Proceedings of the 2005 IEEE
International Conference on Robotics and Automation
Barcelona, Spain, April 2005

Robust Real-Time Instrument Tracking in Ultrasound Images for Visual Servoing

T. Ortmaier *
German Aerospace Center
Institute of Robotics and Mechatronics
D - 82234 Wessling, Germany
Email: Tobias.Ortmaier@dlr.de

M.-A. Vitrani, G. Morel, and S. Pinault
Robotics Lab of Paris (LRP) CNRS
University of Paris 6, 18, Route du Panorama - BP 61
92 265 Fontenay aux Roses Cedex, France
{*vitrani, morel*}*@robot.jussieu.fr*

***Abstract*— Minimally invasive surgery in combination with ultrasound (US) imaging imposes high demands on the surgeon's hand-eye-coordination capabilities. A possible solution to reduce these requirements is minimally invasive robotic surgery in which the instrument is guided by visual servoing towards the goal defined by the surgeon in the US image. This approach requires robust tracking of the instrument in the US image sequences which is known to be difficult due to poor image quality. This paper presents computer vision algorithms and results of visual servoing experiments.**

Adaptive thresholding according to Otsu's method allows to cope with large intensity variations of the instrument echo. Subsequently applied morphological operations suppress noise and echo artefacts. A fast labelling algorithm based on run length coding allows for realtime labelling of the regions. A heuristic exploiting region size and region velocity helps to overcome ambiguities. The overall computation time is less than 10 ms per frame on a standard PC. The tracking algorithm requires no information about texture and shape which are known to be very unreliable in US image sequences. Experimental results for different instrument materials (polyvinyl chloride, polyurethane, nylon, and plexiglas) are given, illustrating the performance of the proposed approach: when chosing the appropriate material the reconstructed trajectories are smooth and only few outliers occur. As a consequence, the visual servoing loop showed to be robust and stable.

***Index Terms*— ultrasound tracking, visual servoing, minimally invasive surgery**

I. Introduction

Ultrasound (US) is an important imaging modality for medical examinations and surgical interventions, as it is cheap, harmless, and allows for real-time image acquisitions. Minimally invasive surgery (MIS) is an operation technique in which the surgeon works with long instruments through small holes. This reduces pain and trauma due to smaller incisions as compared to open surgery. Therefore, a combination of US imaging and MIS leads to a very gentle form of surgery. Unfortunately, this technique imposes high demands on the surgeon, as hand-eye-coordination becomes very difficult. Therefore, this operation technique is limited to simple interventions, such as needle punctures. A possible solution to enlarge the application field of this approach is minimally invasive robotic surgery (MIRS) in combination with visual servoing: the surgeon chooses the point of interest in the image plane and the robot moves the instrument towards this goal. The setup is sketched in Fig. 1. The proposed approach requires real-time tracking of surgical instruments in US images, which is difficult due to speckles, artefacts, and intensity variations.

A lot of work on tracking and detecting contours in US images can be found in the literature, see for example [1]–[3]. Most of the algorithms work with active contours allowing to deal with deformable structures (e.g. tissue). Usually, these algorithms require the minimization of an energy function which involves often a gradient based approach as no analytical solution exists. Unfortunately, a gradient based approach contradicts realtime capabilities as the convergence of the optimization within a certain number of iterations cannot be guaranteed.

Only few articles can be found dealing with real-time tracking in US images. Most notable is [4], where different techniques for real-time tracking are compared: correlation and sum of squared differences (SSD) yield good results but are not applicable in the case presented here, as the echo shape and texture change significantly. The same holds for other tracking techniques based on texture information. The Star and the Star/Kalman [4] approaches rely on the detection of edges along rays emanating from a point interior the instrument echo. This approach is sensitive to ambiguities due to edges arising from speckles or echo artefacts. In the work presented here a robust real-time approach is described which is able to cope with large intensity variations and requires neither texture nor geometry information.

The reminder of the article is organized as follows: the next section describes the experimental setup in detail. In Sect. III, the robust vision algorithms developed for real-time instrument tracking in US image sequences are given. Section IV presents experimental results gained during visual servoing experiments, validating the chosen approach. The last section concludes this article and presents directions for further research.

