Product of Tracking Experts for Visual Tracking of Surgical Tools

Suren Kumar, Madusudanan Sathia Narayanan, Pankaj Singhal, Jason J. Corso and Venkat Krovi
State University of New York (SUNY) at Buffalo
{surenkum,ms329,psinghal,jcorso,vkrovi}@buffalo.edu

Abstract— **This paper proposes a novel tool detection and tracking approach using *uncalibrated monocular surgical videos* for computer-aided surgical interventions. We hypothesize surgical tool end-effector to be the most distinguishable part of a tool and employ state-of-the-art object detection methods to learn the shape and localize the tool in images. For tracking, we propose a Product of Tracking Experts (PoTE) based generalized object tracking framework by probabilistically-merging tracking outputs (probabilistic/non-probabilistic) from time-varying numbers of trackers. In the current implementation of PoTE, we use three tracking experts – point-feature-based, region-based and object detection-based. A novel point feature-based tracker is also proposed in the form of a voting based bounding box geometry estimation technique building upon point-feature correspondences. Our tracker is causal which makes it suitable for real-time applications. This framework has been tested on real surgical videos and is shown to significantly improve upon the baseline results.**

I. INTRODUCTION

The field of surgical robotics had witnessed tremendous advancements over the last decade, transforming both the operating surgical teams as well as the operating rooms. Increasingly surgeries are being performed by teleoperated devices on patients with remote manipulators through small incisions while providing the surgeon at master end with 'look-and-feel' of an open surgery. Such robotic laparoscopic (or minimally invasive) procedures result in minimal pre- and post- surgical trauma and faster recovery for the patients. However, at the current stage, this robotic surgery paradigm does not adequately leverage the 'information-assist' possible by analyzing the enormous data-stream generated from endoscopic cameras and recorded manipulator motions.

At the same time, adoption of the robotic-surgery continues to raise serious questions about shortcomings in patient safety and surgical training. Major causes of concern including reduced field of view (FoV), loss of depth perception, lack of force-feedback and more importantly, lack of surgical training and assessment still remain unsolved [22]. With increasing number of legal claims due to robotic surgical failures and relatively outdated surgical training curriculum, the current situation in robotic surgeries can only be expected to worsen. Therefore, devising robust, feasible and advanced technologies for enhancing surgical safety and aiding the decision-support for surgeons without requiring major modifications to existing system has received greater attention.

Recently there has been significant interest in video understanding techniques applied to recorded or online surgical video streams (for use in anatomic reconstruction, surface registration, hand and/or tool motion tracking, etc.). Such techniques have potential use in providing in-vivo surgical guidance information and semantic feedback to surgeons, thus improving their visual awareness and target reachability, thereby enhancing the overall patients safety in robotic surgeries. The complexities posed by typical surgical scenarios (such as tissue deformations, image specularities and clutter, tool open/closed states, occlusion of tools due to blood and/or organs and tool out-of-view) offer the usual constraints hindering implementation of a robust video-based tool tracking method.

A few authors have begun exploration of video-analysis and understanding methods to enhance the interactions. Jun et al. [13] discuss how video-based motion analysis methods can be useful in surgical expertise evaluation. Voros et al. [24] use tool location data for identifying surgical gestures and providing context specific feedback. These approaches show promise for developing a unified plan to tackle a range of challenges in modern surgical robotics workflow based on reliable video understanding. In this work, we are motivated by the diverse perspectives to develop a novel, robust and efficient tool tracking solution for use in challenging real surgical videos.

II. RELATED WORK

Tracking surgical tools in general has been used for a wide range of applications including safety, decision-support as well as skill assessment. Most tool tracking approaches are either colour marker based or based on the geometric model of the instrument . Former techniques employ fiducial markers on the tool, using color marker and thresholding in HSV space to detect tools [11], attaching light emitting diodes to the tip of instruments and then detecting these markers in endoscopic images [15], color coding tips of surgical instruments and using a simple color segmentation [25]. Using such marker-based methods for surgical tool tracking has issues with manufacturing, bio-compatibility and additional instrumentation. While geometry based approaches use knowledge of the model of tool to find its pose from images [7].

