Particle filtering in the Hough space for instrument tracking

Joan Climent[a], Roberto A Hexsel[b]

[a] *Dept. of Computer Eng., Technical University of Catalonia, Barcelona, Spain*
[b] *Depto. de Informática, Universidade Federal do Paraná, Curitiba, Brazil*

Abstract

In this article we present a real-time tracking system of surgical instruments in laparoscopic operations. We combine CONDENSATION tracking, with the Hough Transform in order to obtain an efficient and accurate tracking. The CONDENSATION algorithm performs well in heavy clutter, and the Hough Transform is robust under illumination changes, occlusion and distractions.

The Hough array is computed using the gradient direction image obtained by means of a Principal Component Analysis. This improves accuracy in the determination of edge orientation and speeds up computation of the Hough Transform.

The experiments on image sequences of actual laparoscopic surgical operations show that the instrument tip is located even in the presence of smoke, occlusions or motion blurring.

Keywords: Instrument tracking, laparoscopic surgery, principal components analysis, Hough Transform, CONDENSATION algorithm.

1. Introduction

Among minimally invasive surgical techniques, laparoscopic surgery has become very popular in recent years. The surgery is performed with the help of an endoscopic camera (a laparoscope) and an assortment of long and thin rigid instruments.

Email address: juan.climent@upc.edu (Joan Climent)

Tubes, called 'trocars', are placed through small incisions and the laparoscope and instruments are inserted and handled through the trocars. The abdominal cavity is inflated with gas to make room for moving the camera and instruments. The surgeon performs the surgical procedure while an assistant holds the laparoscope. This type of procedure avoids surgical opening and reduces the recovery time for the patient. On the other hand, a large number of repetitive gestures are needed which require much effort from the assistant surgeon. Thus, the main drawback of laparoscopy stems from the assistant surgeon having to control the laparoscope manually, which can be rather tiresome. Also, the coordination between the assistant and the surgeon can be difficult, and the images may not be stable. In order to reduce the surgeon's burden and to increase accuracy, robotic laparoscopic systems have been proposed and built [1]. Some examples are the *Zeus* medical robot [2], *Aesop*, *EndoAssist* [3], or the *Da Vinci* system [4].

In order to automatically position the laparoscope, the robot controller has to maintain the spacial coordinates of the instrument's tip. To achieve an accurate positioning of the instruments, several problems must be overcome, such as the inherent complexity of the abdominal cavity scene, time-varying conditions, moving background, presence of reflections, occlusion, presence of smoke, and blood spots on the instruments. Classical attempts to minimize these problems involve the use of colored instruments [5, 6]. This however creates the additional problem of finding artificial colors that do not appear in the scene. The approaches presented in [5, 6] add constraints and further complexity to the laparoscopic procedure: all the instruments have to be marked and this in turn raises sterilization issues.

Rather than using color information, the authors of [7] and [8] employ monochrome patterned marks stamped on the instruments. Line detection algorithms are used to segment the marks, and detect the presence of the instruments. As above, the need to mark all instruments is a major drawback. Active marks have also been used; and the authors of [9] describe how a light emitting diode is placed in the tool tip, projecting laser dots onto the surface of organs. The optical markers are detected in the endoscopic image and help in locating the instrument on the scene.

The authors of [10] present a fast segmentation algorithm of gray regions in color images, based on a joint hue-saturation color feature, which is adapted to gray or metallic instruments. In [11], adaptive algorithms are presented for tracking deforming surfaces from stereo video streams. Another approach, that does not need the presence of artificial marks on the instruments, can be found in [12]: the gradient of the image color components is used, but the 3D position given by the LER camera-holder [13] is needed to determine the tip location. This approach can be problematic when dealing with specular reflections. The removal of reflections on medical images has been addressed in [14], but the real-time constraints are difficult to meet in the context of laparoscopy.

A more recent work dealing with augmented reality for laparoscopic surgical applications can be found in [15]: the authors segment the tool's images by means of edge detector operators, with the tip being detected by color analysis. The main drawback of this technique is the inaccurate edge detection in the presence of blurred edges, which are caused by the movement of the instruments.

In this article we present a system that tracks and detects instruments and can be applied to ordinary surgical instruments without the need of any special marking. The system aggregates four main techniques: (1) the tracking of instruments in the Hough space instead of in the image space; (2) the use of the CONDENSATION algorithm for tracking; (3) the use of Principal Component Analysis for improving accuracy in determining tool orientation; and (4) the use of median filtering to locate the instrument's tip. The prototype, as implemented in a desktop computer, analyzes scenes at a rate of 12 frames per second, fast enough for application in real-time.

The image processing system for locating the instruments employs the Hough Transform (HT) to detect straight lines in the scene. Since the instruments have very structured shapes, the HT is a powerful tool to segment them out from the unstructured organic shapes. The edge orientation image is filtered prior to the Hough array computation by means of a Principal Component Analysis. This improves the accuracy in determining the direction of lines, and minimizes the effects of edges that are blurred by motion. Section 2 presents the line detection procedure.

The technique used for tracking the instruments is the CONDENSATION (Conditional Density Propagation) algorithm. The CONDENSATION algorithm was chosen for tracking because it is known to give good results even in the presence of cluttered background or occlusion [16]. The algorithm is applied to the Hough space, instead of the usual approach of applying it to the image plane. Section 3 describes the tracking algorithm in more detail.

Section 4 presents the method used to locate the instrument tip, using median filtering along the profile of the instrument edge, in the orientation image. Some experimental results are presented in Section 5, showing the system at work on actual sequences of surgical images. Finally, our conclusions are stated in Section 6.

2. Instrument Detection

A robust solution to the detection of straight lines in a scene is the Hough Transform (HT) [17]. The HT yields good results in detecting straight lines and other shapes even in the presence of noise and occlusion. Since the surgical instruments mostly have very structured shapes – straight lines – the HT is an appropriate tool to pick them out of a scene. The HT has been widely used in medical imaging applications, including laparoscopic instruments tracking [5, 12, 18]. A good survey on the Hough Transform can be found in [19].

The normal parameterization of the HT (Equation 1) is used to extract the most significant straight lines in the scene.

$$x \cos \alpha + y \sin \alpha = r \tag{1}$$

where r and α are the length and orientation of the normal vector to the line, from the image origin. Each straight line is uniquely defined by r and α, and for every point in the original image (x, y), it is possible to create a mapping from feature to the parametric space.