II. Experimental Setup

The experimental setup is shown in Fig. 2. As robot MC2E (french acronym for *compact manipulator for endo-*

*This research was carried out when T. Ortmaier was with the LRP.

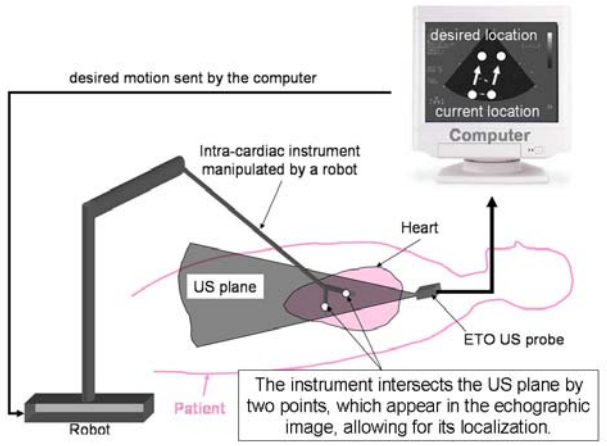
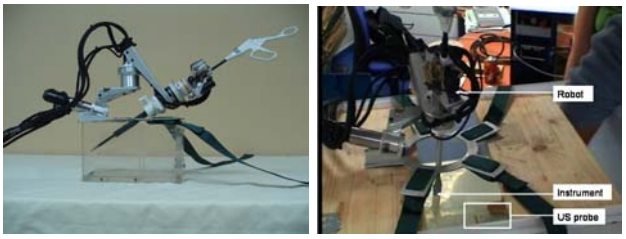


Fig. 1. Visual servoing concept.

scopic surgery, developed at the Laboratoire de Robotique de Paris) is used [5]. This robot is especially suited for minimally invasive robotic surgery applications and provides, with its spherical structure, four degrees of freedom (DoFs) at the instrument tip. The US probe is manually kept under water.



(a) MC2E robot (b) Experimental setup

Fig. 2. Overview of the experimental site.

The robot is equipped with custom made instruments (materials: polyvinyl chloride (PVC), polyurethane (PUR), nylon, and plexiglas; see Fig. 3 for PVC instrument) which are moved in a plastic box filled with water. The robot is placed such that the instrument axes intersect with the ultrasound plane. The instrument axes being parallel to the image plane is a special case and is not considered here. The instrument shape is similar to a pair of forceps, therefore two distinct echoes (thereafter also refered as regions or blobs) can be seen in the US image (see Fig. 5). The goal of the image processing as presented in Sect. III is to reliably track these two echoes. Their position is used later to close the visual servoing loop.

As ultrasound device an HP SONOS 5500 is used. The video output is connected to the BT878 frame grabber card of the image processing computer (Linux PC, P4 at 2.8 GHz with 1 GByte of RAM). Images are captured at frame rate (40 ms, 25 Hz) and converted into gray images

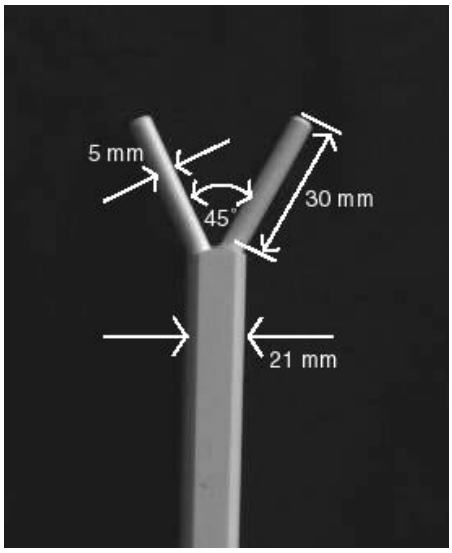


Fig. 3. Instrument made of PVC.