Other approaches (that do not necessarily modify the tool itself) include: using color space for classification of pixel into instruments and organs, performing shape analysis of these classification labels and then predicting the location of tool in next frame using an Auto Regressive Model [23]. McKenna et al. [19] use similar method for classification of surgical tools but use particle filter to track instruments in a video. These approaches are limited to detecting a tool

when a significant area of tool is present in an image frame and there is a good distinction between its background and instruments in color space. Other recent work focuses on locating specific landmarks on the surgical tools by learning a Random Forest based classifier to classify these landmarks from images and using an Extended Kalman Filter (EKF) to smooth the tool poses [21]. However this method requires knowledge of 3D CAD model and extensive image labeling for a single tool.

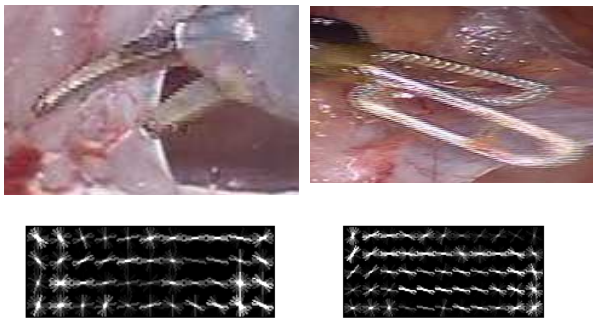
For surgical tool tracking to succeed and be widely employed in a typical surgical setting, there are several key challenges that need to addressed as summarized below:

- *Tool Detection:* Tool tracking approaches need to robustly determine the presence of different surgical tools in images as surgeons move their tools in- and out-of the FoV of an endoscopic camera. It becomes important for a tracking framework to incorporate this knowledge to reduce the number of false alarms. This is a critical problem especially in markerless tracking as color segmentation [11], [25], [7] will produce outliers in tool end-effector detection in presence of tool-tissue interaction, blood stains and many other factors. We address this problem robustly by learning state-of-the art object detectors [9] for different tool types.
- *End-Effector Pose:* Some model based approaches [7] ignore the pose of end-effector while tracking the tool. Since, end-effector in many surgical tools is articulated and always the point of contact with tissues, it is vital to track the end-effector and its articulation in the tracking framework. Our approach to detection and tracking directly models end-effector which is in general most distinguishable part of a surgical tool and employ a detection method that captures its articulation.
- *Generalized Approach:* The tracking algorithm also needs to be generalizable to different types of tools used in various surgical procedures. Model based approaches which model end-effector are very specific [21] to a particular surgical tool. In contrast, our approach is easily generalizable as it only needs annotated bounding boxes to learn a detector for that specific tool [9] and tracking is invariant to tool-types.
- *Tool Tracking Framework:* Different tool-tracking methods have been proposed to solve the problem with clear trade-off. The important issue that needs to be addressed is the effective combination of trackers that can optimally combine the strengths of various methods.

We aim to model the tracking task independent of features/ types of individual tracker and focus on optimally fusing the information from all the available trackers. Hence, we propose a Product of Tracking Experts (PoTE) based generalized object tracking framework which probabilistically merges tracking outputs from time-varying number of trackers to produce robust identity-maintained tracking under varied conditions in unconstrained scenarios.

