EKF monocular SLAM with relocalization for laparoscopic sequences

Oscar G. Grasa, Javier Civera and J. M. M. Montiel

Abstract— **In recent years, research on visual SLAM has produced robust algorithms providing, in real time at 30 Hz, both the 3D model of the observed rigid scene and the 3D camera motion using as only input the gathered image sequence. These algorithms have been extensively validated in rigid human-made environments –indoor and outdoor– showing robust performance in dealing with clutter, occlusions or sudden motions.**

Medical endoscopic sequences naturally pose a monocular SLAM problem: an unknown camera motion in an unknown environment. The corresponding map would be useful in providing 3D information to assist surgeons, to support augmented reality insertions or to be exploited by medical robots. In this paper we propose the combination EKF Monocular SLAM + 1-Point RANSAC + Randomised List Relocalization to process laparoscopic sequences –abdominal cavity images–. The sequences are challenging due to: 1) cluttering produced by tools; 2) sudden motions of the camera; 3) laparoscope frequently goes in and out of abdominal cavity; 4) tissue deformation caused by respiration, heartbeats and/or surgical tools. Real medical image sequences provide experimental validation.

I. INTRODUCTION

SLAM (Simultaneous Localization and Mapping) is a classical problem in mobile robotics: let be a mobile sensor following an unknown trajectory while observing an unknown environment; the goal is to estimate simultaneously both the environment structure and the sensor location with respect to that map. Recent SLAM research has focused on monocular cameras as the unique sensorial input. 30 Hz real-time systems estimating full 3D camera motions and maps of 3D points using commodity cameras are widely available nowadays.

On the other hand, laparoscopic –endoscopy inside the abdominal cavity– techniques have been one of the major advances in surgery, entailing numerous advantages for patients (shorter hospitalization and rehabilitation time, reduced morbidity and less aesthetic impact than traditional surgery) and for surgeons (shorter time and safer operations). The purpose of the paper is to apply monocular visual SLAM methods to analyse laparoscopic images. One of the main weakness of current SLAM methods is its dependence on scene rigidity, this justifies the focus on laparoscopic scenes that are mainly rigid except for temporary deformation.

EKF SLAM has been successfully applied in [1] to deal with laparoscopic sequences without motion clutter nor sudden camera motions, and little cavity deformation. Our contribution is to adapt the state-of-the-art EKF monocular SLAM [2], [3] + 1-Point RANSAC [4] + Randomised list relocalization [5] combination in order to cope with real laparoscopic sequences under challenging conditions including sudden camera motions, laparoscope extraction and reinsertion and high spurious rate caused by temporary tissue deformation or surgical tools cluttering. In-vivo laparoscopic sequences of a human ventral hernia repair provide experimental validation for the proposed method.

EKF based monocular SLAM proposed in [2], [3] was the first method to show real time performance. Scene rigidity and smooth camera motion were assumed. Experimental performance was reported on man made rigid scenes. Inverse depth (ID) initialization proposed in [6], [7] improved the orientation estimation and hence the overall system performance. In [8] the scene rigidity is tightly enforced by means JCBB test (Joint Compatibility Branch and Bound) [9], resulting in an improved robust performance and the ability to remove a moderate level of motion clutter in the scene. More recently, in [4] RANSAC was combined with EKF resulting in a boost in the capability to detect and reject a high number of spurious in real time. This is exploited in the current paper to deal with the high spurious rates coming from the temporal cavity deformation and tools motion clutter.

A known weakness of EKF visual SLAM is its lack of ability to recover from a lost of track, mainly due to camera sudden motion. In [5] it is proposed a randomized list approach to relocate the camera with respect to the map after a severe track lost, this new addition greatly improved the performance. We propose to apply this relocation method to deal with lost of track due to sudden laparoscope motion and also to cope with laparoscope reinsertion inside the abdominal cavity. In [10] and [11] relocation methods based on reduced resolution keyframe images are proposed, we find that feature based ones such as [5] are more expensive but with better ability to deal with scene motion clutter, what justifies our selection.

In a seminal work [12], Burschka et al. proposed the first monocular SLAM system that processes medical images. The system produces a map composed of a reduced number of points in sinus surgery, enough for registering preoperative CT scan with the images gathered by the endoscope. In [13] and [14] a non moving stereo endoscopes provides 3D structure of a moving organ. Moving stereo endoscopes have been successfully used in visual SLAM in [15] assuming smooth camera motion and scene rigidity validated over real sequences, no usage of spurious detection is reported. In [1] it is used EKF+JCBB SLAM to process real moving monocular laparoscopic sequences; however, the system can not

Supported by Spanish Ministerio Ciencia e Innovación grant DPI2009-07130

Oscar G. Grasa, Javier Civera and J. M. M. Montiel are with Instituto de Investigación e Ingeniería de Aragón, Universidad de Zaragoza, Spain.{**oscgg, jcivera, josemari**}**@unizar.es**

cope with sudden camera motions or laparoscope reinsertion. In [16], a dynamic model for coding periodic organ motion is learnt, introducing the scene non-rigidity in the model.