If we divide the parametric space into a number of discrete accumulator cells, we can collect 'votes' in the (r, α) space from each data point in the (x, y) space. Peaks in (r, α) space define the equations of co-linear points in the (x, y) space.

Hough array size was fixed at 128 bins for r and 360 bins for α. The maximum orientation error using the PCA approach, described in Section 2, is 0.94^o. Thus, the resolution for α was fixed at 1^o per bin. We use the maximum resolution for α because angular errors in the determining the line orientation cause large errors in the localization of the tip, specially when it is away from the image center. The upper bound for r, $r_{\max}$, can easily be deducted from the image size. Resolution in distance has been fixed to $0.5r_{\max}$, that is, one bin for every two distance-pixels, since two pixels is an admissible distortion when locating tip position – a human would probably not be able to locate the tip with that resolution – and the error is bounded and does not depend on the position of the tip on the scene. This space sampling is a good tradeoff between efficiency and accuracy. Efficiency is directly related to the number of particles needed, and this in turn depends on the array size. Often, when using the rough Hough table, discretization problems arise because every sample must be counted in just one bin, and often these counts are spread onto neighboring bins. Classic solutions involve smoothing the array; we do not apply any smoothing to the Hough table because the factored sampling – described in Section 3 – already does it, and hence the Hough table is used in its rough form.

For every pixel in the image, the gradient direction must be determined prior to the HT computation. Since real-time operation is mandatory in robotic laparoscopy, we can make use of orientation information, and simplify the computation of the HT considerably, as shown in [20]. Each pixel then only votes for one bin in the accumulator, because the orientation is fixed by the edge direction. However, the local gradient direction is rarely accurate enough to allow a reliable implementation of this scheme. There are a number of ways to obtain more accurate edge directional information from local gradients, such as the edge direction histogram, which is based on a simple edge detection algorithm [21], or directional fields which are based on Principal Component Analysis [22]. To derive the orientation map we use the Principal Component Analysis (PCA) technique, commonly used in fingerprint image processing but never used before in automatic instrument tracking. The analysis is then carried out by sampling a small neighborhood of each pixel and determining the direction in which the greatest variability in the image intensity is observed. This allows the detection of lines with blurred edges, caused by instrument motion. There are other applications in which this technique might be applied, such as video compression [23], and mark tracking in car crash tests [24].

PCA computes a new orthogonal base, given a multidimensional data set, such that the variance of the projection on one of the axes of this new base is maximal, while the projection on the other is minimal. When applying PCA to the auto-covariance matrix of the gradient vectors, it provides the 2-dimensional Gaussian joint probability density function of these vectors. From this function, the main direction of the gradients can be calculated.

For any subregion of the image, it can be shown that the direction $\theta_{\max}$ for which the weighted mean squared gradient of image intensity is a maximum is given by

$$\theta_{\max} = \frac{1}{2} \arctan \frac{2\,\overline{g_x\, g_y}}{\overline{g_x^2} - \overline{g_y^2}} \tag{2}$$

where g_x is the image gradient in the x direction; g_y is the image gradient in the y direction; and $\overline{g_x^2}$, $\overline{g_y^2}$ are the weighted means of these quantities over the image subregion. Here we take the weighted means by applying a 2-dimensional Gaussian filter.

Figure 1 illustrates the technique. Figure 1.b shows the edge orientation image obtained from applying the PCA technique to the original image in Figure 1.a. The Hough table in the (r, α) space is shown in Figure 1.c, and Figure 1.d contains its 3D representation. Peaks in the (r, α) space correspond to the presence of straight lines in the scene. In our previous work [25], the maximum peak was selected as the longest straight line in the image.

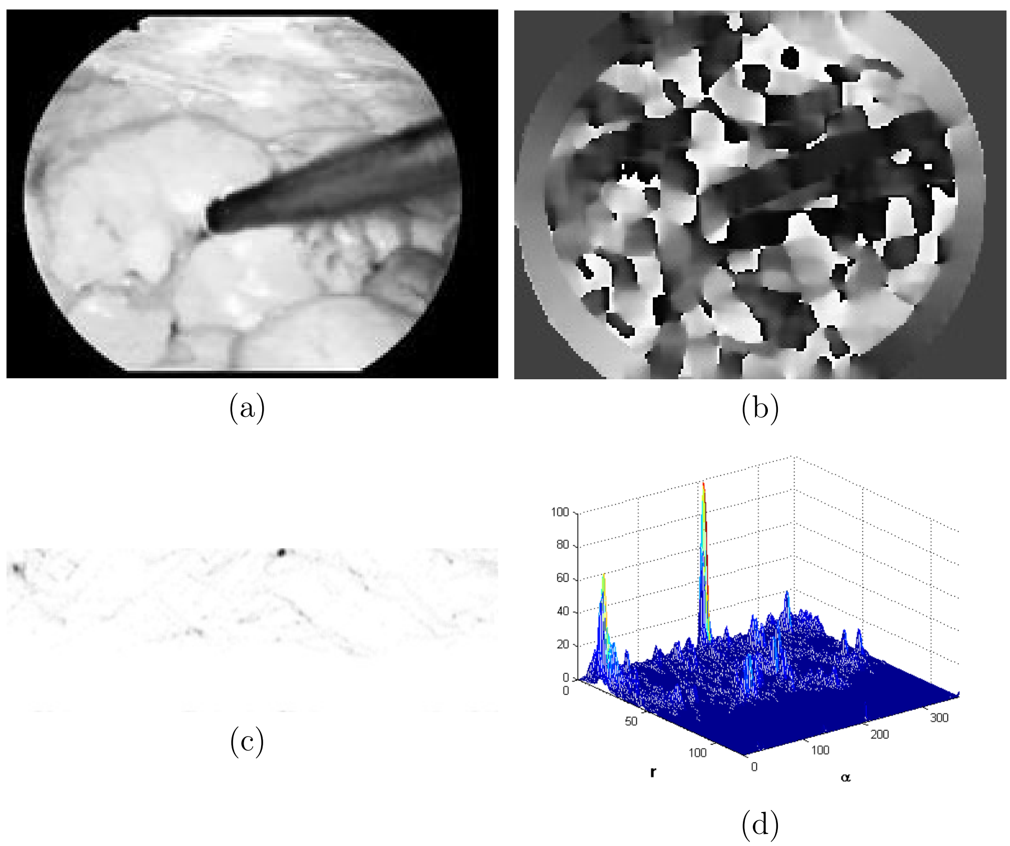


Figure 1: (a) Original image (b) edge orientation image (c) Hough image (d) 3D representation of Hough image

Unfortunately, the longest line in the scene does not always correspond to the instrument. It may happen that a muscle, or some part of other organic tissue also presents a structured shape, and the selection of the longest line in the scene does not guarantee the detection of the surgical instrument.