with pixel intensity between 0 and 255 and a resolution of 384×288 pixels. The vision algorithms described in Sect. III are applied to calculate the center of gravity

$$\boldsymbol{p}_1 = [x_1, y_1]^T \quad \text{and} \quad \boldsymbol{p}_2 = [x_2, y_2]^T \tag{1}$$

of the two regions corresponding to the instrument echos. The instrument position error in the US image (i.e. the position errors of the two regions in the image) is calculated according to

$$\boldsymbol{e}_1 = k(\boldsymbol{g}_1 - \boldsymbol{p}_1) \quad \text{and} \quad \boldsymbol{e}_2 = k(\boldsymbol{g}_2 - \boldsymbol{p}_2) \ , \tag{2}$$

with $\boldsymbol{g}_i$ denoting the desired position (goal) for the regions in the image plane and k being the conversion rate from pixel to meter. The value of k depends on the spatial resolution of the US device and on the resolution of the frame grabber card. It was experimentally identified to $k = 0.000825\ m/px$. The errors are then transmitted via a unidirectional TCP/IP socket to the RT Linux PC controlling the robot. There, the desired velocity of the instrument tip pose $\dot{\boldsymbol{x}}$ is calculated, using a linearizing proportional control law:

$$\dot{\boldsymbol{x}} = \boldsymbol{K}\boldsymbol{J}_{\text{image}}^{-1} \begin{bmatrix} \boldsymbol{e}_1 \\ \boldsymbol{e}_2 \end{bmatrix} \ , \tag{3}$$

with $\boldsymbol{J}_{\text{image}} \in \mathbb{R}^{4\times4}$ being the (invertible) image Jacobian and $\boldsymbol{K} \in \mathbb{R}^{4\times4}$ being a diagonal matrix with positive gains k_{ii}. The image Jacobian $\boldsymbol{J}_{\text{image}}$ depends on the current setup (which is supposed to be known) and describes the relation beween the position errors of the regions $[\boldsymbol{e}_1, \boldsymbol{e}_2]^T$ and the desired velocity of the instrument pose $\dot{\boldsymbol{x}}$. The desired pose velocity is then used to calculate the desired joint velocities $\dot{\boldsymbol{\Theta}}_d$ of the robot:

$$\dot{\boldsymbol{\Theta}}_d = \boldsymbol{J}^{-1}\dot{\boldsymbol{x}} = \boldsymbol{J}^{-1}\boldsymbol{K}\boldsymbol{J}_{\text{image}}^{-1} \begin{bmatrix} \boldsymbol{e}_1 \\ \boldsymbol{e}_2 \end{bmatrix} \ , \tag{4}$$

with $\boldsymbol{J} \in \mathbb{R}^{4\times4}$ being the (invertible) robot Jacobian matrix, which maps the robot joint velocities to the endeffector pose velocity. The computation of $\boldsymbol{J}_{\text{image}}$ as well as a detailed analysis of the robustness of the visual servoing approach with respect to parameter uncertainties is given in [6], [9] and is not detailed here. The joint velocities $\dot{\boldsymbol{\Theta}}_d$ are then sent to the robot controller hardware.

The complete visual servoing loop is sketched in Fig. 4.

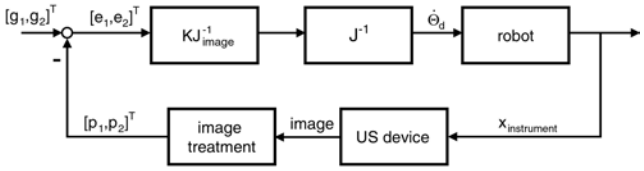


Fig. 4. Visual servoing loop.

III. Image Processing

The goal of the image treatment algorithms is to reliably track the instrument echoes in the image sequence, see Fig. 5. As these echoes (as well as the artefacts) have a high intensity, the first step is to calculate a robust threshold to detect these regions of interest. Due to large variations of the instrument echo intensity a fixed threshold leads to unsatisfactory results. Furthermore, the image treatment should be independent of the current parameter setting of the ultrasound device, e.g. the chosen depth correction. (The depth correction can be chosen by the surgeon and allows for an amplification of the US echo depending on the penetration depth of the ultrasound wave.)