Computer vision methods based on a discrete set of images –instead of image sequences as reported above– have been also applied to medical images, assuming scene rigidity, in order to just compute the 3D structure of the cavity. In [17], the classical two view RANSAC structure from motion is applied to mannequin images to determine the 3D structure; a constraint-based factorization 3D modelling method produces a dense 3D reconstruction in near real-time. Finally, structure from motion is used in [18] to build a photorealistic 3D reconstruction of the colon; in a first stage images are processed pairwise to produce an initial 3D map, in a second stage all the maps are joined in a unique photorealistic 3D cavity model.

II. MONOCULAR SLAM

Our SLAM system builds on the inverse depth EKF SLAM proposed in [7]. It is based on a probabilistic representation of the world map and the camera location maintained in an unique state vector modelled as a multivariate Gaussian, $\mathbf{x}$, updated by an Extended Kalman Filter (EKF).

The state vector at step k is composed of camera location and velocity, $\mathbf{x}_v$, and all points in the map, $\mathbf{y}_i$:

$$\mathbf{x} = \left(\mathbf{x}_v^\top, \mathbf{y}_1^\top, \mathbf{y}_2^\top, \dots \mathbf{y}_n^\top\right)^\top \quad . \tag{1}$$

The camera state, $\mathbf{x}_v$, is formed from position, $\mathbf{r}^{WC}$, orientation encoded in a quaternion, $\mathbf{q}^{WC}$, and linear and angular velocities, $\mathbf{v}^W$ and ω^C:

The map is composed of n point features $(\mathbf{y}_1^\top \cdots \mathbf{y}_n^\top)^\top$ which are coded either in Euclidean coordinates, $\mathbf{y}_i = \begin{pmatrix} X_i & Y_i & Z_i \end{pmatrix}^\top$, or in inverse depth (ID), $\mathbf{y}_i = \begin{pmatrix} x_i & y_i & z_i & \theta_i & \phi_i & \rho_i \end{pmatrix}^\top$, where x_i, y_i and z_i represent the 3D camera position when the feature was observed for first time, θ_i and ϕ_i code the ray corresponding to the observed point and ρ_i is the inverse depth along the ray.

We use a dynamic constant velocity model to code the camera smooth motion:

$$\mathbf{f}_v = \begin{pmatrix} \mathbf{r}_{k+1}^{WC} \\ \mathbf{q}_{k+1}^{WC} \\ \mathbf{v}_{k+1}^{W} \\ \omega_{k+1}^{C} \end{pmatrix} = \begin{pmatrix} \mathbf{r}_{k}^{WC} + \left(\mathbf{v}_k^W + \mathbf{V}_k^W\right)\Delta t \\ \mathbf{q}_k^{WC} \times \mathbf{q}\left(\left(\omega_k^C + \Omega^C\right)\Delta t\right) \\ \mathbf{v}_k^W + \mathbf{V}^W \\ \omega_k^C + \Omega^C \end{pmatrix} , \tag{2}$$

where $\mathbf{q}\left(\left(\omega_k^C + \Omega^C\right)\Delta t\right)$ is the quaternion defined from the rotation vector $\left(\omega_k^C + \Omega^C\right)\Delta t$. We assume that linear and angular accelerations $\mathbf{a}^W$ and α^C are unknown inputs producing at each step an impulse of linear velocity, $\mathbf{V}^W = \mathbf{a}^W \Delta t$, and angular velocity $\Omega^C = \alpha^C \Delta t$, with zero mean and known Gaussian covariance, currently assumed as diagonal. The scene points are coded as perfectly rigid:

$$\mathbf{y}_{i_{k+1}} = \mathbf{y}_{i_k} \tag{3}$$

The measurement model assumes a pinhole camera combined with a two parameters radial distortion.

New features initialization is directed to produce geometrically well conditioned maps. New features are initialized within an image search window randomly located but favouring less populated areas (image regions with few or no map features.) The strongest FAST [19] corners are sought inside the search window and the most distinctive one for relocation, according with RLR classifier (see section IV), is initialized and inserted in the map. Recently initialized features are parametrized in ID. As the estimation evolves and the features estimation improves, features with small uncertainty are converted to Euclidean, what reduces the state size and hence the EKF computation overhead.

