Motion Compensated SLAM for Image Guided Surgery

Peter Mountney and Guang-Zhong Yang
Department of Computing and Institute of Biomedical Engineering
Imperial College, London SW7 2BZ, UK
{peter.mountney05 , g.z.yang}@imperial.ac.uk

Abstract. The effectiveness and clinical benefits of image guided surgery are well established for procedures where there is manageable tissue motion. In minimally invasive cardiac, gastrointestinal, or abdominal surgery, large scale tissue deformation prohibits accurate registration and fusion of pre- and intra-operative data. Vision based techniques such as structure from motion and simultaneous localization and mapping are capable of recovering 3D structure and laparoscope motion. Current research in the area generally assumes the environment is static, which is difficult to satisfy in most surgical procedures. In this paper, a novel framework for simultaneous online estimation of laparoscopic camera motion and tissue deformation in a dynamic environment is proposed. The method only relies on images captured by the laparoscope to sequentially and incrementally generate a dynamic 3D map of tissue motion that can be co-registered with pre-operative data. The theoretical contribution of this paper is validated with both simulated and *ex vivo* data. The practical application of the technique is further demonstrated on *in vivo* procedures.

Keywords: Image Guided Surgery, Minimally Invasive Surgery, Tracking, Simultaneous Localization And Mapping (SLAM), Augmented Reality

1 Introduction

For Minimally Invasive Surgery (MIS), the use of pre- and intra-operative image guidance has well established benefits. However, its application to procedures with large tissue deformation, such as those encountered in cardiovascular, gastrointestinal and abdominal surgery, is still limited. In order to establish a common *in vivo* material frame of reference that follows tissue deformation, *in situ* 3D reconstruction is necessary. In this regard, the use of fiducial markers and optical tracking, as well as intra-operative imaging such as ultrasound, MR and x-ray fluoroscope have been explored extensively. However, the use of vision techniques based on images from laparoscopes/endoscopes during MIS has clear advantages. It does not require the introduction of additional equipment to what is already a very complex surgical setup. Furthermore, it defines a single co-ordinate system for intra-operative 3D reconstruction, imaging device localization and visualization, therefore removing the need for registration between multiple data streams to a global coordinate system.

The vision based techniques used for MIS currently include Simultaneous Localization And Mapping (SLAM) [1, 2] and Structure from Motion [3, 4]. They have been applied to a variety of anatomical settings such as the abdomen [1, 2],

colon [3], bladder [4] and sinus [5], with the assumption that the structure is relatively static. Structure from Motion has been formulated for used in non-rigid environments however, it requires offline batch processing, thus making it difficult for real-time applications. In [6], for example, it is used to estimate a static cardiac surface at a pre selected point in the cardiac cycle. The 3D structure of the deforming cardiac surface is estimated online in [7] by tracking regions of interest on the organ. It is important to note that in approaches such as this, it is assumed that the laparoscopic camera is fixed, which is not realistic for *in vivo* applications.

The purpose of this paper is to present a novel online approach for simultaneous estimation of camera motion and deforming tissue structure. The system presented extends the current SLAM framework, not only to cope with camera motion, but also to learn a high level model for compensating periodic organ motion. The learnt motion model is explicitly incorporated into the statistical SLAM framework, enabling dynamic tissue motion to be estimated even when it is outside the camera's current field-of-view. The basic steps of the proposed algorithm is schematically illustrated in Fig 1, those specific steps for dealing with dynamic tissue motion are highlighted in yellow. We term this Motion Compensated SLAM (MC-SLAM), which to our knowledge, is the first work for simultaneous online estimation of camera motion and dynamic structure. To assess the accuracy of the proposed framework, the proposed method is validated with synthetic and *ex vivo* data and its *in vivo* application is also demonstrated.

2 Methods

2.1 Motion Modeling

To illustrate the practical use of the MC-SLAM framework, we will use MIS liver surgery as an example. It has been shown that the motion of the liver is correlated to the periodic motion of the diaphragm and therefore respiration [8]. In this work we exploit this correlation to create a high level model of respiration that can be used to predict the dynamic 3D position of tissue in the abdomen. The respiration model is estimated by measuring the 3D motion of points on the liver, as shown in Fig 1b), with a stereo laparoscope. The 3D position of points on the liver are estimated by matching regions of interest in the left and right stereo images and triangulating using the camera's intrinsic and extrinsic parameters. The temporal motion of the 3D points is estimated by tracking the regions of interests along time using an approach outlined in [9]. This approach learns what information is unique about a region and how best to distinguish it from its surroundings, making it well suited to MIS data.