III. System Overview

Our aim is to achieve robust tracking in challenging testing scenarios in a causal way i.e. not using any information from future frames. To address the challenge of detecting presence of surgical tool in images, we learn different detectors for each type of surgical tool end-effectors using state-of-the-art object detector [9]. This object detector essentially captures the shape of the object by using Deformable Part Models (DPM) consisting of star-structured pictorial structure model which links root of an object to its parts using deformable springs. Hence this model captures articulation which is invariably present in the surgical tool and allows for learning a detector for different tool end-effector configurations. We annotated surgical tools in real surgical videos obtained from da Vinci Surgical System (dVSS) and learn a Latent Support Vector Machine (LSVM) [5] classifier by extracting Histogram of Oriented Gradients (HOG) [6] from annotated bounding boxes. This type of learning classifiers is highly generalizable and extensible, enabling one to find tools in videos without making any restrictive assumptions about the type, shape, color of tool, view etc. Figure 1 shows HOG



(c) *Tool* **(d) *Clamp***

Fig. 1: Learned HOG Templates with a representative bounding box for different tool types

template model learned using ground truth annotations.

A high level flow chart of proposed framework is shown in Figure 2. Our method bootstraps from high confidence detection to start tracks of various tools. We empirically obtain a very low false positive rate by increasing the detection threshold on learned object detector for different tools.

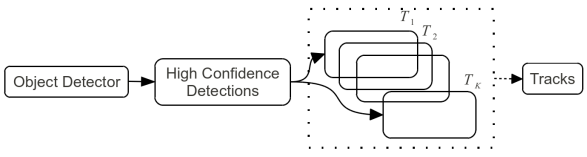


Fig. 2: System Flow Diagram

Each entity is tracked independently by various trackers $T_1, T_2, ..., T_K$. These trackers could be based on either discriminative (data association techniques [10],detector confidence etc.), generative (particle filter [20], Kalman Filter

[26], KLT [17] etc.), model based or a combination of generative and discriminative techniques [14], [3]. Tracking solely by using either generative or discriminative approaches in unconstrained scenes is hard because generative approaches make assumptions about the motion of entity whereas discriminative approaches make assumption of having a robust detector.

IV. Product of Tracking Experts

We adapt a time evolving Product of Experts (PoE) [12] model to optimally fuse hypothesis from various trackers at each instant in time to propose a Product of Tracking Experts (PoTE). We consider each tracker $T_1, T_2, ..., T_K$ as experts for predicting the location of target center. Product of experts model for tracker ensures that the resulting model for track is explained by all the experts. Let θ_k be the parameters associated with probability distribution of each expert ($= [\mu^k, \Sigma^k]^T$ in current case). Probability of any point $\underline{x}$ to be true center of a bounding box as explained by all the expert trackers is given by Equation 1.

$$p(\underline{x}|\theta_{T_1}, \theta_{T_2}, ..., \theta_{T_K}) = \frac{\prod_{k=1}^{K} p_k(\underline{x}|\theta_k)}{\int \prod_{k=1}^{K} p_k(\underline{x}|\theta_k) \mathrm{d}\underline{x}} \tag{1}$$

Denominator in Equation 1 is a normalization constant and can be ignored to choose best $\underline{x}$. This model is very useful for robust tracking because it allows to incorporate (or leave out) arbitrary number of trackers. For example, for a discriminative classifier, detection score is commonly used to guide tracking. This classifier could be included in the tracking mix by modeling its distribution using an indicator function, which determines if this detection score is greater than a predetermined threshold. Tracking frameworks using particle/Kalman filter provide a probability distribution as output which is very suitable for this method. Additionally, any individual tracker is not required to give a probabilistic output but is only required to give a bounding box which could then be modeled as a probability distribution using Equation 4. Breitenstein et al.[3] propose using continuous class confidence density because current object detectors such as HOG [6] provide a score at discrete spatial locations and scale, this could be easily incorporated in the presented framework. Hence, our proposed framework is highly general method of combining results from different types of trackers and easily extensible.