We performed experiments with a set of 512 images, extracted from four actual video sequences to test the quality of the detection system, and to assess its accuracy. Table 1 shows the position (in terms of line length) corresponding to the instrument edge. In only 81% of cases, the longest line detected identifies the correct instrument. The remainder are false detections corresponding to random alignments or to other instruments which are not the one being tracked.

Longest line in scene	% of correct identification
1st	81.44
2nd	8.00
3rd	4.10
4th	3.91
5th	1.95
6th	0.60
7th	0.00

Table 1: Percentage of correct instrument identifications

3. Instrument Tracking

The detection results from static images, as presented above, aren't accurate enough for a real application. The peaks in the Hough table identify the straight lines present in the scene, which may not necessarily be the object of interest – which was somewhat problematic in our previous work [25]. To improve the accuracy, and achieve reliable instrument detection, the information contained in a sequence of images must be taken into account: the instruments must be tracked according to their position in past frames. The Hough array must now be viewed as an approximation to the Probabilistic Hough Transform [12, 26], with each array element providing an estimate of the probability that a line with parameters within a given range is present in the image. Peaks in the accumulator array are tracked, taking advantage of the insensitivity of HT to illumination changes and occlusion.

The estimation of an object's location is usually achieved through a two-phase process consisting of a prediction phase and its correction by observation. The Kalman filter [27] and the CONDENSATION algorithm [25] are the commonly applied estimators. Both are probability based and estimate the state of a process. The Kalman filter aims at predicting, while minimizing the mean of the squared error, so that the new model constructed from what was already known is the most probably correct. The CONDENSATION algorithm is a particle filter which combines stochastic methods with learned object dynamics model to estimate the next state.

Particle filters are a popular class of numerical methods for the solution of optimal estimation problems in non-linear non-Gaussian scenarios. Since the probability density function defined by the Hough array is surely non-Gaussian (see Figure 1.d), the CONDENSATION algorithm is a better choice than Kalman filtering for our application. In comparison with standard approximation methods, such as the Extended Kalman Filter (EKF), particle methods have the advantage of not relying on any local linearization technique, nor any crude functional approximation. Usually, EKF is the technique applied where both the dynamic and observation models are non-linear but the noise is additive and Gaussian. For severe nonlinearities, the EKF can be very unstable and performs poorly.

Another difference between these estimators is that the Kalman filter assumes Gaussian densities, which are unimodal and thus cannot represent simultaneous hypotheses of the location of multiple objects, whereas CONDENSATION can be used for tracking multiple targets.

The technique of particle filtering is popular in tracking applications [16, 28–31], being effective in estimating the state of dynamic systems from sensor information. The key idea is to represent probability densities by set of samples. A number of particles, or hypotheses, are considered, each providing an estimate of state parameters. These are propagated to give an updated estimate in the next frame.

As a result, particle filtering provides a real-time estimation of the state of non-linear, non-Gaussian dynamic systems. This technique was originally developed to effectively track objects in cluttered scenes [16, 30]. The state of a tracked object at time t is described by a vector X_t, while the vector Y_t represents all samples of observations $\{y_1, y_2, \cdots, y_t\}$. The posterior density $P(X_t|Y_t)$ and the observation density $P(Y_t|X_t)$ are often non-Gaussian.

The particle filters are essentially a sample-based variant of Bayes filters, that approximate the probability density distribution by a weighted sample set $S_t = \{(x_t^{(i)}, w_t^{(i)}) \mid i = 1 \cdots N_p\}$. Each sample $x_t^{(i)}$ represents an hypothetical state of the object, and $w_t^{(i)}$ represents the corresponding discrete sampling probability of the sample $x_t^{(i)}$ such that $\sum_{i=1}^{N_p} w_t^{(i)} = 1$.

In its basic form, the particle filtering performs the recursive Bayes filter according to a sampling procedure, often called sequential importance sampling with resampling (SISR) [31]. The iterative evolution of the sample set is described by propagating each sample according to a system model. Each sample element in the set is weighted in terms of the corresponding observations, and N_p samples are drawn with replacement by choosing a particular sample with posterior probability $w_t^{(i)} = P(Y_t | X_t = x_t^{(i)})$.

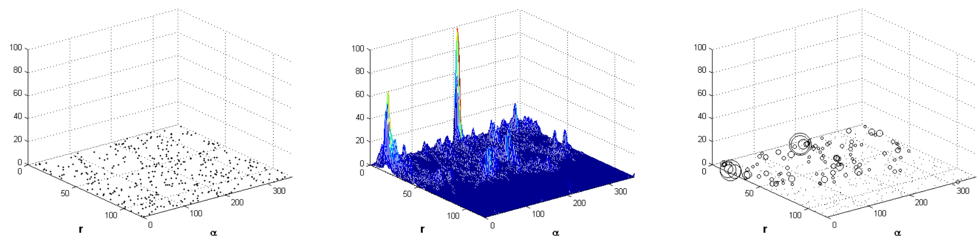
Since it models uncertainty, as posterior probability density, particle filtering provides a robust tracking framework suitable for robotic laparoscopy. Rather than attempting to fit a specific equation to the observed sensory data, it uses the N_p weighted samples to approximate the state described by the observed data. When applied to tracking, each sample represents the *state* of the object being tracked, *e.g.* its velocity and location in the Hough space. Given a randomly sampled state S_t at time t, a prediction of a new state S_{t+1} at time $t+1$ is made using a predictive model.

The Hough array is employed as the observation density in order to weight the sample set. For each sample in the set, a weight is computed according to the value of the corresponding Hough array location. Iterations of factored sampling applied to a Hough array sequence, that correspond to an image sequence, allow sampling to be drawn from the neighboring area of the Hough array locations of higher values, thus increasing the tracking accuracy.

An iteration of the tracking algorithm is illustrated in Figure 2. Figure 2.a shows a new sample set. We measure the weights for each sample location using the Hough table shown in Figure 2.b, and Figure 2.c shows the weighted sample set where the circle size is proportional to particle weight.