In this work, the liver is assumed to move and deform freely in 3D. Fig. 2a) illustrates the 3D coordinates of a region on the liver surface. The data was collected from a static laparoscope during an in vivo porcine experiment. The signal is periodic in all three axes. The point is periodically moving along a path or principal axis in 3D space which corresponds to the superior-inferior axis [10] and can be approximated as locally linear as shown in Fig 1b). Each point on the liver has an individual principal axis of motion. However, its position on that axis is correlated to the respiration cycle.

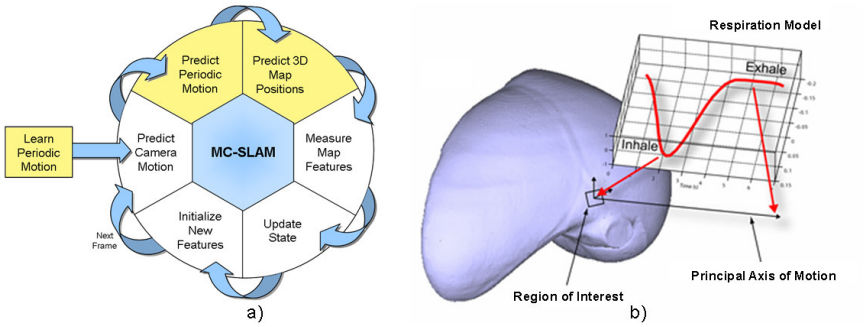

Fig. 1. **(a)** A schematic illustration of the main steps of the proposed MC-SLAM framework. Additional steps for dealing with dynamic tissue motion are highlighted in yellow. **(b)** An example illustration of respiratory modeling from organ motion, which involves: 1) the motion of a region or feature point (of a liver) is tracked temporally in 3D, 2) the principal axis of motion (a vector representing the dominant direction of organ motion) is estimated, 3) the periodic motion along this axis is examined and a respiration model is fitted to the data.

By determining the principal axis of motion and observing the temporal motion characteristics, a model of respiration can be inferred.

In order to relate the 3D coordinate space to the principal axis of motion, Principal Component Analysis (PCA) is used. The result of PCA for the data in Fig. 2a) is shown in Fig. 2b). The first component of PCA is shown in blue, which clearly represents respiration induced tissue motion. The second component contains a small variance caused by hysteresis. A typical respiratory cycle is asymmetrically periodic [10] with a longer dwell time at exhalation as shown in Fig. 2c). This can be represented as

$$z(t) = z_0 - b\cos^{2n}(\frac{\pi t}{\tau} - \phi) \qquad \textbf{(1)}$$

where z_0 is the position of the liver at the exhale, b is the amplitude, τ is the respiration frequency, ϕ is the phase and n describes the shape or gradient of the model. Eq. 1 is used to model the data in the first component of PCA (shown in Fig. 2b). The parameters of Eq. 1 are estimated using Levenberg-Marquardt minimization algorithm where the problem is posed as a least squares curve fitting.

The respiration cycle can be estimated using any point on the liver, assuming it can be tracked throughout the respiratory cycle. The transformation from the global coordinate system to the respiration coordinate system is unique to each point. This means that points on the surface of the liver can move and deform in independent directions but share the same respiration model. Given a model of respiration, it is therefore possible to estimate the dynamic tissue motion using the inverse PCA transformation matrix and a given point in the respiration cycle.

In MIS, respiration is normally regulated by a ventilator. The respiration cycle can therefore be considered periodic with small fluctuations in the frequency and amplitude caused by the ventilator. In the following section, we show how the periodic respiration and associated ventilator noise can be modeled in an Extended Kalman Filter (EKF) to prevent error propagation and synchronization issues.

