Proceedings of the 2010 3rd IEEE RAS & EMBS
International Conference on Biomedical Robotics and Biomechatronics,
The University of Tokyo, Tokyo, Japan, September 26-29, 2010

Contour-based Surgical Instrument Tracking Supported by Kinematic Prediction

C. Staub, C. Lenz, G. Panin, and A. Knoll
Robotics and Embedded Systems
Technische Universität München
{staub|lenz|panin|knoll}@in.tum.de

R. Bauernschmitt
Department of Cardiovascular Surgery
German Heart Center Munich
bauernschmitt@dhm.mhn.de

Abstract— **Surgical tool tracking is an important key functionality for many high-level tasks in both robot-assisted and conventional minimally invasive surgery. Though the fields of application are similar in both surgery techniques (i.e. visually servoed instruments, workflow analysis or augmented reality), the kind of available information about the position and orientation of the surgical tool differ. In conventional laparoscopic surgery additional information to the images provides by the endoscopic camera can only be obtained by an external tracking system. In contrast, robotic systems provide angular informations from encoder readings that allow for a sufficient pose estimation and initialization of an image-based tracking algorithm. Our approach utilizes both encoder readings and visual information, in order to stabilize tracking in image space. The image-based tracking is supervised by means of the kinematic information and reinitialized in case of conflicting results. As tracking modality we utilize the Contracting Curve Density (CCD) algorithm that looks for maximal separation of local color statistics along the contour of a model.**

I. INTRODUCTION

In the last years, robot-assisted minimally invasive surgery (MIS) has attracted many researches and significant efforts have been made in the development of surgery systems and instruments. Because MIS is performed through small incisions or ports in the body, patients considerably benefit from reduced tissue trauma, recovery time and pain, compared to conventional "open" surgery. Reduced dexterity and perception, known from long instruments with fewer degrees of freedom (DoF) and missing force or tactile feedback in non-robot assisted procedures have been replaced by teleoperated *slave* manipulators that are commanded from a *master* console. These systems seek to restore the feel of a regular surgery by providing the surgeon an intuitive interface, including 3D imaging and advanced tools, providing fully six degrees of freedom manipulation. Although robotic surgery systems such as DaVinci^{TM}[1] are in wide use for a variety of abdominal, pelvic, and thoracic procedure, delicate maneuvers are still cumbersome and time-consuming, even the systems provide additional support with features such as tremor filtering to alleviate the handling.

Recently, automation of error-prone and recurrent (sub-) tasks that yield to a quick fatigue of surgeons and account noticeable for a higher overall surgery time have drawn the attention of researchers. Given that knot-tying occurs frequently during surgery, automating this challenging subtask is tackled by several groups (e.g. [2], [3], [4]). Furthermore, techniques for assisting the surgeon with visually guided instruments ([5], [6], [7], [8]) and autonomously navigated endoscopic cameras have been developed (e.g. [9], [10]). For the reason of documenting and benchmarking surgical interventions, and to anticipate potential mistakes in the surgical workflow, modeling and analyzing these procedures has become an active field of research [11], [12].

Despite the manifold of challenges in minimally invasive surgery and the above mentioned achievements in partly autonomous navigation and manipulation, the visual identification, segmentation, and tracking of operated surgical tools during surgery is a crucial requirement for developing techniques that assist the surgeon. As most of the methods require pose information of the surgical instrument, a robust and precise automatic detection is the first step towards higher level functionality. Many of the proposed instrument tracking approaches rely on image processing techniques that use either pure color information or additional geometrical knowledge. Wei et al. [9] analyzed the typical color distribution in laparoscopic images to identify an adequate color that can be used for optical markers that are attached at the distal end of the instrument. The marker is segmented in HSV color space and background noise is filter at a rate of 17Hz. Uckert et al. [13] includes additional geometrical shape information about the shaft to fit a bounding box to the color-classified pixels. Two different shapes are used, a trapezoidal for near-field cases and a rectangular for far-field cases. In [14] it was taken advantage of the metallic appearance of the shaft to track gray regions by joint hue saturation color features. A seeded region growing method was implemented, operating at 13fps. Therefore, the fulcrum is estimated with a series of images in order to project an approximated instrument direction and shape into the image. Voros et al. [15] also reduces the search space by considering the insertion point of the instrument. At the beginning of the procedure, the fulcrum has to be visible in the image and is marked with a "vocal mouse". They state that any kind of surgical instrument can be detected since no color information is used, but only the gradients of the instrument edges, constrained by the incision point. To enhance the computation speed, the image resolution is reduced to 200×100 pixels. The precision of the predicted tip position ranges around 11 pixels. The *Center for Computer Integrated Surgical Systems and Technology* (CISST, Johns Hopkins University, Baltimore) deals with the articulated DaVinci^{TM}instruments. Burschka et al. [16] used template