If all the experts have normal distribution with $\underline{\mu_k}, \mathbf{\Sigma_k}$ as mean and covariance matrix, the resulting best location of center of bounding box $\underline{x}$ can be obtained analytically because product of normal distributions yields a normal distribution.

$$p(\underline{x}|\theta_{T_1}, \theta_{T_2}, ..., \theta_{T_K}) = \frac{\prod_{k=1}^{K} \frac{1}{2\pi|\mathbf{\Sigma_k}|^{\frac{1}{2}}} \exp(-\frac{1}{2}[\underline{x} - \underline{\mu_k}]^T \mathbf{\Sigma_k}^{-1}[\underline{x} - \underline{\mu_k}])}{\int \prod_{k=1}^{K} p_k(\underline{x}|\theta_k) \mathrm{d}\underline{x}} \tag{2}$$

The resulting probability density function (pdf) can be obtained after some algebraic manipulation as

$$p(\underline{x}|\theta_{T_1}, \theta_{T_2}, ..., \theta_{T_K}) \sim \mathcal{N}(\underline{\mu}, \mathbf{\Sigma}), \text{ where}$$
$$\mathbf{\Sigma}^{-1} = \sum_{k=1}^{K} \mathbf{\Sigma_k}^{-1}, \underline{\mu} = \mathbf{\Sigma} \left(\sum_{k=1}^{K} \mathbf{\Sigma_k}^{-1} \underline{\mu_k} \right) \tag{3}$$

Intuition behind this model is shown in Figure 3. In the current implementation, we focus on a PoTE based generative - discriminative tracking method. We exploit combination of interest points, dense optical flow for tool end-effector tracking from low-level tool detection algorithms. Bounding box produced by each tracker (BB^k) is represented as a 2D spatial Gaussian $BB^k \sim \mathcal{N}(\mu^k, \Sigma^k)$. We hypothesize location of the bounding box to be normally distributed in the image plane with its centroid $[x^k_{CB}, y^k_{CB}]^T$ as mean, and width (w_B) and height (h_B) as uncertainty in its location (6 $\times$ variance).

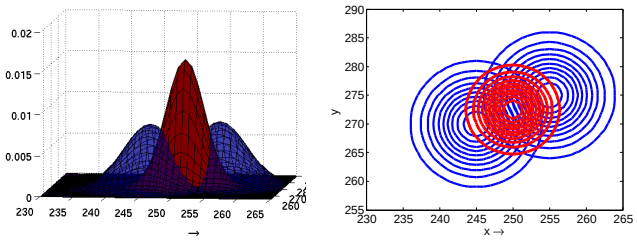


Fig. 3: Two tracking experts on sides and resulting POTE model in middle with Gaussian pdfcontours on right. One tracker has associated Gaussian with mean = $[245, 270]^T$ and variance = $diag([16.66, 25])$, second tracker has associated Gaussian with mean = $[255, 275]^T$ with same variance, when combined using PoTE model results into a Gaussian with mean = $[250, 272.5]^T$ and variance =$diag([8.33, 12.5])$

$$\mu^k = [x^k_{CB}, y^k_{CB}]^T, \Sigma^k = \frac{1}{6} \begin{bmatrix} w^k_B & 0 \\ 0 & h^k_B \end{bmatrix} \tag{4}$$

Hence, outputs of point tracking and dense optical flow tracking is converted to probability distributions using Equation 4. DPM based detector (det) provides many bounding boxes along with their corresponding detection score τ as output. A detection is included in tracking experts mix only if it is deemed reliable, which is evaluated by an indicator function. This indicator function $\mathcal{I}$ is defined in Equation 5 as

$$\mathcal{I} = \begin{cases} 1 & \text{if}(\tau \geq \tau_{thresh}) \wedge (BB_{det} \bigcap BB^e_{t-1}) \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

In Equation 5, τ_{thresh} is a predetermined threshold on detection scores for including only reliable detections. Additionally, only relevant detections are considered to track a tool e on current frame (time t) by evaluating whether a particular detector predicted bounding box BB_{det} intersects with bounding box of tool end-effector in last frame (time $t-1$). If multiple detections have indicator function $\mathcal{I}$ as 1, we select the bounding box which has the maximum score.