The old sample set is defined as $(S_{t-1}^{(i)}, w_{t-1}^{(i)}, C_{t-1}^{(i)})$, for $i = 1 \cdots N$, where, at time $t-1$, $S_{t-1}^{(i)}$ are the samples in the Hough array; $w_{t-1}^{(i)}$ are their associated weights; and $C_{t-1}^{(i)}$ are cumulative weights (probability). An iteration of our tracking algorithm constructs a new sample-set $(S_t^{(i)}, w_t^{(i)}, C_t^{(i)})$, for $i = 1 \cdots N$, using the following steps:

1. Generate a sample S'_t by
   (a) Select a random number $r \in [0, 1]$ using an uniform distribution;
   (b) Find the smallest j for which $C_{t-1}^{(j)} \geq r$ using dicotomic search; and



(a) (b) (c)

Figure 2: (a) New sample set (b) measurement function (c) weighted sample set

   (c) Setting $S_t^{\prime(n)} = S_{t-1}^{(j)}$;

2. Predict by sampling from $P(X_t|X_{t-1} = S_t^{\prime(n)})$ to choose each $S_t^{(n)}$. A dynamic stochastic model is applied for this process, namely the second order auto-regressive process (ARP) [16]. Such a model expresses the state X_t at time t as a linear combination of the previous two states. In order to account for noise and deviations from the model, a random Gaussian element is added to each prediction;

3. Measure weights of the new sample positions in terms of feature Y_t, that is $w_t^{(n)} = P(Y_t|X_t = S_t^{(n)})$. Weights are computed using

$$w_t^{(n)} = \frac{H(S_t^{(n)})}{\sum_i H(i)}$$

   where $H(S_t^{(n)})$ is the value in the Hough array corresponding to the sample $S_t^{(n)}$; and the normalization value $\sum_i H(i)$ is the sum of all the accumulators of the Hough array;

4. Calculate the cumulative probability $C_t^{(n)}$. The weights are normalized so that $\sum w_t^{(n)} = 1$, and cumulative probabilities are $C_t^{(0)} = 0$ and $C_t^{(n)} = C_t^{(n-1)} + w_t^{(n)}$, for $n = 1 \cdots N$.

Results of the tracking process in the Hough space are shown in Figure 3. The factored sampling technique yields accurate sampling during tracking as shown in Figure 3.a. Figure 3.b shows the weights calculated for the sample points in Figure 3.a, overlaid on the Hough image. The circle sizes represent the value of the corresponding weights. Figure 3.c shows the two main straight lines located in the Hough space, overlaid on the original image.

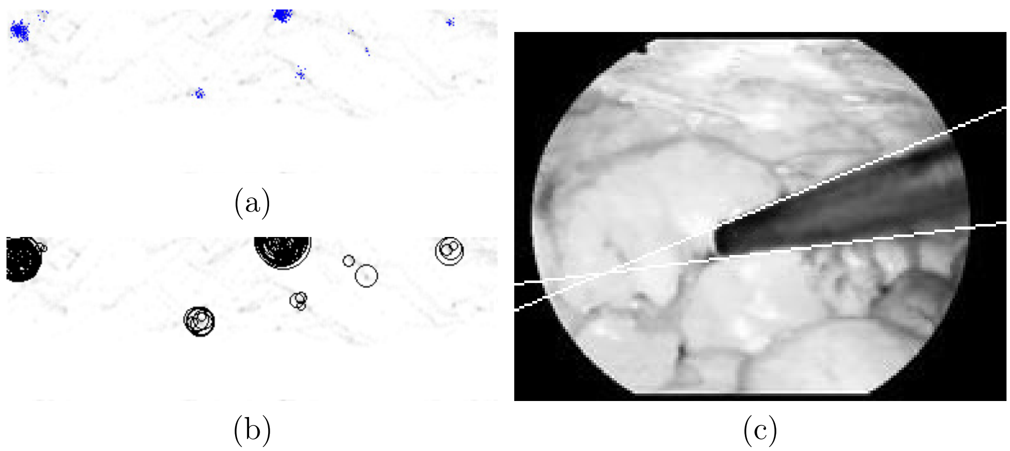


Figure 3: (a) Sample set (b) weighted samples (c) two main straight lines in the scene

In the Hough table (Figure 1.d), there are two peaks, corresponding to the two instrument edges. This suggests that a more complex dynamic model could be defined, one that considers the instrument position as a peak-pair in the Hough table. Defining the dynamics of this model would not be trivial since the straight lines are not parallel; they are related by a point conic, and this relation changes dynamically with perspective because of the three-dimensional movements of the instruments.

Not surprisingly, increased model complexity results in decreased performance. In this application, the cost might not be justified by increased accuracy in tracking. The main reason being that real-time control is mandatory, and the computational cost of the CONDENSATION algorithm increases exponentially with the number of state variables. Yet another problem arises in the initialization stage, when all possible peak-pairs are candidates to represent an instrument. This would either make the initialization process intractable, or involve human assistance so that the tracking process can start up in a reasonable time.

4. Tip Localization

Once one of the instrument edges has been detected, finding the position of it's tip is an easy task: the location of the tip is simply determined by a loss of continuity in the orientation along the instrument edge. Pixels in the orientation image are traced along the line profile until their individual orientations present a significant change with respect to the line orientation.

For every pixel in the line, we compute the difference between its orientation and the line orientation that has already been determined in the HT computation. These differences are considered as the orientation error.

Figure 4 contains an example of this computation. Figure 4.b shows the orientation of the pixels along the line profile shown in Figure 4.a. Only a segment of this line corresponds to the instrument. It can be seen that the continuity in orientation decreases sharply at the ends of the instrument. Figure 4.c shows the deviation between the computed orientation of the pixels along the line profile, and the actual edge orientation.

A useful property of the HT is that the pixels which lie on the line are not necessarily contiguous. This means that there is no need of continuity for detecting a straight line, thus making the line detection process robust to partial occlusions of the instrument edges caused by smoke, blood spots or noise. However, it can also produce misleading results when, by chance, some pixels happen to be aligned. Pixels which are not a part of the instrument, but noise pixels, might be present in the image with the same line parameters as non-noise pixels. For this reason, the line profile must be filtered to suppress the orientation values corresponding to partial occlusions or randomly aligned pixels. This effect is shown in Figure 4.c. We consider these pixels as isolated noise, and use a median filter to suppress their effect on locating the tip.