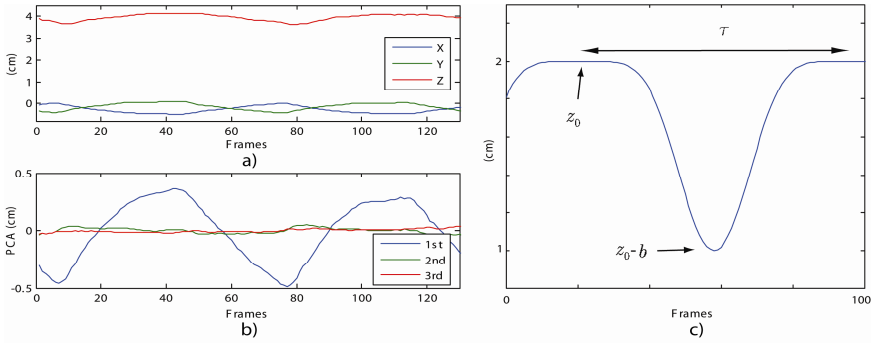

Fig. 2. **(a)** The 3D global coordinates of a region on the surface of the liver illustrating periodic organ motion. **(b)** Result of PCA as applied to (a), illustrating the respiration cycle extracted from organ motion in (a). The first 3 PCA components are shown and the first component corresponds to the principal axis of motion. **(c)** is a graphical representation of the asymmetric respiration model described by Eq. 1.

2.2 Motion Compensated SLAM (MC-SLAM)

In SLAM [1, 2], it is generally assumed that the map is rigid. In MC-SLAM, a periodic motion model is introduced to compensate for the dynamic tissue deformation, thus enabling dynamic mapping and accurate localization of the camera. Conceptually, this introduces three novel steps into the SLAM framework shown in Fig. 1a). The first is to learn an initial estimate of the periodic respiration model using the method described in the previous section. The second and third steps are the prediction of the respiration motion model and prediction of the tissue motion within the map. In conjunction with these steps, we have introduced a new state vector, prediction model and measurement model.

MC-SLAM's **probabilistic framework** is an Extended Kalman Filter (EKF). The state vector $\hat{x}$ is composed of three elements representing the camera $\hat{x}_v$, the periodic respiration model $\hat{m}$ and the map $(\hat{y}_1 \cdots \hat{y}_i)$. The covariance matrix P is square where $P_{\hat{x}_v \hat{m}}$ is the covariance between state elements $\hat{x}_v$ and $\hat{m}$.

$$\hat{x} = \begin{pmatrix} \hat{x}_v \\ \hat{m} \\ \hat{y}_1 \\ \hat{y}_2 \\ \vdots \end{pmatrix} \quad P = \begin{pmatrix} P_{\hat{x}_v\hat{x}_v} & P_{\hat{x}_v\hat{m}} & P_{\hat{x}_v\hat{y}_1} & P_{\hat{x}_v\hat{y}_2} & \cdots \\ P_{\hat{m}\hat{x}_v} & P_{\hat{m}\hat{m}} & P_{\hat{m}\hat{y}_1} & P_{\hat{m}\hat{y}_2} & \cdots \\ P_{\hat{y}_1\hat{x}_v} & P_{\hat{y}_1\hat{m}} & P_{\hat{y}_1\hat{y}_1} & P_{\hat{y}_1\hat{y}_2} & \cdots \\ P_{\hat{y}_2\hat{x}_v} & P_{\hat{y}_2\hat{m}} & P_{\hat{y}_2\hat{y}_1} & P_{\hat{y}_2\hat{y}_2} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \tag{2}$$

The camera's state vector $\hat{x}_v$ contains the position r^W, orientation R^{RW}, translational velocity v^W and angular velocity w^R of the camera. The periodic respiration model $\hat{m} = (\alpha, \tau, b, z_0)^T$ is represented by the parameters derived from Eq. 1, such that $z(t) = z_0 - b\cos^{2n}(\alpha)$ where $\alpha = \pi t / \tau$, t is the time step, z_0 is the exhale position of the liver, b is the amplitude, τ is the frequency and $n = 3$ in accordance with [10]. Phase ϕ is not included as the system is initialized at $\phi = 0$.

The i^{th} map feature $\hat{y}_i = (\overline{y}, eig)$ is derived from the PCA transformation. $\overline{y} = (y_x, y_y, y_z)$ is the mean position of the feature in 3D space during a respiration cycle and $eig = (eig_x, eig_y, eig_z)$ is the eigenvector describing the transformation from 3D space to the periodic respiration model.