Active search is applied to exploit the camera smooth motion and the scene rigidity during data association by means of the EKF prediction step. The innovation covariance 3σ gate defines an elliptical search window around the map point prediction where the match is going to be exhaustively searched using normalized image correlation. All map features are strongly correlated so prediction errors are not independent. Simple active search treats each match independently but does not ensure the global compatibility of the correspondences thus, incompatible matches can be fed in the EKF corrupting the map. In order to detect and remove incompatible matches 1-Point RANSAC has been used, details are given in next section III.

After EKF prediction step and previous to the match by correlation, an accurate estimation of the camera location with respect to every map point is available, this allows to accurately synthesize an appearance patch around the map point in the image, compensating rotation and scale what improves the matching performance and allows long reobservation tracks for the map points. This patch appearance compensation, frequent in SLAM systems, is also applied in our experiments.

III. SPURIOUS DETECTION

Abdominal tissues suffer temporary deformations caused by respiration, heartbeats, or surgical tools interaction. Furthermore, surgical tools clutter the scene occluding map features. These two problems result in a high spurious rate that must be coped with in real time.

JCBB [9] within EKF framework, proposed in [8], has been successfully used in the robotic community. However, JCBB presents two main limitations. First, it operates over the prediction of the measurements derived from a linearized dynamic model, a measurement model and Gaussian assumptions, whereupon it does not correspond with the true state of the system. Second, the JCBB presents an exponential computational complexity in the number of spurious hence, although it produces small CPU overhead with small spurious rates, it becomes intractable at high spurious rates. As we need to cope with high spurious rate, we select the 1-point RANSAC spurious detector [4].

In traditional RANSAC usage, model proposals are computed from scratch, selecting a minimum set of measurements. The complexity is exponential in the minimum set cardinality. Instead of computing the proposal from scratch

(cardinality 5), 1-point RANSAC algorithm combines the prior estimate produced by the Kalman filter plus just one measurement to compute a proposal (cardinality 1). It renders a spurious rejection low computation overhead – about 10% of the EKF cost–, even when facing high spurious rate. The EKF and RANSAC combination is detailed in algorithm 1.

Algorithm 1 1-Point RANSAC EKF

IN: $\hat{\mathbf{x}}_{k-1|k-1}, \mathbf{P}_{k-1|k-1}$ {EKF estimate at step $k-1$}
  th {Threshold for low-innovation points.}
OUT: $\hat{\mathbf{x}}_{k|k}, \mathbf{P}_{k|k}$ {EKF estimate at step k}

{A. EKF prediction and individually comp. matches}
$[\hat{\mathbf{x}}_{k|k-1}, \mathbf{P}_{k|k-1}] = EKF\_pred(\hat{\mathbf{x}}_{k-1|k-1}, \mathbf{P}_{k-1|k-1})$
$[\hat{\mathbf{h}}_{k|k-1}, \mathbf{S}_{k|k-1}] = measure\_pred(\hat{\mathbf{x}}_{k|k-1}, \mathbf{P}_{k|k-1})$
$\mathbf{z}^{IC} = search\_IC\_matches(\hat{\mathbf{h}}_{k|k-1}, \mathbf{S}_{k|k-1})$

{B. 1-Point hypotheses generation and evaluation}
$\mathbf{z}^{li\_inliers} = [\,]$
$n_{hyp} = 1000$ {Initial value. Updated in the loop}
for $i = 0$ to n_{hyp} **do**
  $\mathbf{z}_i = select\_random\_match(\mathbf{z}^{IC})$
  $\hat{\mathbf{x}}_i = EKF\_state\_update(\mathbf{z}_i, \hat{\mathbf{x}}_{k|k-1})$ {Only state; NO covariance}
  $\hat{\mathbf{h}}_i = predict\_all\_measurements(\hat{\mathbf{x}}_i)$
  $\mathbf{z}_i^{th} = find\_matches\_below\_a\_thres(\mathbf{z}^{IC}, \hat{\mathbf{h}}_i, th)$
  if $size(\mathbf{z}_i^{th}) > size(\mathbf{z}^{li\_inliers})$ **then**
    $\mathbf{z}^{li\_inliers} = \mathbf{z}_i^{th}$
    $\epsilon = 1 - \frac{size(\mathbf{z}^{li\_inliers})}{size(\mathbf{z}^{IC})}$
    $n_{hyp} = \frac{log(1-p)}{log(1-(1-\epsilon))}$
  end if
end for

{C. Partial EKF update using low-innovation inliers}
$[\hat{\mathbf{x}}_{k|k}, \mathbf{P}_{k|k}] = EKFupdate(\mathbf{z}^{li\_inliers}, \hat{\mathbf{x}}_{k|k-1}, \mathbf{P}_{k|k-1})$