978-1-4244-7709-8/10/$26.00 ©2010 IEEE

images of the instrument to detect the position of the forceps in stereo images, enriched with additional information and orientation information derived from the trajectory provided by the robot. The method works in real-time, but they report that the kinematic data suffers from significant rotational and translational errors. More recently, the CISST reported a general purpose articulated object tracker [17] and demonstrated its application to surgical scenarios. The geometry and kinematics of the objects have to be known a priori. The appearance of different body parts is modeled by a class-conditional probability and compared with the image after rendering the target object geometry. So far, images are hand segmented to train the appearance model and computation time is around $5sec$ per frame at a resolution of 640×480.

II. System Setup

Hardware and software of the system itself have already been introduced to the research community [3]. Therefore, we constrict the following description to an extent necessary for understanding the subsequent sections.

A. Robotic System

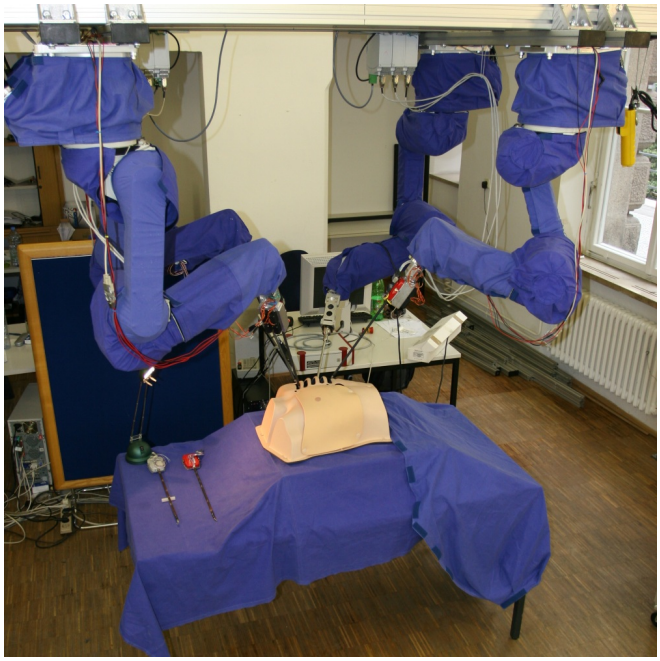

Fig. 1. **Hardware Setup.** Ceiling mounted robots with surgical instruments

As illustrated in Figure 1, the slave manipulator of the system consists of four ceiling-mounted robots which are attached to an aluminum gantry. The robots have six degrees of freedom and are equipped with either a 3D endoscopic stereo camera or with minimally invasive surgical instruments, which are originally deployed by the DaVinci^{TM}system. The surgical instruments have 3DoF. A micro-gripper at the distal end of the shaft can be rotated and adaption to pitch and jaw angles is possible. Fast and easy interchangeability is ensured by a magnetic clutch which releases in case of forces at the instruments exceeding a certain limit (e.g. during a collision). To measure forces during operation the instruments are augmented with strain gauge sensors. The master-side manipulator is mainly composed of a 3D display, some foot switches for user interaction (like starting and stopping the system or executing the piercing process) and of the main in-/output devices, two PHANToM^{TM}haptic displays. On one hand, the devices are used for 6DoF control of the slave manipulator and on the other hand provide a 3DoF force feedback derived from the measurements at the instruments. The control software of the system realizes trocar kinematics, whereby all instruments will move about a fixed fulcrum after insertion into the body. For computer vision tasks a NVIDIA^{TM}Quadro FX 580 graphics card is used for acceleration.

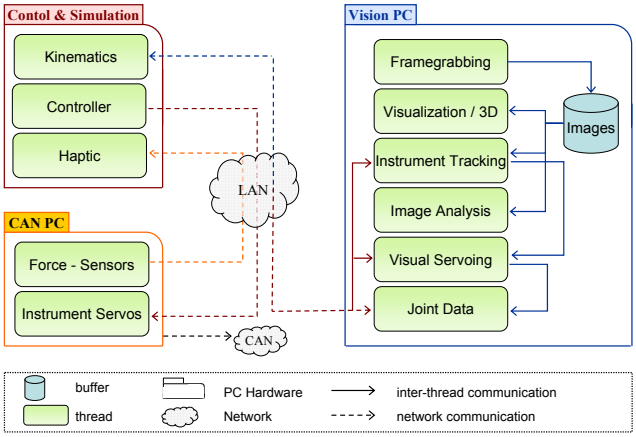

Fig. 2. **Software Architecture.** The software of the system is distributed to 3 PC's that communicate via network connections