The **state prediction model** predicts camera, respiration and map motion. The camera motion is predicted using a "constant velocity, constant angular velocity" model. The state prediction model includes the addition step to predict the point in the respiration cycle and subsequently the motion in the map. The prediction model f_v^m and process noise covariance Q_v^m for the periodic respiration $\hat{m}$ are

$$f_v^m = \begin{bmatrix} 1 & t & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad Q_v^m = \begin{bmatrix} \frac{\Phi_\tau t^3}{3} & \frac{\Phi_\tau t^2}{2} & 0 & 0 \\ \frac{\Phi_\tau t^2}{2} & \Phi_\tau t & 0 & 0 \\ 0 & 0 & \Phi_b t & 0 \\ 0 & 0 & 0 & \Phi_{z_0} t \end{bmatrix} \tag{3}$$

where Φ_τ is noise in the frequency, Φ_b is noise in the amplitude and Φ_{z_0} is noise in the exhale position. The predicted position of the i^{th} map feature in the world coordinate system y_i^W is computed using the predicted respiration parameters and $\hat{y}_i$ such that $y_i^W = eig(z_0 - b\cos^{2n}(\alpha)) + \overline{y}$.

The **measurement model** transforms the state space into the measurement space. Features in the map are measured relative to the camera. A feature's position in the camera coordinate system is estimated using $h_L^R = R^{RW}(y_i^W - r^W)$, where R^{RW} and r^W are the predicted camera rotation and position in the world coordinate system. y_i^W is the position of the feature in the world coordinate system as predicted by the respiration model. The measurement model is

$$h_L^R = R^{RW}(eig(z_0 - b\cos^{2n}(\alpha)) + \overline{y} - r^W) \tag{4}$$

and the partial derivatives (used in the EKF) of the measurement model with respect to $\hat{m}$ are

$$\frac{d\hat{m}}{d\alpha} = R^{RW} eig(nb\sin(\alpha)\cos(\alpha)^{n-1}), \qquad \frac{d\hat{m}}{d\tau} = 0 \tag{5}$$

$$\frac{d\hat{m}}{db} = -R^{RW} eig(\cos(\alpha)^n), \qquad \frac{d\hat{m}}{dz_0} = R^{RW} eig \tag{6}$$

The features are tracked in the image using [9]. During **system initialization** the camera is assumed static for one respiration cycle. Once initialized, new features can easily be added to the system when the camera is moving.

Table 1. Parameters for respiration modeling.

	τ **(Frames)**	b **(cm)**	z_0 **(cm)**
Simulated Estimated	32.38	3.09	0. 95
Simulated Ground Truth	31.83	3	1
Ex Vivo Estimated	52.47	0.85	0.33
Ex Vivo Ground Truth	52	0.9	0.3

4 Results

The proposed MC-SLAM framework is validated on a simulated data set with known ground truth. It is also validated on an *ex vivo* tissue experiment with induced deformation and applied to *in vivo* footage. For quantitative validation with the simulated data, a virtual stereo laparoscope was navigated through a 3D virtual environment with periodic motion applied to a 3D mesh using Eq. 1 and the parameter settings shown in Table 1. The mesh was textured with an image of the liver.

Camera localization is evaluated in Figs. 3a-c) where MC-SLAM is compared to the ground truth and results from the static SLAM framework. The mean position error and standard deviation are 0.25cm and 0.19cm for MC-SLAM and 1.31cm and 0.6cm for static SLAM respectively. As the map and camera position are simultaneously estimated, accurate camera estimation is essential. The position error in MC-SLAM is attributed to rapid changes in acceleration of the camera's motion which are not well modeled. In this simulation, the dominant map motion is in the Z axis and it is evident that this map motion is absorbed into the static SLAM's estimation of the camera's position as there is a periodic error in the Z axis (Fig. 3c). Rotation is accurately recovered by both systems. To validate the method for modeling respiration, the estimated parameters are compared to the ground truth in Table 1.

For the *ex vivo* experiment performed, the ground truth position of the laparoscope was obtained using the approach in [11]. An *ex vivo* porcine liver sample was used and tissue motion was induced with a custom mechanical device. The device consisted of a cam, stepper motor and a sliding tray. Asymmetric motion is induced using Eq. 1 where z, n and b are defined by the cam profile and τ is defined by the stepper motor. Quantitative evaluation of laparoscope localization is shown in Fig 3d-f). The recovered motion using MC-SLAM closely follows the ground truth. Static SLAM however, periodically oscillates away from the ground truth. The mean error and standard deviation are 0.11cm and 0.07cm for MC-SLAM and 0.56cm and 0.25cm for static SLAM respectively. In addition, static SLAM is more prone to data association errors as shown in Fig. 3d-f) between frames 800-1000. The estimated and ground truth parameters of the respiration model are compared in Table 1 and Fig. 4a).

