Real-time augmented reality for ear surgery

Raabid Hussain[1,✉], Alain Lalande[1], Roberto Marroquin[1], Kibrom Berihu Girum[1], Caroline Guigou[2], and Alexis Bozorg Grayeli[1,2]

[1]Le2i, Universite de Bourgogne Franche-Comte, Dijon, France
[2]ENT Department, University Hospital of Dijon, Dijon, France
✉ Raabid.Hussain@u-bourgogne.fr

Abstract. Transtympanic procedures aim at accessing the middle ear structures through a puncture in the tympanic membrane. They require visualization of middle ear structures behind the eardrum. Up to now, this is provided by an oto endoscope. This work focused on implementing a real-time augmented reality based system for robotic-assisted transtympanic surgery. A preoperative computed tomography scan is combined with the surgical video of the tympanic membrane in order to visualize the ossciles and labyrinthine windows which are concealed behind the opaque tympanic membrane. The study was conducted on 5 artificial and 4 cadaveric temporal bones. Initially, a homography framework based on fiducials (6 stainless steel markers on the periphery of the tympanic membrane) was used to register a 3D reconstructed computed tomography image to the video images. Micro/endoscope movements were then tracked using Speeded-Up Robust Features. Simultaneously, a micro-surgical instrument (needle) in the frame was identified and tracked using a Kalman filter. Its 3D pose was also computed using a 3-collinear-point framework. An average initial registration accuracy of 0.21 mm was achieved with a slow propagation error during the 2-minute tracking. Similarly, a mean surgical instrument tip 3D pose estimation error of 0.33 mm was observed. This system is a crucial first step towards keyhole surgical approach to middle and inner ears.

Keywords: Augmented reality, transtympanic procedures, otology, minimally invasive, image-guided surgery

1 Introduction

During otologic procedures, when the surgeon places the endoscope inside the external auditory canal, the middle ear cleft structures, concealed behind the opaque tympanic membrane (TM), are not directly accessible. Consequently, surgeons can access these structures using the TM flap approach [1] which is both painful and exposes the patient to risk of infection and bleeding [2].

Alternatively, transtympanic procedures have been designed which aim at accessing the middle ear cleft structures through a small and deep cavity inside the ear. These techniques have been used in different applications such as ossicular chain repair, drug administration and labyrinthe fistula diagnosis [3, 4].

The procedures offer many advantages: faster procedure, preservation of TM and reduced bleeding. However, limited operative space, field of view and instrument manoeuvring introduce surgical complications.

Our hypothesis claims that augmented reality (AR) would improve the procedure of middle ear surgery by providing instrument pose information and superimposing preoperative computed tomography (CT) image of the middle ear onto the micro/endoscopic video of TM. The key challenge is to enhance ergonomy while operating in a highly undersquared cylinderical workspace achieving sub-millimetric precision. To our knowledge, AR has not been applied to transtympanic and otoendoscopic procedures, thus the global perspective of the work is to affirm our hypothesis.

In computer assisted surgical systems, image registration plays an integral role in the overall performance. Feature extraction methods generally do not perform well due to the presence of highly textured structures and non-linear biasing [5]. Many algorithms have been proposed specifically for endoscope-CT registration. Combinations of different intensity based schemes such as cross-correlation, squared intensity difference, mutual information and pattern intensity have shown promising results [6, 7]. Similarly, feature based schemes involving natural landmarks, contour based feature points, iterative closest point and k-means clustering have also been exploited [8, 9].

Different techniques involving learned instrument models, artificial markers, pre-known kinematic and gradient information using Hough transform have been proposed to identify instruments in video frames [10]. If the target is frequently changing its appearance, gradient based tracking algorithms need continuous template updating to maintain accurate position estimation. Analogously, reliable amount of training data is required for classifier based techniques. Although extensive research has been undertaken for identification of instruments in image plane, limited work has been accomplished to estimate 3D pose. Trained random forest classifier using instrument geometry as a prior and visual servoing techniques employing four marker points have been proposed [11, 12]. Three point perspective framework involving collinear markers has been also suggested [13].

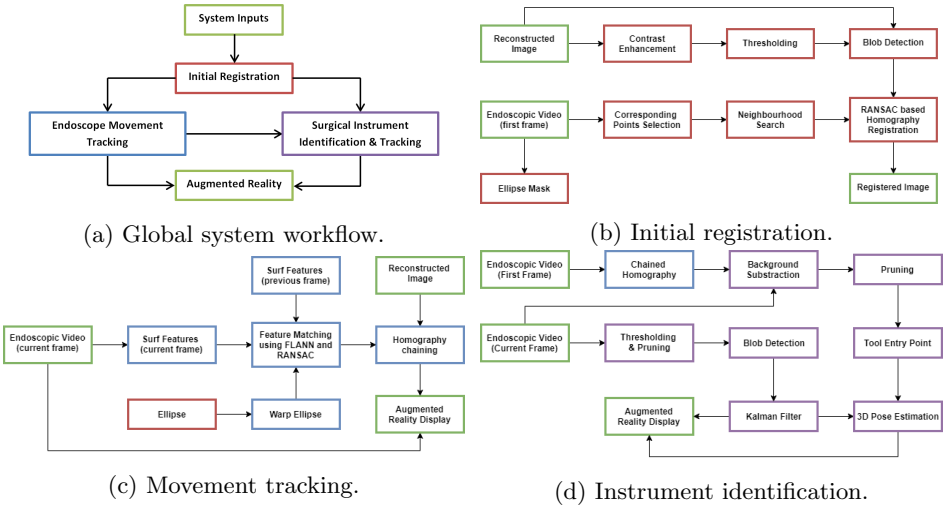
Our proposed approach initially registers the CT image with the microscopic video, based on fiducial markers. This is followed by a feature based motion tracking scheme to maintain synchronisation. The surgical instrument is also tracked and its pose estimated using 3-collinear-point framework.