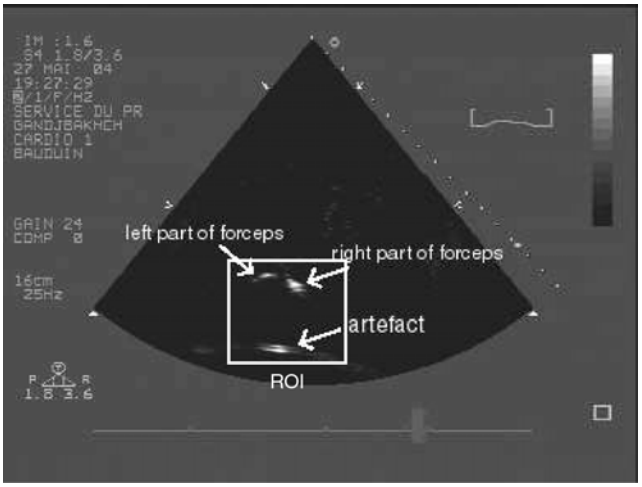


Fig. 5. Instrument echo, artefacts, speckles, and ROI.

First, a median filter is used to suppress small artefacts and noise. Afterwards, an adaptive threshold according to [7] is applied in real-time to account for intensity variations: the histogram of the region of interest (ROI) is calculated. Thereafter, the histogram is divided into two parts by chosing a certain threshold k (see Fig. 6). The values $\omega_i(k)$ and $\mu_i(k)$ are calculated for each group, with $\omega_1(k)$ being the number of pixels with intensity inferior or equal to k and $\omega_2(k)$ being the number of pixels with intensity superior to k. The mean intensities of these groups are denoted $\mu_1(k)$ and $\mu_2(k)$, respectively. The optimal threshold is calculated according to

$$k_{\text{opt}} = \arg \max_{\text{k}\in\{0,..,255\}} \omega_1\omega_2(\mu_2 - \mu_1) \ , \qquad (5)$$

by simply running through all possible values of k. This can be easily done in realtime. The optimal threshold k_{opt} guarantees that the two mean values μ_1 and μ_2 differ significantly and that both groups have a significant number of pixels. Therefore, k_{opt} is quite robust with respect to noise and intensity changes.

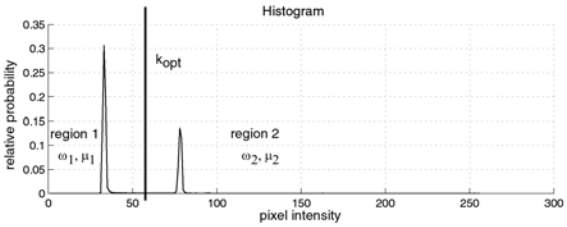


Fig. 6. Histogram and parameter for Otsu's method.

After thresholding, erosion and dilation (i.e. opening) are applied. This separates the instrument echoes from close artefacts and thus helps to increase tracking accuracy. Furthermore, small speckles with high intensity which were unintentionally detected are suppressed. Results for the different image processing steps are shown in Fig. 7.

Thereafter, a fast labelling algorithm based on run length coding is applied [8]. Thanks to its efficient implementation in MMX (intel's multi media extension) the algorithm allows for real-time labelling of regions.

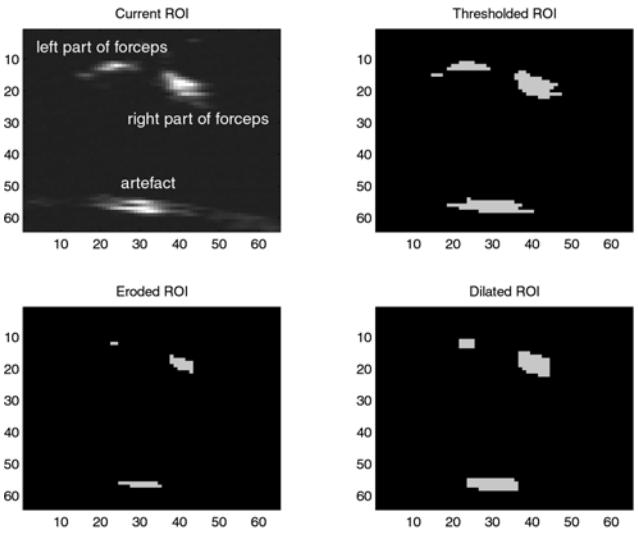


Fig. 7. Overview of image processing steps (all axes in pixel).