Linear filters have undesirable blurring effects. In [25], average filtering is used to suppress noise, but the result thus obtained is degraded by the effect of noise values, as should be expected when using linear filtering. Median filters are robust, and can furthermore eliminate the effect of noise values with extremely large magnitudes. Figure 4.d shows the result of median filtering the profile in Figure 4.c. Note how the large contrasts along the orientation error profile were suppressed. A simple derivative operator is used to detect the end of the tool and to locate the instrument tip, as shown in Figure 4.f.

5. Results

This section presents the results of applying the techniques described above to actual video sequences from laparoscopic interventions. A 2GHz Pentium PC was used to perform the computations. The images were acquired at a resolution of 768x576 pixels, and this resolution was reduced to 384x288 pixels so that real-time operation could be achieved. The number of particles was fixed to 400, and that yields a computation time under 80ms

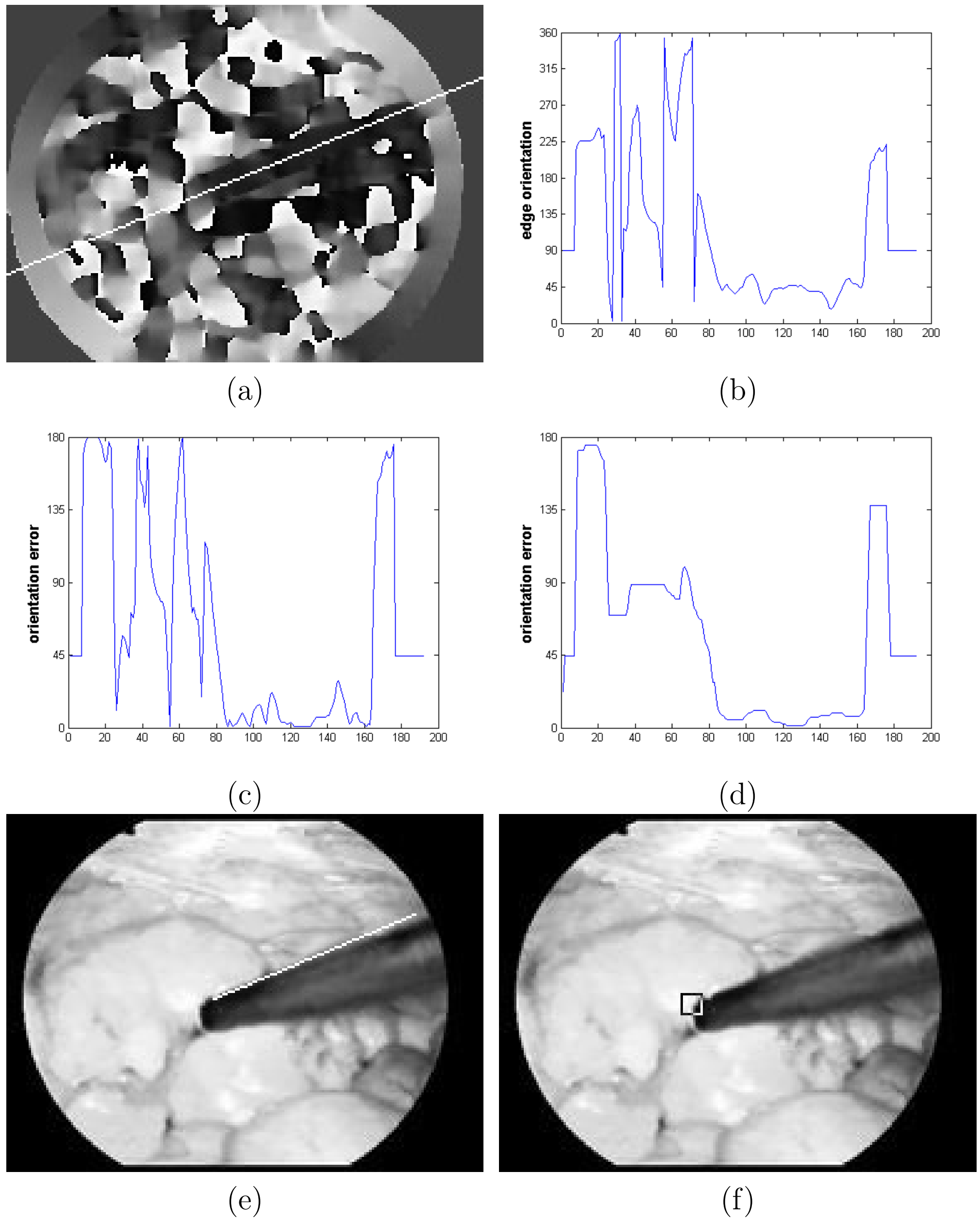


Figure 4: (a) Orientation image (b) line profile (c) deviation between computed and actual orientation (d) median filtering of deviation (e) line segment (f) tip location

per frame (a rate of 12 frames per second) which is fast enough for real interventions.