{D. Partial EKF update using high-innovation inliers}
$\mathbf{z}^{hi\_inliers} = [\,]$
for every match $\mathbf{z}^j$ above a threshold th **do**
  $[\hat{\mathbf{h}}^j, \mathbf{S}^j] = point\_j\_pred\_and\_cov(\hat{\mathbf{x}}_{k|k}, \mathbf{P}_{k|k}, j)$
  $\nu^{\mathbf{j}} = \mathbf{z}^j - \hat{\mathbf{h}}^j$
  if $\nu^{\mathbf{j}\top} \mathbf{S}^{\mathbf{j}-\mathbf{1}} \nu^{\mathbf{j}} < \chi^2_{\mathbf{2,0.01}}$ **then**
    $\mathbf{z}^{hi\_inliers} = add\_match\_to\_inliers(\mathbf{z}^{hi\_inliers}, \mathbf{z}^j)$
  end if
end for
$[\hat{\mathbf{x}}_{k|k}, \mathbf{P}_{k|k}] = EKFupdate(\mathbf{z}^{hi\_inliers}, \hat{\mathbf{x}}_{k|k}, \mathbf{P}_{k|k})$

The 1-point RANSAC is merged with EKF to maximize the CPU efficiency. The algorithm has four main stages.

The first stage is the classical active search, including state and measurement EKF prediction by means the dynamic and measurement models (section II):

$$\begin{aligned} \hat{\mathbf{x}}_{k|k-1} &= \mathbf{f}_k(\hat{\mathbf{x}}_{k-1|k-1}) \\ \mathbf{P}_{k|k-1} &= \mathbf{F}_k \mathbf{P}_{k-1|k-1} \mathbf{F}_k^T + \mathbf{G}_k \mathbf{Q}_k \mathbf{G}_k^T \end{aligned} \quad (4)$$

$$\begin{aligned} \hat{\mathbf{h}}_i &= \mathbf{h}_i(\hat{\mathbf{x}}_{k|k-1}) \\ \mathbf{S}_i &= \mathbf{H}_i \mathbf{P}_{k|k-1} \mathbf{H}_i^T + \mathbf{R}_i \end{aligned} \quad (5)$$

where $\mathbf{F}_k$ is the Jacobian of $\mathbf{f}_k$ with respect to the state vector $\mathbf{x}_{k|k-1}$ at step k, $\mathbf{Q}_k$ is the covariance of the zero-mean Gaussian noise assumed for the dynamic model and $\mathbf{G_k}$ is the Jacobian of this noise with respect to the state vector $\mathbf{x}_{k|k-1}$. $\mathbf{H}_i$ is the Jacobian of the measurement $\mathbf{h}_i$ with respect to the state vector $\mathbf{x}_{k|k-1}$ and $\mathbf{R}_i$ is the covariance of the Gaussian noise assumed for each individual measurement. These measurement predictions are used for defining the search windows. The matches $\mathbf{z}_i$ are sought inside these windows. This stage produces the individually compatible matches: $\mathbf{z}^{IC} = (\mathbf{z}_1 \cdots \mathbf{z}_n)^T$.

The second stage is hypotheses generation and consensus. Updated state hypotheses, $\hat{\mathbf{x}}_i$, are generated based on a single randomly selected match, $\mathbf{z}_i$, out of $\mathbf{z}^{IC} = (\mathbf{z}_1 \cdots \mathbf{z}_n)^T$ and on the predicted state $\mathbf{x}_{k|k-1} \sim \mathcal{N}(\hat{\mathbf{x}}_{k|k-1}, \mathbf{P}_{k|k-1})$. A hypothesis $\hat{\mathbf{x}}_i$ is obtained by updating the EKF state processing the selected match $\mathbf{z}_i$ but without updating the state covariance –to avoid the most expensive step–. After that, hypothesis support is computed by projecting the updated state into the image and counting measurements inside a threshold defined by χ^2 and the measurement covariance. It it has been conservatively assumed that all the correlated error is corrected by integrating the selected point and hence all the error can be modeled as measurement error.

The third stage is the partial update using low-innovation inliers $\mathbf{z}^{li\_inliers}$, those supporting the most voted hypothesis. It is a full update including the covariance update:

$$\begin{aligned} \mathbf{K}_k &= \mathbf{P}_{k|k-1} \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^T + \mathbf{R}_k)^{-1} \\ \hat{\mathbf{x}}_{k|k} &= \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k (\mathbf{z}^{li\_inliers} - \mathbf{h}'(\hat{\mathbf{x}}_{k|k-1})) \\ \mathbf{P}_{k|k} &= (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_{k|k-1} \end{aligned} \quad (6)$$

In any case, some other inliers have been misclassified as outliers. The fourth stage is a second active matching process, like stage 1 but based on the last low-innovation update. Once all high-innovation inliers are rescued, $\mathbf{z}^{hi\_inliers}$, a second update including also the covariance is computed.