Usually, more than the two regions corresponding to the instrument echoes are detected in the selected ROI. To overcome ambiguities and to identify the correct regions a two step heuristic based on region size and velocity of the region center of gravity is applied. First, the region size s_i (in pixel) is considered. Only regions which satisfy

$$s_{\text{lower}} < s_i < s_{\text{upper}} \tag{6}$$

are further examined. This eliminates small regions due to speckles or artefacts as well as large regions due to other surgical devices or prominent organ structures. Second, the distances between the current region positions and the region positions in the precedent frame (i.e. the velocity of the regions) are considered. The two regions having the smallest distance are selected as the instrument echoes. This is based on the fact, that in MIRS the instrument moves slowly. During the experiments this heuristic proved to be robust with respect to the beforementioned disturbances.

Finally, the center $\boldsymbol{c} = [c_x, c_y]^T$ of the ROI is updated for the next image treatment, taking the position of the detected instrument echoes $\boldsymbol{p}_i$ into account:

$$\boldsymbol{c} = 1/2(\boldsymbol{p}_1 + \boldsymbol{p}_2) \ . \tag{7}$$

This allows to keep the region size small and therefore reduces computation time and possible ambiguities.

IV. Experimental Results

This section presents image processing results for the different materials used, as well as trajectories resulting from the closed visual servoing loop.

Examples for the ultrasound images are given in Fig. 8. Instrument echoes are indicated by white arrows. Polyurethane (PUR) and nylon yielded distinctive echoes. Polyvinyl chloride gave rise to frequent speckles and artifacts as a high gain tuning of the US device was necessary to obtain significant and stable echoes. In this case, it was hardly possible to close a stable visual servoing loop as tracking was prone with frequent outliers. The instrument made of plexiglas yielded only one large echo instead of two, so it was impossible to close the visual servoing loop.

For the algorithms presented in Sect. III the following parameters were used: the filter mask of the median filter was 3×3 pixels and so was the mask for the dilation. Erosion was not applied, as it was experimentally observed to be unnecessary. The region of interest (ROI) was 32×32 pixels. This was tuned as a function of the known instrument size, i.e. the known distance between the two jaws of the pair of forceps. The labelling algorithm tolerated gaps up to a length of three pixels in one line. Image processing results for the PUR instrument are given in Fig. 9: the two detected blobs, the ROI, and the desired position (g_1 and g_2) are shown. Furthermore, the histogram of the ROI and the calculated threshold according to [7] are presented. In this case the image processing time is approx. 10 ms (the time indicated in Fig. 9 also includes the time necessary to store the image, which leads to an overall treatment time of approx. 100 ms).

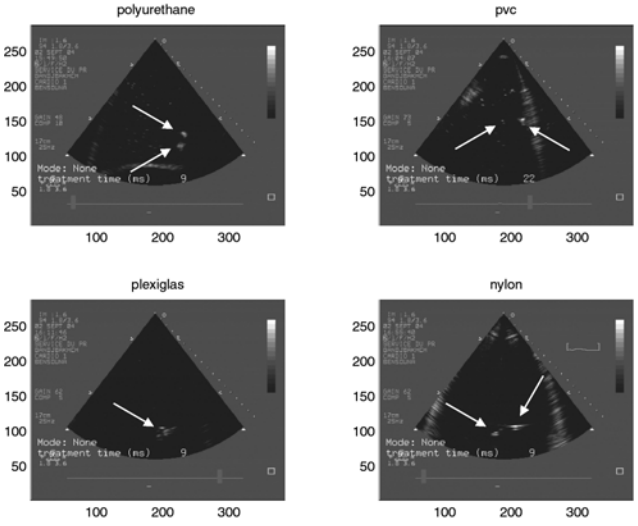


Fig. 8. Instrument echoes for materials used.

