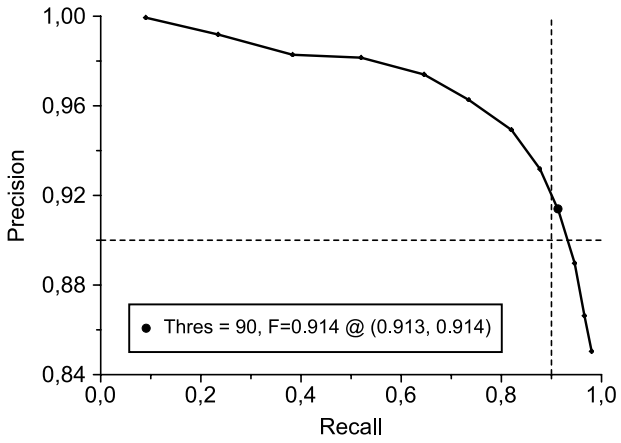

Fig. 9: PR-curve of the algorithm on the validation bank

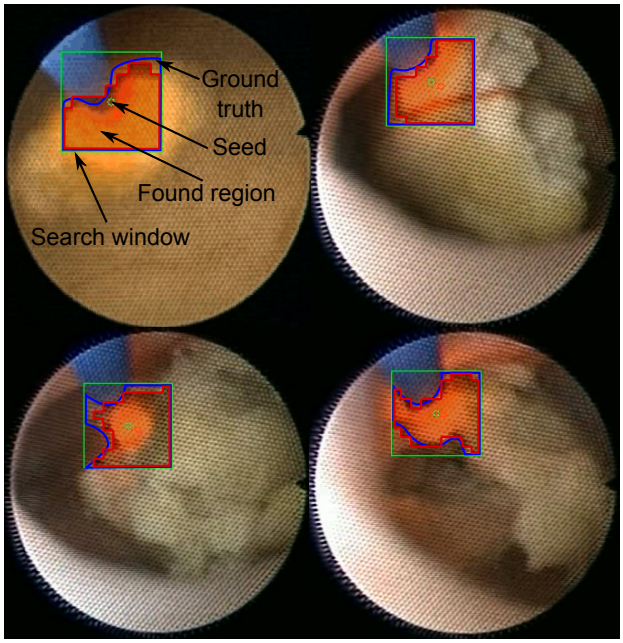

Fig. 10: Examples of segmentations obtained on real video images with the final set-up of the parameters

and 0.914. Hence, the threshold was set to the single value of 90.

Finally, the results found in the set-up were refined using the validation image bank. The final algorithm used the C_{RGB} criterion with a window of 9x9 pixels and a threshold of 90. With this set-up, the algorithm results were encouraging: an expert observer analyzed the results of the algorithm on the images of the set-up and validation datasets and concluded that out of the 923 images, 871 were correctly segmented (94.3%). Examples of segmentation results can be found on Figure 10.

3.3 Computational time

The algorithm has a maximum computational time of 62 ms for a single image, which allows it to run at a framerate of 15 Hz. This is much faster than the laser shooting rate of 0.8 Hz, and thus, the algorithm matches its speed constraint. It must be taken into account that owing to the use of an encoded video file, the computational time includes the decoding time. However, in the final application, a video acquisition will be used, and hence, the decoding time will be replaced by an acquisition time. It must also be noted that the OpenCV library is optimized for Intel processors. On an Intel Core 2 Duo P8700 Processor, the algorithm runs at more than 30 Hz (video framerate: 25 Hz).

4 Analysis and Discussion

4.1 Value of the threshold

As shown in Sections 3.1 and 3.2, the threshold is set to 90. This single value is interesting, because it avoids burdening the surgeon with a set up of the algorithm. Although this parameter was set after a study made on a dataset composed of more than 900 images, this set-up could lead to uncertain results in particular cases. A solution would be to preset the algorithm with the found parameters, and to allow the surgeon to adjust the parameters around this preset at the beginning of the operation. This must be taken into account when designing the end-user application and the tests.

4.2 Identified bias

The proposed method is not bias-free. The ground truth has been traced by hand, which is prone to variability. However, this has not been examined further since it is not the aim of this paper.

Moreover, the region growing seed and red-spot detection were examined. This spot is produced by a laser light, very coherent with a determined wavelength, and hence, easy to detect. The detection rate for the red spot is near 100%, but neither the impact of specular reflections on the surface of the lithiasis, nor the impact of eventual blood drops flowing through the kidney was examined.

The processed videos presented calculi with irregular shapes. This is a consequence of the current surgical method in which the surgeon, owing to the lack of maneuverability, often aims at the same points while shooting with the laser, digging holes in the calculi and fragmenting them. It should be noted that the algorithm will be included in a system that will gradually vaporize the calculus surface, which should improve the results of the algorithm.

5 Conclusion and perspectives

The aim of this work was to propose an algorithm for segmentation of calculi on ureteroscopic images. A region growing algorithm was developed and set up for the most likely clinical scenario. Manual segmentations were made by an expert on real video images to establish ground truth, and a quantitative error measurement led us to fix the algorithm parameters. An adaptation of the *Local Homogeneity Criterion* was used in RGB color space. The stopping criterion was a threshold of 90, and the window size was set to 9x9 pixels. With this set-up, the algorithm was able to obtain good results on a validation image bank composed of images coming from different clinical situations: a F-score of 0.91, with a Relative Compactness indicator of 1.01 and Yasnoff measure values inferior to 13 in 90% of the cases and inferior to 15.5 in 95% of the cases. Moreover, the algorithm was fast enough to run at a framerate of 31 Hz. However, the algorithm has still drawbacks and bias that must be studied further to improve its results and robustness.