IV. RELOCALIZATION

Active search is one of the system strengths, it enables the system real time operation, but it is also one of its weakness. The system works fine provided that the mapped features are found inside the elliptical search window. However if the camera suffers from sudden motions, the image is blurred, or there are large occlusions tracking will fail because no features are matched within several frames.

The relocalization used in the paper – Randomized List Relocalization (RLR) – proposed in [5] is based on Randomized Trees [20]. The complete reference can be found in [5], a brief summary is included with the aim of readability.

The relocalization system must detect lost of track and stop EKF integration, to avoid map corruption due to incorrect data associations, and then enable a recovery procedure. The tracking is deemed lost if all attempted matches in a frame

have been unsuccessful, the camera pose uncertainty has grown too large, or if all the predicted mapped features are out of the predicted camera field of view. The relocalization finds matches between the current image and the already estimated map in a data-driven manner without assuming priors about the camera localization with respect to the map.

The image to map matching is cast as a classification problem. A two stage on-line training is applied for every map feature. Firstly at feature initialization, 400 warped versions of the texture patch around the feature are GPU synthesised from the first image and used to train the classifier. The second stage harvests texture patches during EKF SLAM operation that are used for online training. The classifier is also exploited for selecting most distinctive features at initialization, those scoring lowest in the classifier with respect to features already in the map. Doing so, the map features are trackable, locally salient and also distinctive for recognition

When the system has lost track, all FAST features detected in the current image are fed in the classifier. As map features can be similar to each other, multiple feature correspondence hypotheses are considered. Then RANSAC is applied to relocalize the camera with respect to the map. Camera location is hypothesized from three feature correspondences using three-point-pose algorithm proposed in [21]. Each hypothesis is scored according to how many other map features have been matched in the image. Once a good pose hypothesis is found, it is optimized before the SLAM system is reinitialized. If the pose estimate is indeed close enough to the true estimate then one or two fixed map EKF iterations are sufficient to refine the camera pose.

It should be noticed that map integrity is fundamental not only for tracking, but also for relocalization. Hence, both a good spurious detector and a relocalization system are essential for robust performance.

V. EXPERIMENT AND RESULTS

To prove the proposed system performance, a 874 frame laparoscopic sequence at 360x288@25 Hz has been selected (see the accompanying video). The sequence corresponds to an abdominal exploration where a real human ventral hernia (hole) can be seen. It contains some typical challenging issues of laparoscopic sequences: sudden motions, surgical tools clutter, temporary tissue deformation and endoscope extraction and reinsertion into the abdominal cavity.

Endoscope intrinsic parameters were calibrated using a standard planar pattern calibration method, based on [22], and followed by bundle adjustment. A two parameter radial distortion model has been applied.

The experiment was run on an Intel Core i7 processor at 2.67 GHz. The number of features the algorithm tried to measure in each frame was fixed to 45. In order to improve detection and matching of features, all thresholds of feature detection and cross-correlation were relaxed with respect to scenes of human-made environments.

Figure 1 shows for each frame: the map size, the number of measured features, and the number of outliers. A feature is considered as outlier if it is matched by image correlation inside the active search region but deemed as inconsistent by the 1 point RANSAC. Some frames present equal or even higher number of outliers than inliers. This demonstrates the effectiveness of 1-Point RANSAC when facing high outlier rates. Some of this high spurious regions correspond with temporary tissue deformation due to tools interacting with the tissue, or surgeons pushing the cavity from outside (see fig. 2). When this temporary deformation happens, some matches are not found simply because the corresponding points are imaged out of the search window due to the severe deformation, other matches, with smaller deformation are imaged inside the elliptical search window, but eventually are marked as spurious by the RANSAC algorithm.

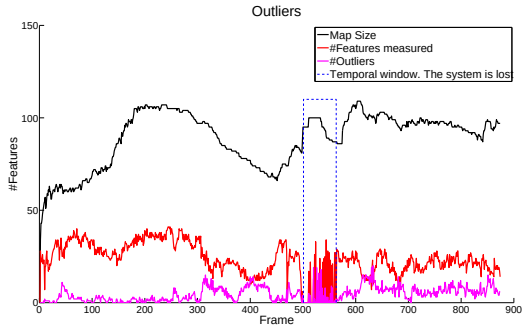

Fig. 1. Map size. Black, total number of map features. Red measured features. Magenta spurious matches i.e. matches found inside the active region search marked as spurious by 1 point RANSAC. Blue dashed rectangle corresponds to frames where track is lost.