B. Distributed Software Environment

The software architecture of our system is distributed over 3 standard PC's. It is partitioned into a *simulation and control* part, a *vision* part and one computer is connected to a *CAN* network (cp. Fig. 2). The commands for the servos-motors that control the joints of the instrument as well as the data that is provided by the amplifiers of the strain gauge sensors are communicated between the simulation PC and the PC that is connected to the CAN network. The GUI of the simulation environment comprises an interface to a 3D model of the scene, which can be manipulated in real time. Parameters of each model can be adjusted and joint angles of the robots can be altered this way. New trajectories can be generated by means of a key framing module, incorporating a collision detection. On one hand, joint data can directly be sent to the robot hardware, on the other hand the poses of the instruments and robots are synchronized with the "Vision PC" for further processing. For this reason, enough computing power can be provided for image analysis, i.e. instrument tracking, visual servoing or augmented reality. Most of the image processing tasks run in individual threads that have access to an image database, which holds up-to-date images provided by the stereoscopic endoscope.

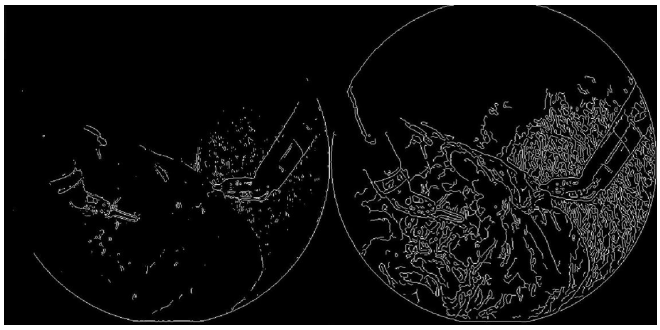

Fig. 3. **Edge detection.** The images show edge detection results of the Sobel (left) and the Canny (right) filter. In both cases the tool shaft can hardly be distinguished from background noise.

III. Tracking supported by Kinematic Prediction

The tracking of surgical tools is particular challenging due to the changing appearance of the background (i.e. background movement through organs, non-uniform and time-varying lightning conditions, smoke caused by electro-dissection and specularities), but also due to the partial occlusion of the instrument and body fluids that may change the appearance of the instrument itself. In many cases of surgical tool tracking the tracking is constricted to a sequential "frame-by-frame detection" (also referred to as *detection*), rather than including a motion model. Accordingly, no optimization of the configuration space or pose prediction is performed over time. In a Bayesian *prediction-correction* context, the state of the object is updated by integrating posterior statistics and therewith knowledge about time-depending characteristics of the movement. This "intelligence" within our tracking pipeline is provided by a Kalman Filter that is running on the output of a contour tracker, known as contracting curve density algorithm (CCD), based on separation of local color statistics [18], [19]. The separation takes place between the object and the background regions, across the projected shape contour of a CAD model under a predicted pose hypothethis. The processing flow of the tracker is shown in Fig. 5.

Bayesian tracking involves a detection step to initialize the system in the very first frame or after encountering a track loss. Instead of relying upon visual data, we take the object pose, given by the kinematic measurement from robot sensor readings. The precision of this coarse approximation is limited due to the absolute accuracy of our system (and also of most complex robotic systems) by 1) the mounting of the robots on an aluminum gantry which is afflicted by several intrinsic aberrations, 2) instruments that exhibit imprecise calibration and unpredictable play, 3) the magnetic clutch that couples the instruments to the robot flange, and 4) limitations of the robotic hardware itself. To attain the best possible result, a precise overall system calibration has been performed [20]. The idea of integrating joint angle measurements for tracking purpose was i.e. also used by Ruf et al. [21] to track a polyhedral tool and simultaneously adapt inaccuracies in the static calibration of the robot. To restrict the initial search from the first frame to a specific region is computational more efficient than a complete image analysis and can also be motivated by biological considerations: Biologically inspired algorithms seek to direct the attention rapidly towards a ROI, using an attention-based type of filter, and only process a smaller amount of the visual input data [22]. *Bottom-up* approaches compute visual salient features, such as regions of high contrast, local scene complexity or high scene dynamics. The second type of visual attention is often referred to as *top-down* attention, as the attention is controlled from higher areas of cognition. Kinematic measurements, which are fed to the visual information processing by another software component (thus, a higher area of cognition), can guide the attention directly to a region of interest. This idea is directly applicable to the proposed method, independent of the utilized type of feature matching.