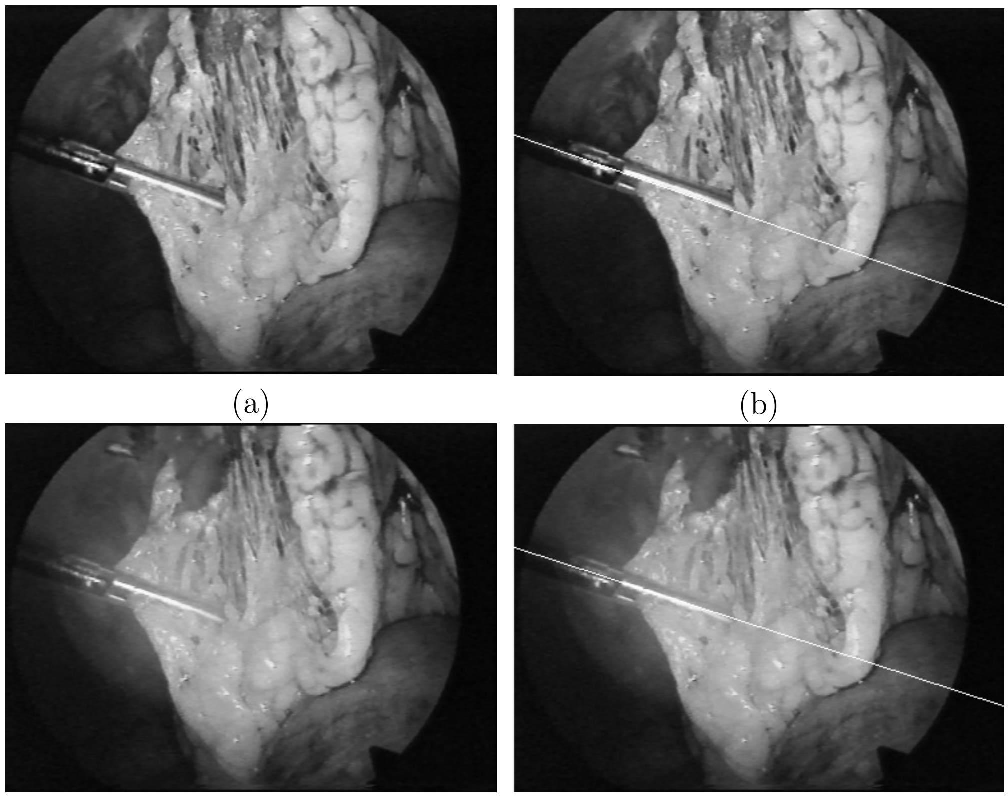
While developing the system, we used images from a real laparoscope "operating" on an artificial testbed, where it was possible to exert complete control of the instruments' actions. While testing the tool tip location algorithm, artificial color marks were placed on the instrument tip. The locations computed by our system were compared against those obtained by straightforward color segmentation, which were used as ground truth[1]. The tests for tip localization with the testbed used 14280 frames, from 22 different sequences. The test results indicate that the mean error between the actual mark position and that determined by our location algorithm is 2.88 pixels, with a maximum error of 11.18 pixels. The sequences used in these tests include reflections and motion blurring. Unfortunately, an actual moving organic background cannot be emulated in our testbed. Realistic smoke is also difficult to emulate.

For more realistic tests, the system was applied to sequences of real interventions on living people. Fifty sequences of 512 frames were used, and some of these sequences show proceedings with two or more instruments. From these, 88 sequences with the tracking of selected different instruments were selected to test the final version of the system. In these sequences, the tip location in 1920 frames was obtained manually and this information was used as ground truth. The location results from the actual procedures were very similar to those obtained on the testbed. The mean error is 3.01 pixels and the maximum error is 11.66 pixels.

We believe that the margin of error is perfectly admissible, for two reasons. First, the precision achieved by a human locating the tool tip manually is hardly better than 2 or 3 pixels. Second, our main objective is visual servoing for a laparoscope, and thus the camera must be kept focused on the region around the tip of the instruments, which is where the surgeons focus their interest. An error of a few pixels in determining where the tip is does not interfere with the surgeon's work. Obviously, the target position has to be filtered before being sent to the robot controller, to avoid image flicking. The robot's control system is outside of the scope of this article.

Figure 5 shows the instrument detection in the presence of smoke. Figure 5.b shows the detection of the instrument while the cavity is 'clean' of smoke,

[1] Active marks, such as LEDs, are not suitable because of saturation effects in the laparoscope.



(a) (b)

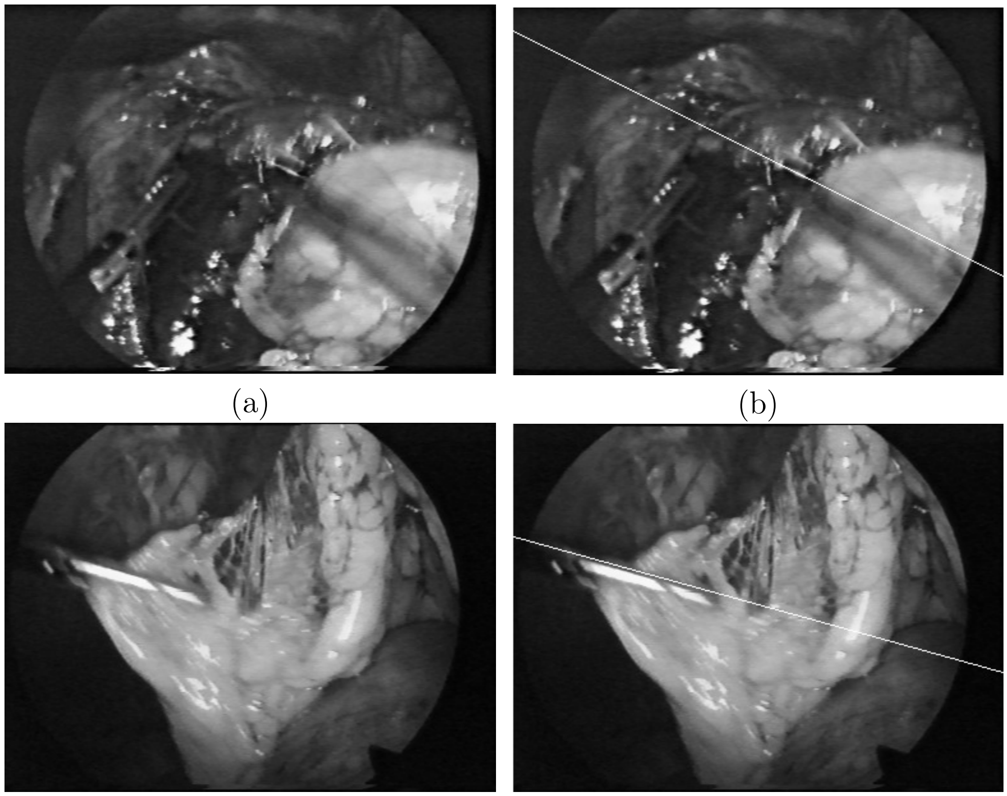
(c) (d)

Figure 5: (a) Original image (b) instrument detection (c) image with presence of smoke (d) instrument detection

and Figure 5.d shows the detection of the same instrument once smoke is formed. Figures 5.a and 5.c are quite similar except for the smoke; there are seven video frames between the two, which were captured before and after the formation of smoke.

Figure 6 shows the instrument detection in images blurred because of motion. As can be seen in Figures 6.b and 6.d, the instrument is accurately detected with images under heavy motion blurring. Furthermore, the specular reflections on the metallic surface do not affect the detection of the instrument.