Additional ways of selecting a bounding box given by a tool detector could be based on velocity and size information as proposed by [3]. Once a particular detection bounding box is selected, the detector is modeled as an expert by using Equation 4. The size of the finally generated track is selected as the size of last associated detection.

The key benefits in our current implementation ensue from complementary nature of the two constituent probabilistically merged approaches – Point feature based tracking, which is robust for small motion and Region based tracking, which works well in case of significant motion. As a result, our tracker shows robust tracking for scenarios which involve unconstrained surgical tool activities as shown in Figure 5. We now describe individual experts apart from tool detection in our current implementation.

A. Point Feature Based Tracking

Kanade-Lucas-Tomasi (KLT) first introduced by Lucas and Kanade [17], is a point feature tracker extensively used for computer vision tasks. This algorithm finds good spatial features to track by locating Harris corners in an image. To track a particular feature, a window centered on feature point in current image is matched in next image by Newton-Raphson method of minimization. This can be made robust by performing the matching across an image pyramid. We use Stan Birchfield's [2] implementation of KLT to achieve tracking of feature points. KLT based tracker needs an initialization from tool detection to identify the region in current image that has to be tracked in subsequent images. Process of tracking using KLT is pictorially depicted in Figure 4.

To initialize the KLT tracker, we evaluate feature points inside the bounding box in current frame. Assuming $[x_B(t), y_B(t), w_B(t), h_B(t)]$ is the axis aligned bounding box, where $[x_B(t), y_B(t)]^T$ is left top corner of bounding box in the image and $(w_B(t), h_B(t))$ specify width and height of the bounding box respectively at frame t. Geometric location of each feature point is obtained relative to top left point of the bounding box, thus encoding relative geometrical location of all the features with respect to bounding box.

$$\begin{aligned} G_x(t,j) &= (x_B(t) - x_f(t,j)), \\ G_y(t,j) &= (y_B(t) - y_f(t,j)) \end{aligned} \tag{6}$$

$G_x(t,j), G_y(t,j)$ stores relative (x,y) location of j th feature on frame t. In second step, these features are tracked in next image using KLT. Since robotic surgical tools are highly articulated systems, only a part of original feature points are tracked and considered after this step.

$$\begin{aligned} x_B(t+1,j) &= (G_x(t,j) + x_f(t+1,j)), \\ y_B(t+1,j) &= (G_y(t,j) + y_f(t+1,j)) \\ x_B(t+1) &= \sum_{j=1}^{J} w_j x_B(t+1,j), \\ y_B(t+1) &= \sum_{j=1}^{J} w_j y_B(t+1,j) \end{aligned} \tag{7}$$

where w_j is the weight associated with each feature point as obtained from the normalized objective residual in KLT minimization such that $\sum_{j=1}^{J} w_j = 1$. Each tracked feature carries the geometrical relationship from previous frame and votes for current location of bounding box by assuming that collection of large number of features can diminish the effect of noise in location of bounding box using (7), where $(x_B(t+1,j), y_B(t+1,j))$ is location of the top-left corner of the bounding box as predicted by j th feature on frame $t+1$. Width and height of bounding box are updated based on the tracking output from the previous time step.

B. Dense Optical Flow

This tracker is based on extracting dense optical flow [4] and predict the bounding box in next frame. Optical flow measures apparent motion of a pixel between two images assuming that its brightness remains constant in both images. We start tracking by using the detections with confidence measure above a given threshold τ_{thresh}. In each frame, we obtain the optical flow between two frames for all the pixels belonging to the desired bounding box. The location of bounding box in next frame is a result of flow in all the pixels and is approximated by the mean flow of all the pixels. Width and height of the bounding box is updated based on tracking output from the previous associated detection with current tool end-effector track.