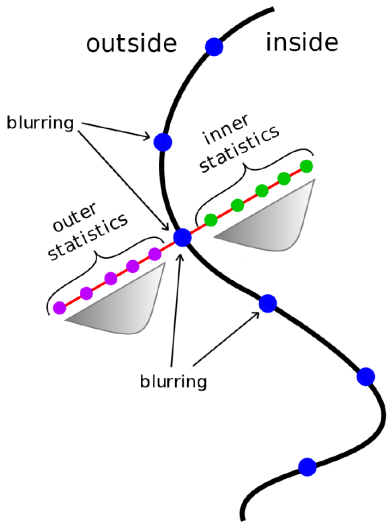

Fig. 4. The CCD algorithm tries to maximize the separation of color statistics between two image regions. The algorithm first samples pixels along the normals for collecting local color statistics.

A. Model Building

Our system is equipped with the EndoWrist™needle driver tools that are originally deployed with the DaVinci™system. The instruments are composed of a long grayish shaft, a wrist joint and two brackets. It is represented as a polygonal mesh model with 6DoF (3 rotations, 3 translations) in word coordinates by a 4×4 transformation matrix. As our main interest is visual servoing, we need a tracking in the image domain. Therefore, a simple rectangular model can be used as target to represent the projection of the shaft cylinder. The pose parameters of the 3D instrument are reduced to a planar roto-translational pose s with scale in image space.

$$s = (t_x, t_y, h, \theta_x) \tag{1}$$

where the rotation θ_x has to be determined newly for the 2D projection. As the 3D pose is given, this can be performed by calculating the intersection angle of the edge contour of

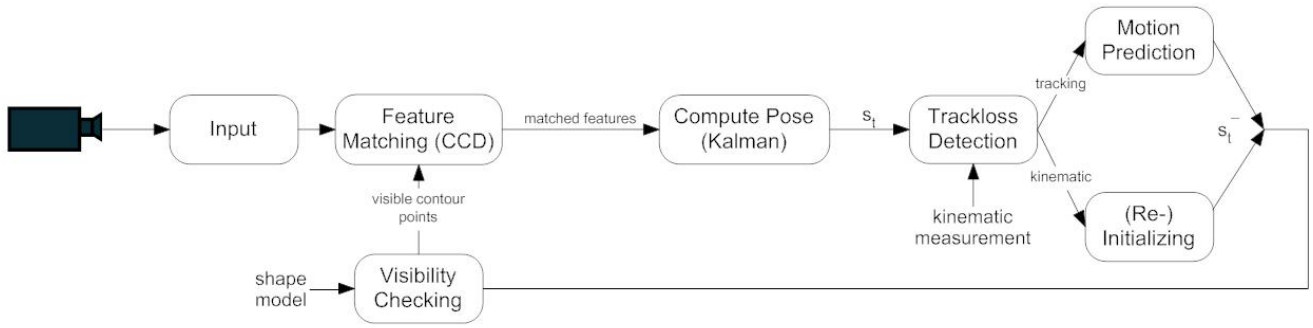

Fig. 5. **Tracking Pipeline.** The camera pose can be obtained after calibrating the extrinsic parameters and the overall system. The kinematic measurement of the instrument in 6 degrees of freedom is transfered to a 2D model with 4 DoF (t_x, t_y, h, θ_x). It is used to (re-)initialize the Contracting Curve Density algorithm.

the shaft projection and the coordinate system of the image domain. The position information can directly be projected to image space.

B. Tracking with CCD

As already mentioned above, tracking in the context of MIS procedures is exacerbate by changing environment conditions. Simple color segmentation approaches often fail due to varying lightening conditions of different light sources or need a sophisticated fine tuning of parameters. Algorithms that are based upon edge detection suffer from the large amount of feature edges from the background. Figure 3 shows a typically intra-operative scene with an artificial heart and tissue in the background. Neither the Sobel- nor the Canny operator can distinguish the instrument shaft reliable from the background.

The amenity of the CCD modality is that the appearance of the model is adjusted over time, since local color statistics are computed in every tracking cycle and maximized according to the shape of the model. Therefore, the method can be applied to marker-based as well as markerless tracking. After setting the initial pose, the Kalman filter generates a prior state hypothesis s_t^- by applying a Brownian motion model to the previous state (s_{t-1}).

$$s_t^- = s_{t-1} + w_t \tag{2}$$

with w being a white Gaussian noise sequence.

The CCD modality requires a sampling of good features for tracking from the object model under the given pose s_t^- and camera view. As a first step, the visible internal and external edges from the polygonal mesh model have to be identified under the current pose hypothesis. In our case, we use the silhouette of a 2D rectangle to represent the shaft. Alongside of this contour a set K of uniformly distributed sampling points $\{h_1, \ldots, h_2\}$ is taken to collect color statistics around each sample position on each side of the contour. The basic idea of CCD is to maximize the separation of local color statistics between the two sides of the object boundaries (object vs. background) [18]. The grayish shaft of the instrument supports this idea by strongly varying from red tissue and organs. Contemporaneously, the algorithm can account for small change of the shaft appearance over time (e.g. from body liquids), since the statistics are updated in every iteration.