The CONDENSATION algorithm has the capability to maintain multiple hypotheses, and once we developed a tracker for one instrument, we obtain a multi-instrument tracker without any additional effort. The ability of the



(a) (b) (c) (d)

Figure 6: (a) Motion blurring (b) instrument detection (c) motion blurring (d) instrument detection

CONDENSATION algorithm to represent multi-modal distributions was tested using image sequences with the presence of two instruments. Figure 7.a shows an image with two instruments. Four edges are easily identified as peaks in Hough table shown in Figure 7.b. Figure 7.c shows the sample set, organized into four clusters around the peaks corresponding to the straight lines shown in Figure 7.d.

Different hypotheses are continuously being considered by the CONDENSATION algorithm while physically tracking the one selected by the surgeon. The surgeon can choose to track a given instrument by manually switching among those present in the scene.

Figures 8 shows the result of the tracking system applied to one of the selected sequences, which displays specular reflections and perspective distor-

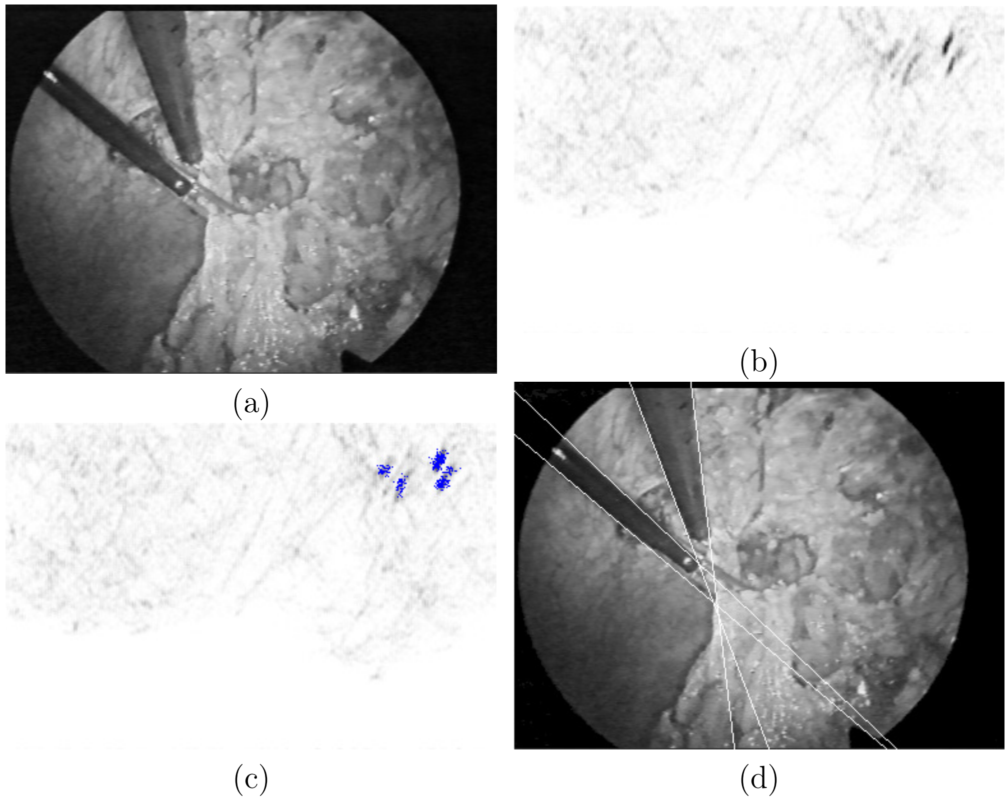


Figure 7: (a) Original image (b) Hough array (c) sample set (d) four main tracked lines.

tion caused by 3D movement of the instrument. The instrument detection, and the tip localization are shown together with the original image sequence.

6. Conclusions

This article describes a set of techniques that can be used to implement a robust and accurate tracking system for surgical instruments in robotized laparoscopic procedures. The instrument locations thus computed are immune to partial occlusions, noise and motion blurring. The contributions in this article are listed below.

The first contribution is the use of the CONDENSATION algorithm for tracking the surgical instruments in the Hough space instead of tracking in the image space. This technique is effective in cluttered backgrounds, and performs well in estimating non-Gaussian and non-linear dynamic systems, as is the case with the dynamics of surgical instruments. Tracking an object

(a) (b) (c)

Figure 8: Results of image sequence processing

through an image sequence typically involves following features detected in one image to their new positions in the next. To perform the tracking with a Hough array sequence has the advantage of making the tracking robust to occlusion and illumination changes.

As a second contribution, the error in determining the tool direction is minimized by a Principal Component Analysis of the gradient orientation prior to the Hough Transform computation. This technique is widely used in fingerprint recognition but has never been used before for instrument tracking in laparoscopy.

Finally, a median filtering has been applied for the localization of the instrument tip, to attain higher accuracy than that achievable with linear filtering.

Last but not least, a further advantage of our the system is that it is can be used with standard surgical tools, obviating the need for marking them with color or stripes.

Our implementation on an ordinary desktop computer can analyze scenes at a rate of 12 fps, which is a reasonable frame rate for real usage.

Acknowledgments

This research was partially supported by *Consolider Ingenio* 2010, project CSD2007-00018, and CICYT project DPI2010-17112. One of the authors (RAH) was supported by grant BEX 1656-09-0, from CAPES, Ministry of Education, Brazil.

References

[1] R. H. Taylor, J. Funda, D. LaRose, M. Treat, A telerobotic system for augmentation of endoscopic surgery, in: Proc IEEE Int Conf on Engineering in Medicine and Biology, Vol. 14, 1992, pp. 1054–1056.

[2] H. Reichenspurner, R. Damanio, M. Mack, D. Boehm, H. Gulbins, C. Detter, B. Meiser, R. Ellgass, B. Reichart, Use of the voice-controlled and computer-assisted surgical system Zeus for endoscopic coronary artery bypass grafting, Journal of Thoracic and Cardiovascular Surgery 118 (1) (1999) 11–16.

[3] P. Ballester, Y. Jain, K. R. Haylett, R. F. McCloy, Comparison of task performance of robotic camera holders EndoAssist and Aesop, in: Proc

15th Int Congress and Exhibition on Computer Assisted Radiology and Surgery, 2001, pp. 1100–1103.