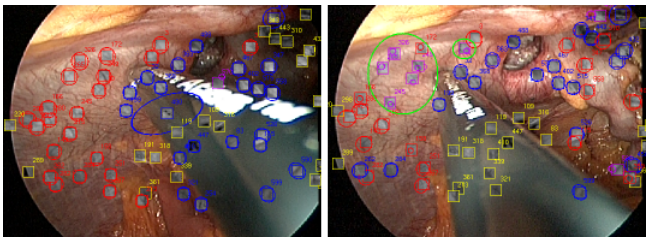

Fig. 2. (left), frame #375 cavity undeformed, most of the map points are successfully reobserved. (right) frame #388 a significant number of map points around the tool-tissue contact point are marked as outliers and hence not measured, avoiding map corruption; green circle encloses points marked as outliers; blue points close to the tool suffer such a big deformation that are imaged out of the search window and hence not matched

Figure 3 presents the total cycle time budged identifying EKF prediction, 1-Point RANSAC, EKF update, EKF update for rescued matches, and initialization and map management. The EKF prediction represents an almost imperceptible share. Also it is worth noting the approximately constant time consumed by 1-point RANSAC, about 10% of the total budget. The low CPU time consumed by the update for rescued matches signals that those rescued matches are just a few however, they are very informative because normally correspond with recently initialized points close to the camera, producing valuable translation information. Map management uses a significant fraction of the computing

budget and needs a more careful optimization.

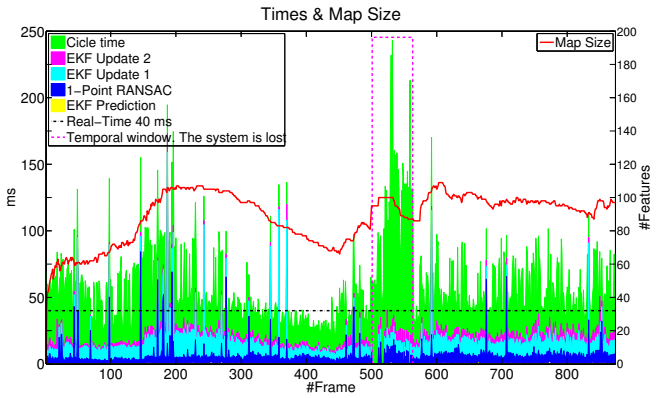

Fig. 3. Computation time budget and map size in double y-axis figure; times are shown in milliseconds (ms) on the left axis. The map number of features is shown on the right axis. Magenta dashed rectangle signals frames where the tracking was lost.

Figure 4 shows the total computation time per frame as a histogram. It can be observed the most of frames take more than 40 ms but less than 100 ms what is coherent with the experiments reported in [4] where 50 features were measured in each frame. This data shows that our system is very close to real time performance.

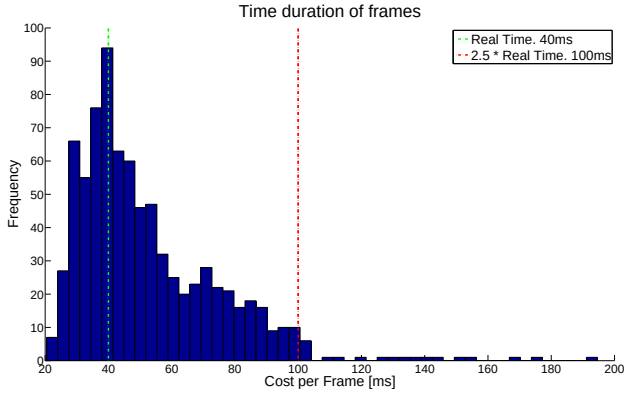

Fig. 4. Histogram showing the computational cost.

During endoscopic procedures, it is frequent extracting and reinserting the endoscope into the body, what represents an extreme situation for relocalization. In the figure 5, four selected frames illustrate the tracking lost and recovery. Before total recovery, an unstable relocation stage is observed.

An important quality of the produced map is the point persistence. The histogram in figure 6 shows how long map features live. It is clear that an important fraction of the 396 initialized features die early because cannot be successfully reobserved. However for a map of about 100 features, there are 54 features ($\approx 50\%$) that have survived more that 600 frames. These persistent features selected in a survival of the fittest way are well spread over the observed cavity, locally salient and suitable for camera relocalization (figure 7).