We first sample points along the respective normals, separately collect the statistics, and afterwards blur each statistic with the neighboring ones (cp. Fig. 4). From each contour position h_i, foreground and background color pixels are collected along the normals n_i up to a distance L (that is manually defined and fix), and local statistics up to the 2^{nd} order are estimated

$$v_i^{0,B/F} = \sum_{d=1}^{D} w_{id} \tag{3}$$

$$v_i^{1,B/F} = \sum_{d=1}^{D} w_{id} I(h_i \pm L\overline{d} n_i)$$

$$v_i^{2,B/F} = \sum_{d=1}^{D} w_{id} I(h_i \pm L\overline{d} n_i) I(h_i \pm L\overline{d} n_i)^T \tag{4}$$

with $\overline{d} \equiv d/D$ the normalized contour distance, where the $\pm$ sign is referred to the respective side, and image values I are 3-channel RGB. The local weights w_{id} decay exponentially with the normalized distance, thus giving a higher confidence to observed colors near the contour.

Single line statistics are afterwards *blurred* along the contour, providing statistics distributed on local areas

$$\widetilde{v}_i^{o,B/F} = \sum_j exp(-\lambda\,|i-j|) v_j^{o,B/F}; o = 0, 1, 2 \tag{5}$$

and finally normalized

$$\overline{I}_i^{B/F} = \frac{\widetilde{v}_i^{1,B/F}}{\widetilde{v}_i^{0,B/F}} \tag{6}$$

$$\overline{R}_i^{B/F} = \frac{\widetilde{v}_i^{2,B/F}}{\widetilde{v}_i^{0,B/F}}$$

in order to provide the two-sided, local RGB means $\overline{I}$ and (3×3) covariance matrices $\overline{R}$

The second step involves computing the residuals and Jacobian matrices for the Gauss-Newton pose update. For this

purpose, observed pixel colors $I(h_i + L\overline{d}n_i)$ with $\overline{d} = -1, \ldots, 1$ are classified according to the collected statistics (7), under a fuzzy membership rule $a(x)$ to the foreground region

$$a(\overline{d}) = \frac{1}{2}\left[erf\left(\frac{\overline{d}}{\sqrt{2}\sigma}\right) + 1\right] \tag{7}$$

which becomes a sharp $\{0; 1\}$ assignment for $\sigma \to 0$; pixel classification is then accomplished by mixing the two statistics accordingly

$$\begin{aligned} \hat{I}_{id} &= a(\overline{d})\overline{I}_i^F + (1 - a(\overline{d}))\overline{I}_i^B \\ \hat{R}_{id} &= a(\overline{d})\overline{R}_i^F + (1 - a(\overline{d}))\overline{R}_i^B \end{aligned} \tag{8}$$

and color residuals are given by

$$E_{id} = I(h_i + L\overline{d}n_i) - \hat{I}_{id} \tag{9}$$

with covariances $\hat{R}_{id}$.
Finally the $(3 \times n)$ derivatives of E_{id} can be computed by differentiating (7) and (9) with respect to the pose parameters

$$J_{id} = \frac{\partial \overline{I}_{id}}{\partial s} = \frac{1}{L}\left(\overline{I}_i^F - \overline{I}_i^B\right)\frac{\partial a}{\partial \overline{d}}\left(n_i^T \frac{\partial h_i}{\partial s}\right) \tag{10}$$

which are stacked together in a global Jacobian matrix $\mathbf{J}_{ccd}$. The state is then updated using a Gauss Newton step:

$$\begin{aligned} s &= s + \Delta s \\ \Delta s &= \mathbf{J}_{ccd}^{+}\mathbf{E}_{ccd} \end{aligned} \tag{11}$$

The optimization is done until the termination criteria is satisfied $(\Delta s \approx 0)$.