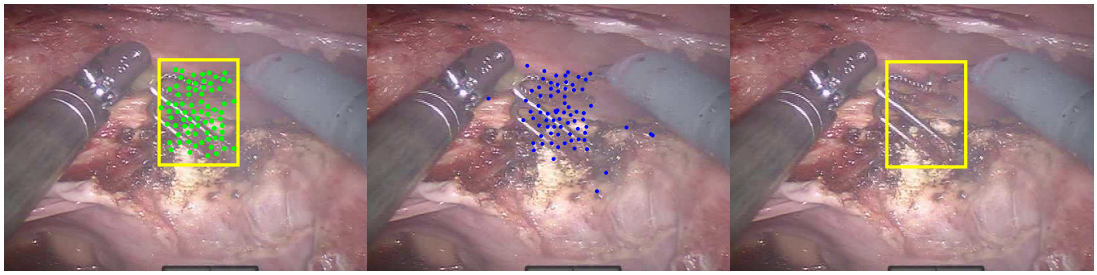
V. Experiments

We propose a new dataset consisting of 8 small sequences (1500 frames) for "Clamp" class and 8 sequences (1650 frames) for "Tool" class acquired while performing Hysterectomy surgery using dVSS to conduct our evaluation. To the best of our knowledge, there are no publicly available datasets for testing our tool tracking algorithm. The proposed dataset has real-world video-sequences with various artefacts including tool articulations, occlusions, rapid appearance changes, fast camera motion, motion blur, smoke and specular reflections. This dataset was then manually annotated for the bounding boxes of the tools in every frame. The overall accuracy of our PoTE method is then evaluated using standard performance measures [18] by calculating True Positive (TP), False Positive (FP), True Negatives (TN) and False Negatives (FN). We treat a bounding box in image to be True Positive if the pascal measure (ratio of area intersection and area union) in image frame is greater than 0.5, which is commonly used for measuring accuracy of object detection detection methods [8]. We test a baseline tracker using detection and KLT tracking for both tool classes on the proposed dataset and PoTE tracker with detection, optical flow and KLT as experts. As shown in Table I, our algorithm

Tool Type	Baseline	PoTE
Clamp	68.27%	**75.81%**
Tool	28.04%	**63.31%**

TABLE I: Accuracy using Baseline and PoTE tracker

outperforms the baseline method on this challenging dataset



(a) ***Step 1 - Corner Detection*** (b) ***Step 2 - Feature Tracking*** (c) ***Step 3 - Reconstruction***

Fig. 4: Flow diagram of tracking using KLT

for both the tool types. Baseline method's performance worsens on "Tool" class because of rapid perceived motion associated camera pose/zoom changes, articulation and tool motion. "Clamp" is usually kept stationary in surgery to hold the tissue while "Tool" is used to perform tissue cutting as can be seen from results in Figure 5. Additionally, the proposed point feature based tracking method is only suitable for rigid motion in image frame as can be observed in reconstruction step in Equation 7. We will release the annotated dataset and our code upon publication to encourage further research into this problem.

VI. DISCUSSION

This paper proposed a novel tool detection and tracking approach using uncalibrated monocular surgical videos for computer-aided surgical interventions. The resulting detection and tracking PoTE framework gives good results by probabilistically-merging tracking outputs. This framework has been tested on real surgeries and shows improvement upon the baseline results. In our future work, we plan to investigate hierarchical coarse-to-fine flow techniques [1] and feature matching techniques [16] that can handle rapid motions for incorporation into PoTE model.

ACKNOWLEDGMENT

This work was partially supported by the National Science Foundation CPS-MEDIUM Grant (CNS-1314484) and the DARPA Mind's Eye program (W911NF-10-2-0062).