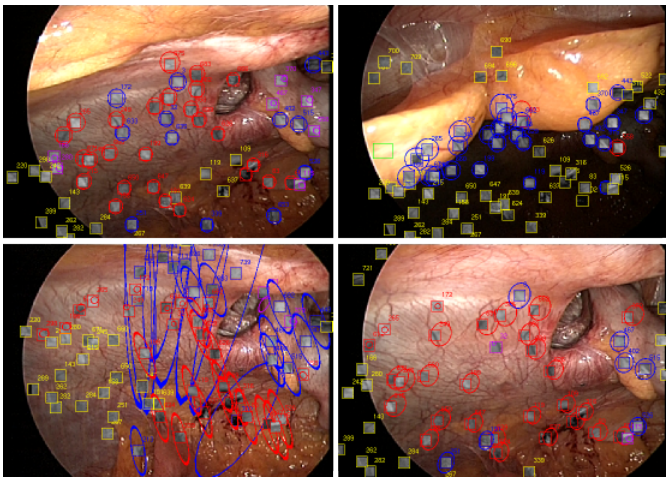

Fig. 5. Relocalization. Upper left, system just before tracking lost. Upper right, endoscope partially out of the cavity. Lower left, unstable relocalization. Lower right, system after total tracking recovery

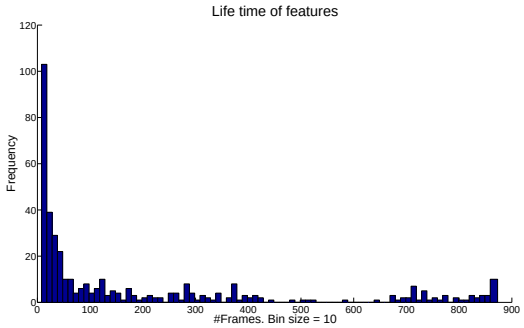

Fig. 6. Histogram displaying feature persistence. 396 are initialized in the experiment, 54 of them survive for more than 600 frames. A new feature is tested for 10 frames, if trackable is kept otherwise is removed, for this reason persistence lower that 10 correspond to non trackable features

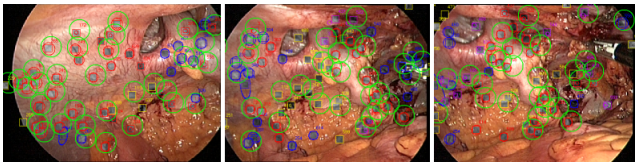

Fig. 7. Features surviving for more than 600 frames are surrounded by a green circle

VI. CONCLUSIONS

This paper presents the combination EKF monocular SLAM + 1-Point RANSAC + Randomized list Relocalization as a suitable to build a map from laparoscopic sequences. This combination has demonstrated to be able to cope with typical challenges in this kind of sequences: sudden motions, surgical tool cluttering, temporary tissue deformation and endoscope extraction and reinsertion in the abdominal cavity.

As the system is based on an EKF filter, the computational cost is quadratic in the map size and linear in the number of measured features in the image. In our experiments, a significant number of map points need to be measured to achieve robust relocalization. In the reported experiments,

the map size was above 100 features and the number of measured features is fixed to 45 resulting times lower than 3 times real-time (120 ms). In the current code there is still room for further improvement so we believe 25Hz real time performance is achievable.

We would like to stress that the proposed algorithm is able to compute a nice summary of the scene after processing the whole sequence. A survival of the fittest process selects what scene features are included in the map. Only locally salient, trackable, and distinctive for relocation points are included in the final map. This rigid map is excellently exploited by relocalization procedure to recover from tracking losses and to relocate at reinsertions. The computed map might well be the starting point for learning priors for processing sequences corresponding to similar procedures performed to different patients.

All results have been validated over a real sequence, so it can be concluded that monocular SLAM in the abdominal cavity is a valid mapping method that does not need any additional sensor but just a standard monocular endoscope and commodity computers.

VII. FUTURE WORK

Experimental validation has been provided for the method, however it would be interesting to compare solution with respect to a ground truth which might be obtained, for example, by an external tracker on the endoscope. It would also be worth a comparison with respect SLAM based on keyframe+bundle adjustment methods such as those proposed in [23].

It is our next goal to research how these maps can be used in surgery, mainly to produce 3D measurements of the cavity and for being the basis of augmented reality insertions with the corresponding clinical validation, and to identify what medical procedures can benefit from the proposed method.

A clear venue for research is the inclusion of elastic models that code the tissue deformations so that the rigidity is replaced by a more realistic model in the in-body environment.

VIII. ACKNOWLEDGMENTS

The authors would like to thank Dr. M. A. Bielsa from Hospital Clínico Universitario "Lozano Blesa", (Zaragoza, Spain), for the ventral hernia image sequence; Imperial College London (Dr. A.J. Davision) for collaboration in visual SLAM software; and University of Oxford (Dr.I. Reid and B. Williams) for visual SLAM and relocalization software.