IV. Experimental Results and Conclusions

A. Experimental Results

The evaluation has been performed on a Intel Xeon QuadCore™2.4Ghz system. Images were taken and processed in real-time with full PAL resolution (768×576) from the framegrabber.
As a first step, the precision of the instrument projection into image space, taken from the kinematic data, was tested. The data is transmitted via network and applied to the geometrical CAD model of the instrument. The pose of the camera is set in a similar fashion. The projected shaft does not have to overlay the image perfectly, but a good match supports a fast initialization of the tracking. Also, the search length along the normals of the sampled contour points can be kept smaller. The search length was determined experimental and has to be set once. Figure 7 depicts the shaft overlay as well as the first tracking steps during the alignment of the model. We provide some experiments, showing the performance of the tracking system for different, more or less crucial poses of the instrument. The tests have been performed twice: with an instrument that has an attached color marker at the distal end, and without any additional markers. The first row of Fig. 8 shows the result of the tracking with marker. Concerning robustness against varying lightning conditions, partial occlusion and reinitialization it outclasses the markerless tracking (Fig. 8, second row). The major problem during markerless tracking is a drift of the rectangular model along the shaft (Fig. 8, second row, last picture). Since statistics are evaluated at all contour edges, also the two edges that actually belong to the shaft get included in the computation. At those edges, no differentiation between object and background can be achieved. Hence, the computation is not unique and the model starts drifting away. Only at the foremost edge (at the distal end) a classification is possible.
Another problem is the limitation to only one scaling factor for the model. Allowing an independent scaling of both sides of the rectangle could prevent a misalignment, such as depicted in Fig. 8, first row, last image. Since the tip of the shaft is still detected correctly, this kind of error does barely affect the tracking.
Fig. 6 provides a comparison between the kinematic prediction and the tracking result. In general, a good agreement of the movement can be observed between measurement and kinematic estimation.
By considering more challenging situations, we also performed tests with fast changing lightning conditions and partial occlusion of the marker have been performed (please see the video file, corresponding to this paper).

B. Conclusion and Future Work

In this paper, we have presented an approach to track surgical instruments during robot-assisted minimally invasive surgery, based on kinematic pose prediction and image analysis. For the image-based tracking, the Contracting Curve Density algorithm was used. For initializing and in case of tracking loss, the instrument pose was estimated via the kinematic chain and projected into image space. During experiments, instruments with applied markers showed robust tracking performance with respect to varying lightning conditions and partial occlusion. During markerless tracking the detection of the shaft was possible most of the time. However the model could not always be matched with the instrument tip, but was drifting on the entire shaft length. In regions with very dark background, the grayish shaft could not be separated correctly and tracking was lost. To improve this situation, a more complex model could be used. On one hand, only edges at the outer side of the shaft as well as the edge pointing towards the instrument tip should be used to sample feature points. On the other hand, an articulated model that introduces a second segment, representing the silver-colored brackets, might correct the drifting issue and fixate the model permanently at the instrument tip during tracking.
Also a different motion model or the fusion of kinematic prediction and feature-based measurement could further improve the robustness of the tracking.

V. ACKNOWLEDGMENTS

This work is supported by the German Research Foundation (DFG) within the Collaborative Research Center SFB 453 on “High-Fidelity Telepresence and Teleaction”.

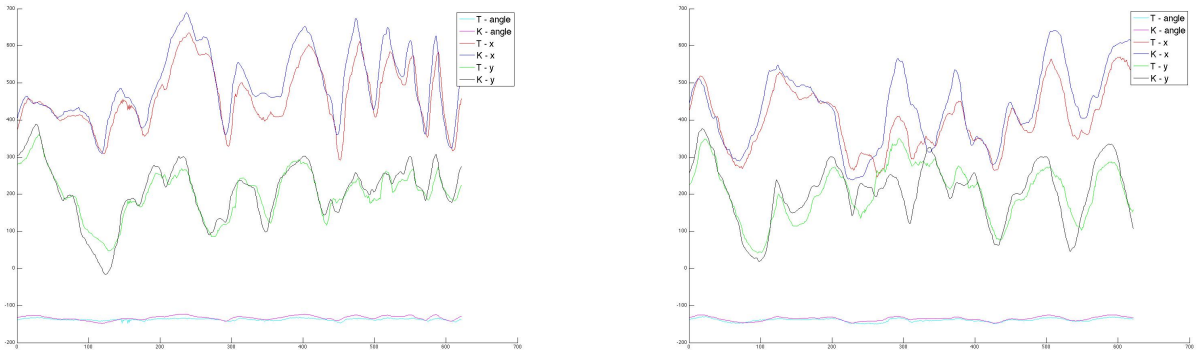

Fig. 6. Tracking results vs. kinematic readings (angle, x- and y-position). The x-axis denotes the time step, the y-axis denotes the position in pixels and the rotation in degrees. The left picture shows the tracking result with the marker applied to the instrument tip. The right plot illustrates the markerless tracking. Conspicuously, the angle still matches the kinematic prediction, but x- and y-coordinates drift apart from time to time. In particular in between the time steps 220 and 430 the position differs from the kinematic readings, while the angular part still matches. This happens, as the rectangular model moves up and down the shaft and does not "snap" to the tip. Please note that a natural displacement of some pixels exists between the kinematic measurement and the tracking, as the marker is not position at the very distal end, but the kinematic outputs the end of the shaft.