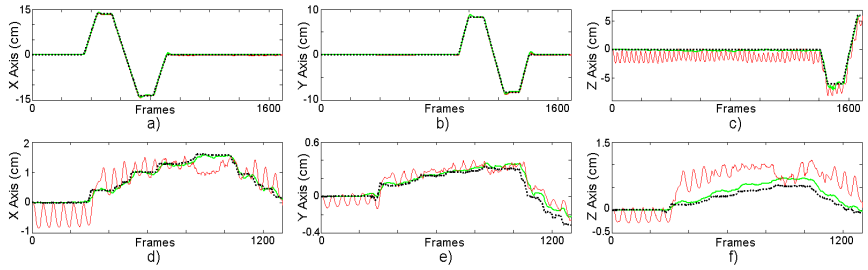

Fig. 3. Quantitative comparison of estimated laparoscope position in the world coordinate frame with MC-SLAM (green), static SLAM (red) and ground truth (black dashed). **(a-c)** simulated data XYZ axes. **(d-f)** *ex vivo* data XYZ axes.

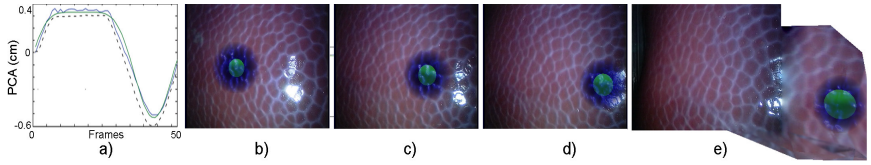

Fig. 4. *Ex vivo* experiment results, **(a)** the estimated respiration model; blue - observed data, green - respiration model, black dashed - ground truth. **(b-e)** Illustration of Image Guided Surgery with pre-operative data visualized [12] intra-operatively. **(a-b)** show a static laparoscope and the tissue at **(b)** exhale and **(c)** inhale position. **(d)** combined laparoscope and tissue motion. **(e)** laparoscope motion results in the target moving outside the current field-of-view. The dynamic target position is estimated relative to the current position of the laparoscope and visualized using view expansion [13].

Figs. 4b-e) show the intra-operative laparoscopic images of the *ex vivo* tissue augmented with a virtual tumor which is manually and rigidly registered to the MC-SLAM map. The tumor is visualized using the Augmented Reality (AR) technique presented in [12]. Figs. 4b-c) are captured from a static laparoscope and illustrate the position of the tumor at full exhale and full inhale. Figs. 4d-e) demonstrate the system working in the presents laparoscopic and tissue motion. In Fig. 4e), the camera navigates away from the tumor and its position is visualized outside the current field-of-view using dynamic view expansion [13]. This illustrates the capability of the system to predict the dynamic 3D position of tissue even when the tissue is not in the camera's current field-of-view.

For the *in vivo* experiment, the ground truth data was not available. The estimated respiration model is illustrated in Fig. 5a) and Figs. 5b-e) illustrate results with the use of AR visualization. A virtual tumor is manually and rigidly registered to the MC-SLAM map. Figs. 5b-c) show intra-operative *in vivo* images captured from a static laparoscope. Figs. 5b) and 5c) show the tissue position at the full exhale and full inhale point in the respiration cycle. This illustrates tissue displacement resulting from respiration which was estimated at 1.08cm. The motion of the augmented tumor demonstrates the dynamic nature of the MC-SLAM map and progression beyond the rigid environment assumption. In Fig. 5d) and Fig. 5e), the laparoscope is navigated by the surgeon to explore the abdomen. Throughout the exploration the augmented data is displayed in a visually consistent manor in the presents of both laparoscope and tissue motion.