REFERENCES

[1] O. Grasa, J. Civera, A. Güemes, V. Muñoz, and M. J.M.M., "EKF monocular slam 3d modeling, measuring and augmented reality from endoscope image sequences," in *5th Workshop on Augmented Environments for Medical Imaging including Augmented Reality in Computer-Aided Surgery. (MICCAI 2009)*, 2009.

[2] A. J. Davison, "Real-time simultaneous localisation and mapping with a single camera," in *Proceedings of the International Conference on Computer Vision (ICCV)*, 2003.

[3] A. J. Davison, N. D. Molton, I. D. Reid, and O. Stasse, "MonoSLAM: Real-time single camera SLAM," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 29, no. 6, pp. 1052–1067, 2007.

[4] J. Civera, O. G. Grasa, A. J. Davison, and J. M. M. Montiel, "1-Point RANSAC for EKF Filtering: Application to Real-Time Structure from Motion and Visual Odometry," *Journal of Field Robotics*, vol. 27, no. 5, pp. 609–631, October 2010.

[5] B. Williams, G. Klein, and I. Reid, "Real-time SLAM relocalisation," in *Proceedings of the International Conference on Computer Vision (ICCV)*, 2007.

[6] J. M. M. Montiel, J. Civera, and A. J. Davison, "Unified inverse depth parametrization for monocular SLAM," in *Proceedings of Robotics: Science and Systems (RSS)*, 2006.

[7] J. Civera, A. J. Davison, and J. M. M. Montiel, "Inverse depth parametrization for monocular SLAM," *IEEE Transactions on Robotics (T-RO)*, vol. 24, no. 5, pp. 932–945, 2008.

[8] L. A. Clemente, A. J. Davison, I. D. Reid, J. Neira, and J. D. Tardós, "Mapping large loops with a single hand-held camera," in *Proceedings of Robotics: Science and Systems (RSS)*, 2007.

[9] J. Neira and J. Tardós, "Data association in stochastic mapping using the joint compatibility test," *IEEE Transactions on Robotics and Automation*, vol. Vol. 17, no. No. 6, pp. pp. 890 – 897, 2001.

[10] G. Klein and D. W. Murray, "Improving the agility of keyframe-based SLAM," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2008.

[11] S. Lovegrove and A. Davison, "Real-time spherical mosaicing using whole image alignment," in *Computer Vision ECCV 2010*, ser. Lecture Notes in Computer Science, 2010, vol. 6313, pp. 73–86.

[12] D. Burschka, M. Li, R. Taylor, H. G. D., and M. Ishii, "Scale-invariant registration of monocular endoscopic images to ct-scans for sinus surgery," *Medical Image Analysis*, vol. 9, no. 5, pp. 413–426, 2005.

[13] D. Stoyanov, A. Darzi, and G.-Z. Yang, "A practical approach towards accurate dense 3d depth recovery for robotic laparoscopic surgery," *Computer Aided Surgery*, pp. 199–208, 2005.

[14] F. Mourgues, F. Devernay, and É. Coste-Manière, "3D reconstruction of the operating field for image overlay in 3D endoscopic surgery," in *IEEE/ACM Symp. Augmented. Reality*, 2001, pp. 191–192.

[15] P. Mountney, D. Stoyanov, A. Davison, and G.-Z. Yang, "Simultaneous stereoscope localization and soft-tissue mapping for minimal invasive surgery," in *Medical Image Computing and Computer-Assisted Intervention*, 2006, pp. 347–354.

[16] P. Mountney and G.-Z. Yang, "Motion compensated SLAM for image guided surgery," in *International Conference on Medical Image Computing and Computer Assisted Intervention, (MICCAI2010)*, 2010.

[17] C. Wu, Y. Sun, and C. Chang, "Three-dimensional modeling from endoscopic video using geometric constraints via feature poisitioning," *IEEE Trans. on Biomedical engineering*, vol. 54, no. 7, 2007.

[18] D. Koppel, C.-I. Chen, Y.-F. Wang, L. H., J. Gu, A. Poirson, and R. Wolters, "Toward automated model building from video in computer assisted diagnoses in colonoscopy," in *Proceedings of the SPIE Medical Imaging Conference*, 2007.

[19] E. Rosten and T. Drummond, "Fusing points and lines for high performance tracking." in *IEEE International Conference on Computer Vision*, vol. 2, October 2005, pp. 1508–1511.

[20] V. Lepetit and P. Fua, "Keypoint recognition using randomized trees," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 28, no. 9, pp. 1465–1479, 2006.

[21] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[22] Z. Zhang, "A flexible new technique for camera calibration." *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000.

[23] G. Klein and D. W. Murray, "Parallel tracking and mapping for small AR workspaces," in *Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR)*, 2007.