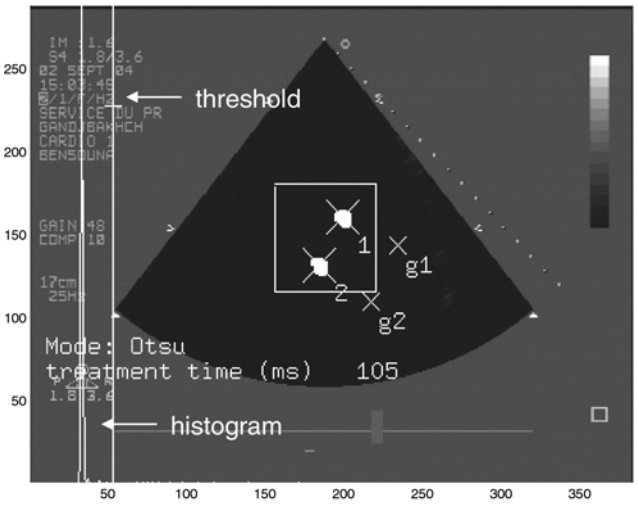


Fig. 9. Example of image processing results for polyurethane.

The calculated thresholds for the PUR instrument and the nylon instrument are given in Fig. 10 and Fig. 11, respectively. Here, the instrument was moving, see Fig. 13 for the trajectories of the PUR instrument. Considering the large variations of the threshold it becomes clear that a fixed threshold leads to unsatisfactory results. Furthermore, the applied method allows for an automated adaptation of the vision algorithms to the chosen instrument material as well as gain and depth correction of the ultrasound device.

The desired positions for blob one and two, as well as their measured trajectories for the PUR instrument are shown in Fig. 12. It can be seen, that the measured trajectories converge well towards the desired position with no remaining offset. During the experiment the gains k_{ii} (diagonal elements of matrix K of the proportional control law Eq. 3) used to calculate the desired veolcity of the

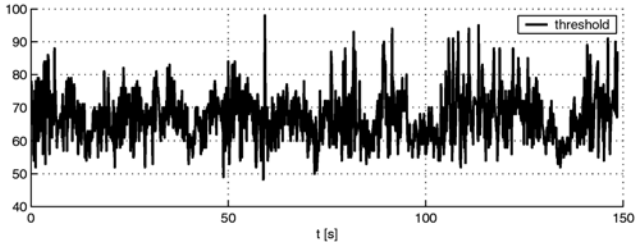


Fig. 10. Calculated threshold for PUR instrument.

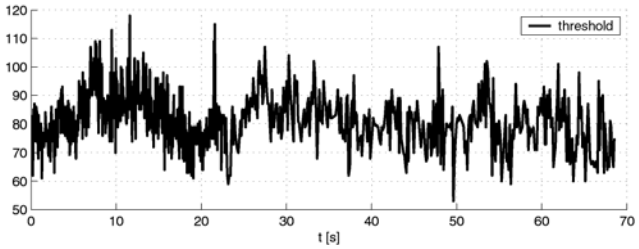


Fig. 11. Calculated threshold for nylon instrument.

instrument pose $\dot{\boldsymbol{x}}$ were increased, from 0.4 s^{-1} to 2.2 s^{-1}. Therefore, the step response of the closed control loop became faster (i.e. the system bandwidth increased). This can be seen in Fig. 12, too. More details on the performed experiments, the computation of the Jacobian $\boldsymbol{J}$, and the robustness of the closed loop system can be found in [9].

The measurement noise is less than ± 1 pixel (see Table I). This leads to an overall system inaccuracy of less than ± 1 mm, as no offset between the desired and measured instrument position remains. Here, 500 measurements for the PUR case and 400 measurements for the nylon case were considered. When the positions of the echoes were recorded the instrument was not moving. Due to low noise trajectories are very smooth. The trajectories of the detected regions in the US image plane are given in Fig. 13. The curves are close to a straight line, thus indicating a good calibration of the system (here: mainly the spatial relation between robot base frame and US probe frame [6]). Trajectories for the nylon instrument are similar and are therefore not shown here.