The developed algorithm was developed to be integrated in an automated system to sweep the surface of renal calculi and vaporize them. The algorithm will be used to compute a path planning and a visual servoing scheme to achieve such results. On the other side, the motorization of the distal tip and its control scheme are currently under development.

Acknowledgements The authors thank the referees for their comments that have helped to improve the paper significantly. They are also grateful to Clément Chastagnol, Mathilde Irida and Pauline Maurice, for their helpful comments on the first draft of this paper.

References

1. R. Bartoletti, T. Cai, N. Mondaini, F. Melone, F. Travaglini, M. Carini, and M. Rizzo. Epidemiology and risk factors in urolithiasis. *Urol Int*, 79 Suppl. 1:3–7, 2007.
2. G. M. Preminger, H. G. Tiselius, D. G. Assimos, P. Alken, A. C. Buck, M. Gallucci, T. Knoll, J. E. Lingeman, S. Y. Nakada, M. S. Pearle, K. Sarica, C. Trk, and J. S. Wolf. 2007 Guideline for the management of ureteral calculi. *Eur Urol*, 52:1610–1631, Dec 2007.
3. C. Abdelshehid, M. T. Ahlering, D. Chou, H. K. Park, J. Basillote, D. Lee, I. Kim, L. Eichel, D. Protsenko, B. Wong, E. McDougall, and R. V. Clayman. Comparison of flexible ureteroscopes: deflection, irrigant flow and optical characteristics. *J Urol*, 173:2017–2021, Jun 2005.
4. G. B. Boline and J. A. Belis. Outpatient fragmentation of ureteral calculi with mini-ureteroscopes and laser lithotripsy. *J Endourol*, 8:341–343, Oct 1994.
5. J. Szewczyk, R. Blanc, A. Gaston, and PH. Bidaud. Active catheters for neuro-radiology. Chambery, France, 2007. 2nd Surgetica conf.
6. Y. Haga, Y. Tanahashi, and M. Esashi. Small diameter active catheter using shape memory alloy. In *Proceedings of IEEE international workshop micro electro mechanical systems*, pages 419–424, 1998.
7. J. R. van Sornsen de Koste, S. Senan, C. E. Kleynen, B. J. Slotman, and F. J. Lagerwaard. Renal mobility during uncoached quiet respiration: an analysis of 4DCT scans. *Int J Radiat Oncol Biol Phys*, 64:799–803, Mar 2006.
8. D. B. Leusmann. A classification of urinary calculi with respect to their composition and micromorphology. *Scand J Urol Nephrol*, 25:141–150, 1991.
9. J. S. Duncan and N. Ayache. Medical image analysis: progress over two decades and the challenges ahead. *IEEE Trans Patt Anal Machine Intell*, 22:85–106, 2000.
10. P. N. Rao. Imaging for kidney stones. *World J Urol*, 22:323–327, Nov 2004.
11. P. H. Vivier, M. Dolores, I. Gardin, P. Zhang, C. Petitjean, and J. N. Dacher. In vitro assessment of a 3D segmentation algorithm based on the belief functions theory in calculating renal volumes by MRI. *Am J Roentgenol*, 191:127–134, Sep 2008.
12. N. Makni, P. Puech, R. Lopes, A. S. Dewalle, O. Colot, and N. Betrouni. Combining a deformable model and a probabilistic framework for an automatic 3D segmentation of prostate on MRI. *Int J Comput Assist Radiol Surg*, 4:181–188, Mar 2009.
13. S. Sridhar, N. Kumaravel, and K. S. Easwarakumar. Segmentation of renal calculi in ultrasound images. *Med Inform Internet Med*, 27:229–236, Dec 2002.
14. J. Xie, Y. Jiang, and H. T. Tsui. Segmentation of kidney from ultrasound images based on texture and shape priors. *IEEE Trans Med Imaging*, 24:45–57, Jan 2005.
15. R. San José Estépar, C. F. Westin, and K. G. Vosburgh. Towards real time 2D to 3D registration for ultrasound-guided endoscopic and laparoscopic procedures. *Int J Comput Assist Radiol Surg*, 4:549–560, Nov 2009.
16. S. Voros, E. Orvain, P. Cinquin, and J. A. Long. Automatic detection of instruments in laparoscopic images: a first step towards high level command of robotized endoscopic holders. In *Proceedings of BioRob*, pages 1107–1112, 2006.
17. R. Adams and L. Bischof. Seeded Region Growing. *IEEE Trans Pattern Anal Mach Intell*, 16(6):641–647, 1994.
18. A. Tremeau and N. Borel. A region growing and merging algorithm to color segmentation. *Patt Recognit*, 30(7):1191–1203, 1997.
19. Y. I. Ohta, T. Kanade, and T. Sakai. Color information for region segmentation. *Comput graphics image process*, 13(3):222–241, 1980.
20. N. R. Pal and S. K. Pal. A review on image segmentation techniques. *Patt recognit*, 26(9):1277–1294, 1993.
21. G. Wyszecki and W. S. Stiles. *Color science*. Wiley New York, 1967.
22. C. J. Van Rijsbergen. *Information retrieval*, chapter 7, pages 114–134. Butterworth-Heinemann, Newton, MA, USA, 1979.
23. E. Claridge, P. N. Hall, M. Keefe, and J. P. Allen. Shape analysis for classification of malignant melanoma. *J Biomed Eng*, 14:229–234, May 1992.
24. T. K. Lee, D. I. McLean, and M. S. Atkins. Irregularity index: a new border irregularity measure for cutaneous melanocytic lesions. *Med Image Anal*, 7:47–64, Mar 2003.

25. W. A. Yasnoff, J. K. Mui, and J. W. Bacus. Error measures for scene segmentation. *Patt recognit*, 9(4):217–231, 1977.