[4] G. S. Guthart, J. K. Salisbury Jr, The Intuitive™ telesurgery system: overview and application, in: Proc IEEE-ICRA Int Conf on Robotics and Automation, 2000, pp. 618–621.

[5] G. Wei, K. Arbter, G. Hirzinger, Real-time visual servoing for laparoscopic surgery, IEEE Engineering in Medicine and Biology Magazine 16 (1) (1997) 40–45.

[6] Y. F. Wang, D. R. Uecker, Y. Wang, A new framework for vision enabled and robotically-assisted minimally invasive surgery, Computerized Medical Imaging and Graphics 22 (1998) 429–437.

[7] A. Casals, J. Amat, D. Prats, E. Laporte, Vision guided robotic system for laparoscopic surgery, in: Proc Int Conf Advanced Robots, 1995, pp. 33–36.

[8] O. Tonet, T. U. Ramesh, G. Megali, P. Dario, Tracking endoscopic instruments without a localizer: a shape-analysis-based approach, in: Studies in Health Technology and Informatics, Vol. 119, IOS Press, 2005, pp. 544–549.

[9] A. Krupa, J. Gangloff, C. Doignon, M. de Mathelin, G. Morel, J. Leroy, L. Soler, J. Marescaux, Autonomous 3-D positioning of surgical instruments in robotized laparoscopic surgery using visual servoing, IEEE Trans on Robotics and Automation 19 (5) (2003) 842–853.

[10] C. Doignon, P. Graebling, M. de Mathelin, Real-time segmentation of surgical instruments inside the abdominal cavity using a joint hue saturation color feature, Real-Time Imaging 11 (5-6) (2005) 429–442. doi:10.1016/j.rti.2005.06.008.

[11] D. Burschkaa, J. J. Corsoa, M. Dewana, G. D. Hagera, W. Laua, L. Ming, H. Lina, P. Marayonga, N. Rameya, Navigating inner space: 3D assistance for minimally invasive surgery, in: Workshop Advances in Robot Vision, IEEE/RSJ Int Conf on Intelligent Robots and Systems, 2004, pp. 67–78.

[12] S. Voros, J.-A. Long, P. Cinquin, Automatic detection of instruments in laparoscopic images: a first step towards high-level command of robotic endoscopic holders, The Int Journal of Robotics Research 26 (11-12) (2007) 1173–1190.

[13] P. J. Berkelman, P. Cinquin, J. Troccaz, J.-M. Ayoubi, C. Létoublon, Development of a compact cable-driven laparoscopic endoscope manipulator, in: MICCAI '02: Proc 5th Int Conf on Medical Image Computing and Computer-Assisted Intervention, Springer-Verlag, 2002, pp. 17–24.

[14] M. Gröger, W. Sepp, T. Ortmaier, G. Hirzinger, Reconstruction of image structure in presence of specular reflections, in: Proc of the 23rd DAGM-Symposium on Pattern Recognition, Springer-Verlag, 2001, pp. 53–60.

[15] A. M. Cano, F. Gayá, P. Lamata, P. Sánchez-González, E. J. Gómez, Laparoscopic tool tracking method for augmented reality surgical applications, in: ISBMS '08: Proc of the 4th Int Symp on Biomedical Simulation, Springer-Verlag, 2008, pp. 191–196. doi:10.1007/978-3-540-70521-5_21.

[16] M. Isard, A. Blake, CONDENSATION—conditional density propagation for visual tracking, Int Journal of Computer Vision 29 (1) (1998) 5–28.

[17] R. O. Duda, P. E. Hart, Use of the Hough Transform to detect lines and curves in pictures, Communications of the ACM 15 (1972) 11–15.

[18] B. F. Allen, F. Kasper, G. Nataneli, E. Dutson, P. Faloutsos, Visual tracking of laparoscopic instruments in standard training environments, in: Medicine Meets Virtual Reality 18 (MMVR), Vol. 163 of Studies in Health Technology and Informatics, 2011, pp. 11–17. doi:10.3233/978-1-60750-706-2-11.

[19] V. F. Leavers, Which Hough transform?, CVGIP: Image Understanding 58 (2) (1993) 250–264. doi:10.1006/ciun.1993.1041.

[20] D. H. Ballard, Generalizing the Hough Transform to detect arbitrary shapes, Pattern Recognition 13 (1981) 111–122.

[21] A. K. Jain, A. Vailaya, Image retrieval using color and shape, Pattern Recognition 29 (1996) 1233–1244.

[22] A. M. Bazen, S. H. Gerez, Systematic methods for the computation of the directional fields and singular points of fingerprints, IEEE Trans on Pattern Analysis and Machine Intelligence 24 (7) (2002) 905–919. doi:10.1109/TPAMI.2002.1017618.

[23] H. Yu, Z. Lin, F. Pan, Applications and improvement of H.264 in medical video compression, IEEE Trans on Circuits and Systems 52 (12) (2005) 2707–2716. doi:10.1109/TCSI.2005.857869.

[24] J. Bigun, T. Bigun, K. Nilsson, Recognition by symmetry derivatives and the generalized structure tensor, IEEE Trans on Pattern Analysis and Machine Intelligence 26 (12) (2004) 1590–1605. doi:10.1109/TPAMI.2004.126.

[25] J. Climent, P. Mares, Automatic instrument localization in laparoscopic surgery, Electronic Letters on Computer Vision and Image Analysis 4 (1) (2004) 21–31.

[26] R. S. Stephens, Probabilistic approach to the Hough Transform, Image and Vision Computing 9 (1) (1991) 66–71. doi:10.1016/0262-8856(91)90051-P.

[27] R. E. Kalman, A new approach to linear filtering and prediction problems, Journal of Basic Engineering 82 (1) (1960) 35–45.

[28] S. Arulampalam, S. Maskell, N. Gordon, T. Clapp, A tutorial on particle filters for on-line non-linear/non-Gaussian Bayesian tracking, IEEE Trans on Signal Processing 50 (2) (2001) 174–188.

[29] C. Kwok, D. Fox, M. Meila, Real-time particle filters, Proc of the IEEE 92 (3) (2004) 469–484.

[30] M. Isard, A. Blake, Contour tracking by stochastic propagation of conditional density, in: Proc European Conf Computer Vision, 1996, pp. 343–356.

[31] A. Doucet, N. D. Freitas, N. Gordon, Sequential Monte Carlo methods in practice, Springer-Verlag, 2001.