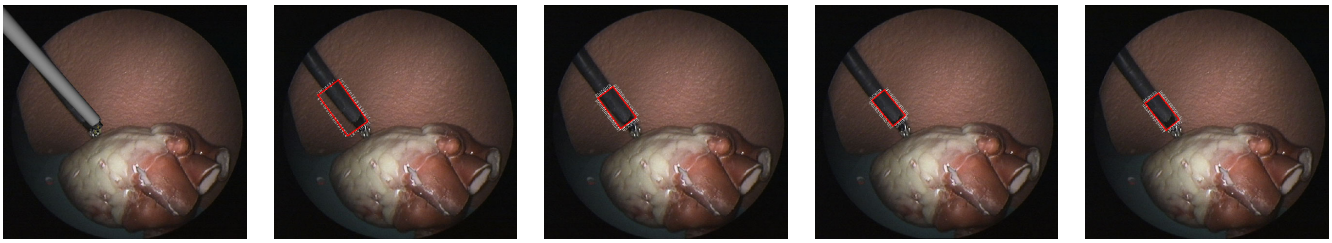

Fig. 7. Initialization of the tracking. The first frame shows the overlay of the kinematic projection with the endoscopic image. It is precise enough for a first pose estimation. The following images are frame numbers $1, 2, 3$ and 5 and show how the shape is adjusted to the shaft.

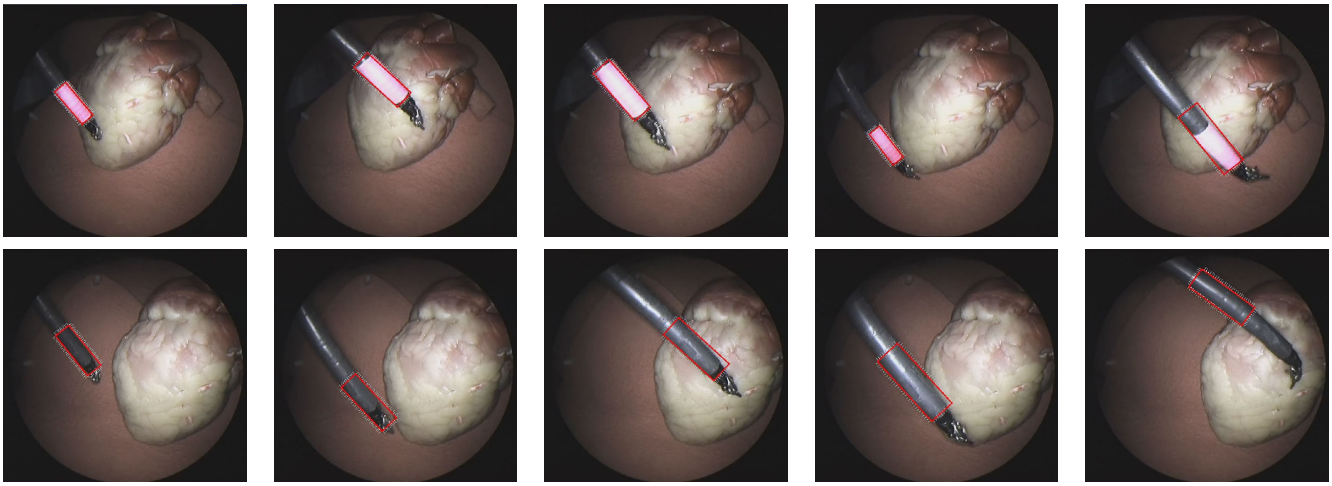

Fig. 8. Top row: Tracking with marker. In certain positions the scaling does not exactly fit to the marker, but the tip of the instrument is still recognized (last frame). Bottom row: Tracking without marker. A major problem is the drifting of the marker along the instrument shaft (images 2 and 5).

REFERENCES

[1] G. Guthart and J. Salisbury, J.K., "The intuitive^{TM}telesurgery system: overview and application," *Robotics and Automation, 2000. Proceedings. ICRA '00. IEEE International Conference on*, vol. 1, pp. 618–621, 2000.

[2] H. Mayer, D. Burschka, A. Knoll, E. Braun, R. Lange, and R. Bauernschmitt, "Human-machine skill transfer extended by a scaffolding framework," may 2008, pp. 2866 –2871.

[3] H. Mayer, I. Nagy, A. Knoll, E. Braun, R. Lange, and R. Bauernschmitt, "Adaptive control for human-robot skilltransfer: Trajectory planning based on fluid dynamics," april 2007, pp. 1800 –1807.

[4] H. Wakamatsu, A. Tsumaya, E. Arai, and S. Hirai, "Manipulation planning for knotting/unknotting and tightly tying of deformable linear objects," in *Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on*, April 2005, pp. 2505–2510.