REFERENCES

[1] J. Bergen, P. Anandan, K. Hanna, and R. Hingorani. Hierarchical model-based motion estimation. In *ECCV*, 1992.
[2] S. Birchfield. KLT: An implementation of the Kanade-Lucas-Tomasi feature tracker. Available: http://www.ces.clemson.edu/ stb/klt/.
[3] M.D. Breitenstein, F. Reichlin, B. Leibe, E. Koller-Meier, and L. Van Gool. Online multiperson tracking-by-detection from a single, uncalibrated camera. *PAMI*, 33(9):1820–1833, 2011.
[4] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, 2011.
[5] C. Cortes and V. Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.
[6] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *CVPR*, volume 1, pages 886–893, 2005.
[7] C Doignon, F Nageotte, and M de Mathelin. Segmentation and guidance of multiple rigid objects for intra-operative endoscopic vision. *Dynamical Vision*, pages 314–327, 2007.
[8] P. Dollár, C. Wojek, B. Schiele, and P. Perona. Pedestrian detection: A benchmark. In *CVPR*, pages 304–311, 2009.
[9] P.F. Felzenszwalb, R.B. Girshick, and D. McAllester. Cascade object detection with deformable part models. In *CVPR*, pages 2241–2248. IEEE, 2010.
[10] T. Fortmann, Y. Bar-Shalom, and M. Scheffe. Sonar tracking of multiple targets using joint probabilistic data association. *IEEE Journal of Oceanic Engineering*, 8(3):173–184, 1983.
[11] M. Groeger, K. Arbter, and G. Hirzinger. Motion tracking for minimally invasive robotic surgery. *Medical Robotics, I-Tech Education and Publishing*, pages 117–148, 2008.
[12] Geoffrey E. Hinton. Products of experts. In *ICANN*, pages 1–6, 1999.
[13] S.K. Jun, M.S. Narayanan, P. Agarwal, A. Eddib, P. Singhal, S. Garimella, and V. Krovi. Robotic minimally invasive surgical skill assessment based on automated video-analysis motion studies. In *BIOROB*, pages 25–31. IEEE, 2012.
[14] Z. Kalal, K. Mikolajczyk, and J. Matas. Tracking-learning-detection. *PAMI*, (99), 2011.
[15] A. Krupa, J. Gangloff, C. Doignon, M.F. 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. *Trans. on Rob. and Auto.*, 19(5):842–853, 2003.
[16] David G Lowe. Object recognition from local scale-invariant features. In *ICCV*, volume 2, pages 1150–1157. Ieee, 1999.
[17] B.D. Lucas, T. Kanade, et al. An iterative image registration technique with an application to stereo vision. In *IJCAI*, 1981.
[18] J. Makhoul, F. Kubala, R. Schwartz, and R. Weischedel. Performance measures for information extraction. In *Proc. DARPA Broadcast News Workshop*, 1999.
[19] S.J. McKenna, H.N. Charif, and T. Frank. Towards video understanding of laparoscopic surgery: Instrument tracking. In *Proc. of Image and Vision Computing, New Zealand*, 2005.
[20] K. Okuma, A. Taleghani, N. Freitas, J.J. Little, and D.G. Lowe. A boosted particle filter: Multitarget detection and tracking. *ECCV*, 2004.
[21] A. Reiter, P. Allen, and T. Zhao. Feature classification for tracking articulated surgical tools. *MICCAI*, pages 592–600, 2012.
[22] F. Soleimani, F. Moll, D. Wallace, J. Bismuth, and B. Geršak. Robots and medicine–shaping and defining the future of surgery, endovascular surgery, electrophysiology and interventional radiology. *Slovenian Medical Journal*, 80(7-8), 2011.
[23] D.R. Uecker, YF Wang, C. Lee, and Y. Wang. Laboratory investigation: Automated instrument tracking in robotically assisted laparoscopic surgery. *Computer Aided Surgery*, 1(6):308–325, 1995.
[24] S. Voros and G. D. Hager. Towards real-time tool-tissue interaction detection in robotically assisted laparoscopy. In *BioRob*. IEEE, 2008.
[25] G.Q. Wei, K. Arbter, and G. Hirzinger. Automatic tracking of laparoscopic instruments by color coding. In *CVRMed-MRCAS'97*, pages 357–366. Springer, 1997.
[26] G. Welch and G. Bishop. An introduction to the kalman filter. *Design*, 7(1):1–16, 2001.