TABLE I

MEASUREMENT NOISE FOR PUR AND NYLON INSTRUMENT.

	blob 1, x	blob 1, y	blob 2, x	blob 2, y
PUR	0.455 px	0.274 px	0.554 px	0.448 px
nylon	0.352 px	0.483 px	0.741 px	0.499 px

Finally, the position errors of the two blobs are shown in Fig. 14. They are transmitted to the robot controller which computes the desired Cartesian velocity of the robot, as given in Eq. 3.

During the experiments the closed control loop proved to be robust with respect to errors in the image Jacobian $\boldsymbol{J}_{\text{image}}$, which occur due to uncertainties of the experimental setup. The spatial relation between US probe and

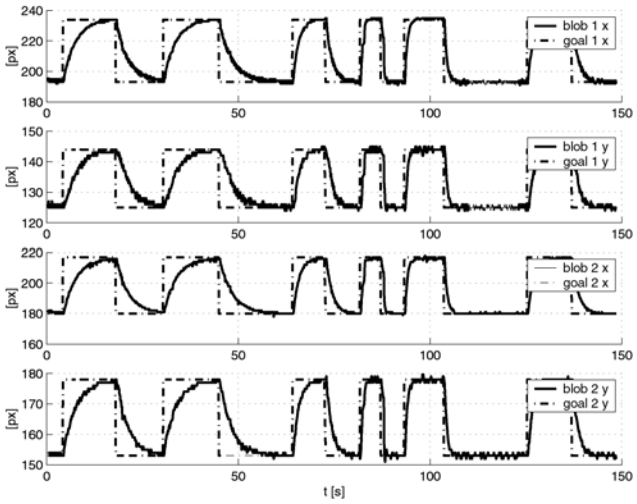


Fig. 12. Desired and measured trajectories for PUR instrument echoes.

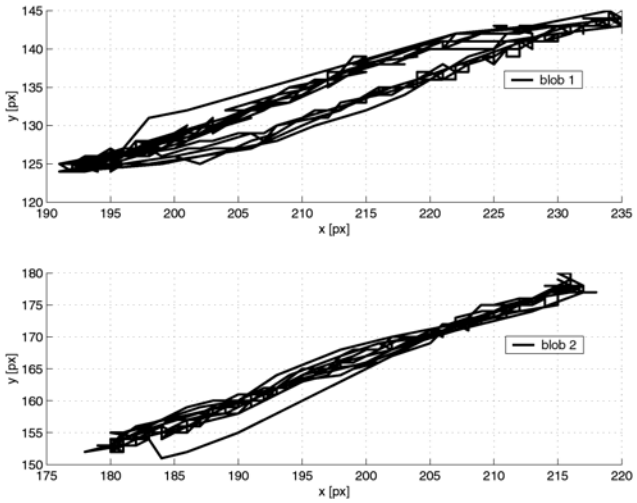


Fig. 13. Measured trajectories for PUR instrument in the US image.

robot base, which is needed to compute $\boldsymbol{J}_{\text{image}}$, is not exactly known. This validates the simulations presented in [6]. Further details on the closed visual servoing loop and its robustness with respect to parameter uncertainties during experiments can be found in [9].

V. DISCUSSION AND CONCLUSION

Reliable tracking of instruments in US image sequences is possible in realtime. During the experiments four different instrument materials were tested: PVC, nylon, plexiglas, and PUR. The instruments made of PUR and nylon yielded good (i.e. significant) echoes. Tracking of the PVC instrument was possible but error prone. The plexiglas instrument produced only one echo instead of two echoes.