[5] C. Staub, T. Osa, A. Knoll, and R. Bauernschmitt, "Automation of tissue piercing using circular needles and vision guidance for computer aided laparoscopic surgery," in *Proceedings of the IEEE International Conference on Robotics and Automation*, May 2010, to appear.

[6] F. Nageotte, P. Zanne, C. Doignon, and M. de Mathelin, "Stitching planning in laparoscopic surgery: Towards robot-assisted suturing," *The International Journal of Robotics Research*, p. 0278364909101786, 2009.

[7] P. Hynes, G. Dodds, and A. Wilkinson, "Uncalibrated visual-servoing of a dual-arm robot for mis suturing," *Biomedical Robotics and Biomechatronics, 2006. BioRob 2006. The First IEEE/RAS-EMBS International Conference on*, pp. 420–425, Feb. 2006.

[8] A. Krupa, J. Gangloff, C. Doignon, M. de Mathelin, G. Morel, J. Leroy, L. Soler, and J. Marescaux, "Autonomous 3-d positioning of surgical instruments in robotized laparoscopic surgery using visual servoing," *Robotics and Automation, IEEE Transactions on*, vol. 19, no. 5, pp. 842–853, Oct. 2003.

[9] G.-Q. Wei, K. Arbter, and G. Hirzinger, "Real-time visual servoing for laparoscopic surgery. controlling robot motion with color image segmentation," *Engineering in Medicine and Biology Magazine, IEEE*, vol. 16, no. 1, pp. 40–45, Jan.-Feb. 1997.

[10] A. Casals, J. Amat, and E. Laporte, "Automatic guidance of an assistant robot in laparoscopic surgery," *Robotics and Automation, IEEE International Conference on*, vol. 1, pp. 895–900 vol.1, Apr 1996.

[11] C. E. Reiley and G. D. Hager, "Task versus subtask surgical skill evaluation of robotic minimally invasive surgery," in *MICCAI (1)*, 2009, pp. 435–442.

[12] H. C. Lin, I. S. ans David Yuh, and G. D. Hager, "Towards automatic skill evaluation: Detection and segmentation of robot-assisted surgical motions," *Computer Aided Surgery*, vol. 11, no. 5, pp. 220–230, September 2006.

[13] D. R. Uecker, C. Lee, Y. F. Wang, and Y. Wang, "Automated instrument tracking in robotically-assisted laparoscopic surgery," *Journal of Image Guided Surgery*, vol. 1, pp. 308–325, 1998.

[14] C. Doignon, F. Nageotte, and M. de Mathelin, "The role of insertion points in the detection and positioning of instruments in laparoscopy for robotic tasks," in *MICCAI*, 2006, pp. 527–534.

[15] S. Voros, J.-A. Long, and P. Cinquin, "Automatic detection of instruments in laparoscopic images: A first step towards high-level command of robotic endoscopic holders," *The International Journal of Robotics Research*, vol. 26, no. 11-12, pp. 1173–1190, 2007.

[16] D. Burschka, J. J. Corso, M. Dewan, W. W. Lau, M. Li, H. C. Lin, P. Marayong, N. A. Ramey, G. D. Hager, B. Hoffman, D. Larkin, and C. J. Hasser, "Navigating inner space: 3-d assistance for minimally invasive surgery," *Robotics and Autonomous Systems*, vol. 52, no. 1, pp. 5–26, 2005.

[17] Z. Pezzementi, S. Voros, and G. D. Hager, "Articulated object tracking by rendering consistent appearance parts," in *Robotics and Automation, 2009. ICRA '09. IEEE International Conference on*, May 2009, pp. 3940–3947.

[18] G. Panin, A. Ladikos, and A. Knoll, "An efficient and robust real-time contour tracking system," *Computer Vision Systems, International Conference on*, vol. 0, p. 44, 2006.

[19] G. Panin, E. Roth, and A. Knoll, "Robust contour-based object tracking integrating color and edge likelihoods," in *VMV*, 2008, pp. 227–234.

[20] C. Staub, A. Knoll, T. Osa, and R. Bauernschmitt, "Autonomous high precision positioning of surgical instruments in robot-assisted minimally invasive surgery under visual guidance," in *Autonomic and Autonomous Systems, IEEE International Conference on*, vol. 0. IEEE Computer Society, March 2010, pp. 64–69.

[21] A. Ruf, M. Tonko, R. Horaud, and H.-H. Nagel, "Visual tracking of an end-effector by adaptive kinematic prediction," in *Intelligent Robots and Systems, 1997. IROS '97., Proceedings of the 1997 IEEE/RSJ International Conference on*, vol. 2, Sep 1997, pp. 893–899 vol.2.

[22] L. Itti and C. Koch, "Computational modelling of visual attention." *Nature reviews. Neuroscience*, vol. 2, no. 3, pp. 194–203, March 2001.