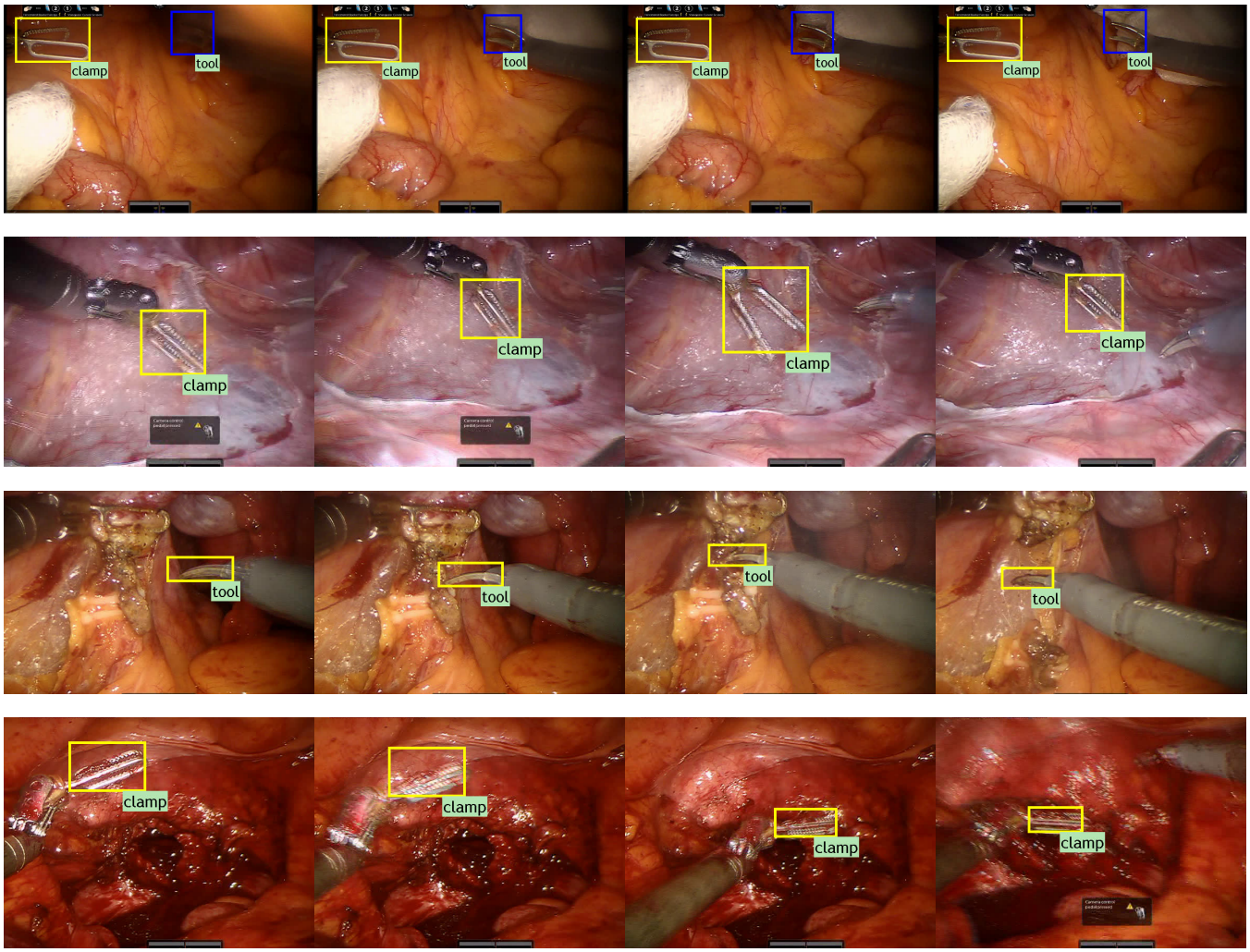


Fig. 5: Tracking results for “Tool” and “Clamp” on various surgical operation videos in proposed dataset. (Please view in color)