The trajectories of the PUR and nylon instrument echoes in the image plane are smooth and only few outliers occurred. Adaptive thresholding is crucial due to large intensity variations in the image sequences and allows to automatically cope with changes of gain and depth

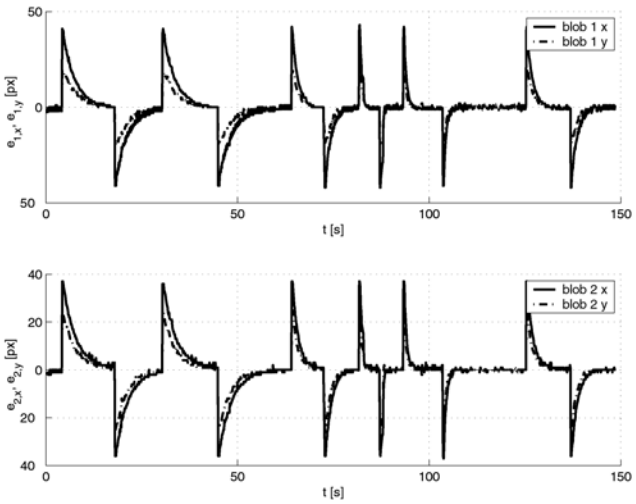


Fig. 14. Position errors for PUR instrument echoes.

correction of the ultrasound device. Median filter and morphological operations suppress noise and small artefacts. A heuristic based on echo size and blob velocity helps to overcome ambiguities. The overall computation time on a standard PC is below 10 ms per image.

Closing the visual servoing loop yielded good results: the system is robust with respect to large errors in the image Jacobian which are due to uncertainties of the experimental setup. Additionally, no remaining offset between the desired instrument position and the measured instrument position occurs. The system bandwidth is satisfactory.

Future work includes experiments to evaluate further the robustness of the overall system with respect to kinematic errors in the setup. Additionally, in vivo experiments are scheduled to examine the performance of the vision algorithms under more realistic circumstances.

ACKNOWLEDGMENTS

The authors thank Dr. U. Frese for the fast MMX labelling algorithm and his great *xdisplay* library. Also, the authors thank Dr. M. Karouia and Dr. N. Bonnet for their support during the experiments. The financial support of the ROBEA program (CNRS/SPI) through the GABIE project is gratefully aknowledged.

REFERENCES

[1] Ivana Mikic, Slawomir Krucinski, and James D. Thomas. Segmentation and tracking in echocardiographic sequences: Active contours guided by optical flow estimates. *IEEE Transactions on Medical Imaging*, 1998.

[2] F. Escolano, M.A. Cazorla, D. Gallardo, and R. Rizo. Deformable templates for tracking and analysis of intravascular ultrasound sequences. In *Proceedings of First International Workshop of Energy Minimization Methods in Computer Vision and Pattern Recognition*, 1997.

[3] Yusuf Sinan Akgul, Chandra Kambhamettu, and Maureen Stone. Task-specific contour tracker for ultrasound. In *IEEE Workshop on Mathematical Methods in Biomedical Image Analysis*, 2000.

[4] P. Abolmaesumi, S. Salcudean, W. Zhu, M. Sirouspour, and S. DiMaio. Image-guided control of a robot for medical ultrasound. *IEEE Transactions on Robotics and Automation*, 18(1), 2002.

[5] N. Zemiti, T. Ortmaier, M.-A. Vitrani, and G. Morel. A force controlled laparoscopic surgical robot without distal force sensing. In *Proc. of the ISER 2004; 9th International Symposium on Experimental Robotics*, Singapore, June 2004.

[6] M.-A. Vitrani, T. Ortmaier, and G. Morel. Automatic guidance of a surgical instrument with US based visual servoing. In *Proc. of the 1st SURGETICA Conference: Computer Assisted Medical and Surgical Interventions*, Chambery - Sovie, France, January 2005.

[7] Otsu N. A threshold selection method from grey-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66, January 1979.

[8] U. Frese, B. Bäuml, S. Haidacher, G. Schreiber, I. Schaefer, M. Hähnle, and G. Hirzinger. Off-the-Shelf Vision for a Robotic Ball Catcher. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1623–1629, 2001.

[9] M.-A. Vitrani, G. Morel, and T. Ortmaier. Automatic guidance of a surgical instrument with ultrasound based visual servoing. In *IEEE International Conference on Robotics and Automation (ICRA)*, Barcelona, Spain, April 2005.