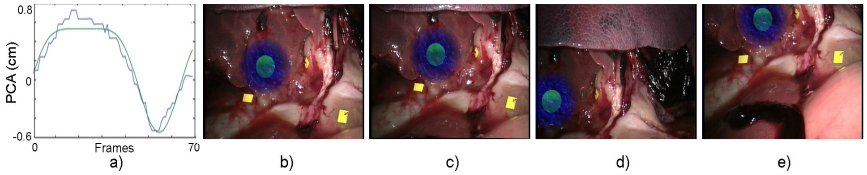

Fig. 5. *In vivo* experiment results; **(a)** the estimated respiration model (green) and observed data (blue); **(b-e)** illustration of Image Guided Surgery on *in vivo* footage with virtual pre-operative data visualized [12] intra-operatively; **(b-c)** images from a static laparoscope with the tissue at **(b)** exhale and **(c)** inhale position; **(d-e)** show combined tissue and laparoscope motion during abdominal exploration.

5 Conclusions

In this paper, we have presented a novel MC-SLAM system for simultaneous laparoscopic localization and dynamic tissue mapping. The system explicitly incorporates a periodic model of respiration into the statistical framework. This enables the system to predict and anticipate changes in tissue structure and estimated organ motion even when it is not in the laparoscope's current field-of-view. The method has been validated on simulated and *ex vivo* data and its clinical relevance has been demonstrated with *in vivo* data. Future work will focus on non-rigid registration of pre-operative data, faster initialization and more sophisticated motion models.

References

1. Mountney P., Stoyanov D., Davison A.J., Yang G.-Z. Simultaneous Stereoscope Localization and Soft-Tissue Mapping for Minimal Invasive Surgery. Medical Image Computing And Computer Assisted Intervention; (2006);347-354.
2. Garcıa O., Civera J., Gueme A., Munoz V., Montiel J.M.M. Real-time 3D Modeling from Endoscope Image Sequences. International Conference on Robotics and Automation Workshop on Advanced Sensing and Sensor Integration in Medical Robotics; (2009).
3. Koppel D., Chen C.-I., Wang Y.-F., Lee H., Gu J., Poirson A., et al. Toward automated model building from video in computer-assisted diagnoses in colonoscopy. In Proc SPIE; (2007).
4. Wu C.-H., Sun Y.-N., Chang C.-C. Three-dimensional modeling from endoscopic video using geometric constraints via feature positioning. IEEE Transactions on Biomedical Engineering. (2007);54(7):1199-1211.
5. Burschka D., Li M., Ishii M., Taylor R., Hager G.D. Scale-invariant Registration of Monocular Endoscopic Images to CT-scans for Sinus Surgery. Medical Image Analysis. (2005);9(5):413 - 426.
6. Hu M., Penney G.P., Rueckert D., Edwards P.J., Bello R., Casula R., et al. Non-rigid Reconstruction of the beating Heart Surface for Minimally Invasive Cardiac Surgery. Medical Image Computing and Computer Assisted Intervention; (2009);34-42.
7. Stoyanov D., Mylonas G.P., Deligianni F., Darzi A., Yang G.-Z. Soft-tissue Motion Tracking and Structure Estimation for Robotic Assisted MIS Procedures. Medical Image Computing and Computer Assisted Intervention; (2005);139-146.
8. Davies S.C., Hill A.L., Holmes R.B., Halliwell M., Jackson P.C. Ultrasound quantitation of respiratory organ motion in the upper abdomen. British Journal of Radiology. (1994);67:1096–1102.
9. Mountney P., Yang G.-Z. Soft Tissue Tracking for Minimally Invasive Surgery: Learning Local Deformation Online. Medical Image Computing and Computer Assisted Intervention; (2008);364-372.
10. Lujan A.E., Larsen E.W., Balter J.M., Haken. R.K.T. A Method for Incorporating Organ Motion due to Breathing into 3D Dose Calculations. Medical Phisics. (1999);26(5):715-720.
11. Noonan D., Mountney P., Elson D., Darzi A., Yang G.-Z. A Stereoscopic Fibroscope for Camera Motion and 3D Depth Recovery During Minimally Invasive Surgery. International Conference on Robotics and Automation; (2009);4463-4468.
12. Lerotic M., Chung A.J., Mylonas G., Yang G.-Z. pq -space Based Non-Photorealistic Rendering for Augmented Reality. International Conference on Medical Image Computing and Computer Assisted Intervention; (2007);102-109.
13. Mountney P., Yang G.-Z. Dynamic View Expansion for Minimally Invasive Surgery using Simultaneous Localization And Mapping. Engineering in Medicine and Biology Conference; (2009);1184-1187.