2 Methodology

The system is composed of three main processes: initial registration, movement tracking and instrument pose estimation. The overall hierarchy is presented in Fig. 1. The proposed system has two main inputs. Firstly, the reconstructed image which is the display of the temporal bone, depicting middle ear cleft structures behind TM, obtained from preoperative CT data through OsiriX 3D endoscopy function (Pixmeo SARL, Switzerland). Secondly, the endoscopic video which is the real-time video acquired from a calibrated endoscope or surgical

microscope during a surgical procedure. The camera projection matrix of the input camera was computed using [14]. The calibration parameters are later used for 3D pose estimation.

There is low similarity between the reconstructed and endoscopic images (Fig. 2), thus the performance of intensity and feature based algorithms is limited in this case. Marroquin et al. [2] established correspondence by manually identifying points in the endoscopic and CT images. However, accurately identifying natural landmarks is a tedious and time-consuming task. Thus six stainless steel fiducial markers ($\approx 1mm$ in length) were attached around TM (prior to CT acquisition).



(a) Global system workflow.

(b) Initial registration.

(c) Movement tracking.

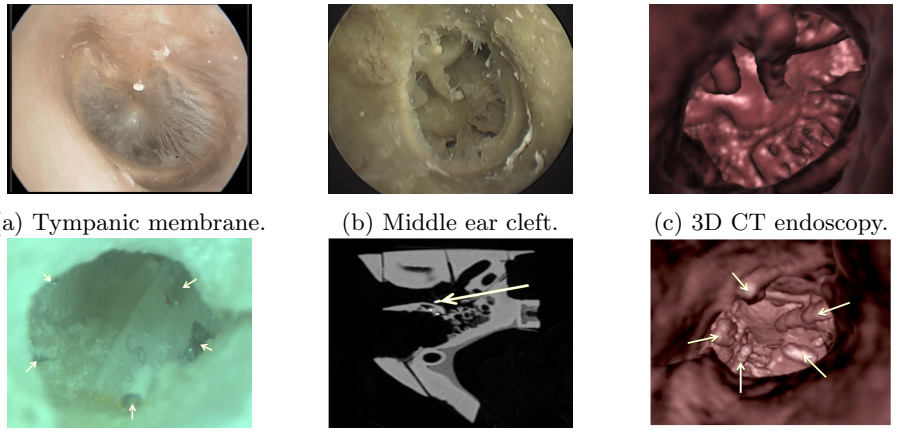
(d) Instrument identification.

Fig. 1: Overall workflow of the proposed system. A colour coding scheme, defined in (a), has been used to differentiate between different processes.

2.1 Initial Semi-automatic Endosocope-CT Registration

Since the intensity of fiducials is significantly higher than that of anatomical regions on CT images, contrast enhancement and thresholding is used to obtain fiducial regions. The centre of each fiducial is then obtained using blob detection. The user selects the corresponding fiducials in the first frame of the endoscopic video. There are very few common natural landmarks around TM, thus the fiducials ease up the process of establishing correspondence. In order to eliminate human error, similar pixels in a small neighbourhood around the selected points are also taken into account. A RANdom SAmple Consensus (RANSAC) based homography [15] registration matrix H_R, which warps the reconstructed image onto the endoscopic video, is computed using these point correspondences.

An ellipse shaped mask is generated using the fiducial points in the endoscopic frame. Since TM does not have a well-defined boundary in the endoscopic



(a) Tympanic membrane. (b) Middle ear cleft. (c) 3D CT endoscopy.

(d) Microscopic image. (e) CT MPR image. (f) Reconstruction image.

Fig. 2: Problem definition. Amalgamation of a reconstructed CT image (c) with the endoscopic video (a) may be used to visualize the middle ear cleft structures (b) without undergoing a TM flap procedure. However, similarity between them is low so fiducial markers are introduced which appear (d) grey in the microscopic image, (e) white in the CT MPR image and (f) as protrusions on the CT reconstructed image.

video, this mask is used as an approximation of TM. The mask is used in the tracking process to filter out unwanted (non-planar) features.

2.2 Endoscope-Target Motion Tracking

Speeded Up Robust Features (SURF) [16] was employed in our system for tracking the movement between consecutive video frames [2]. For an accurate homography, all the feature points should lie on coplanar surfaces [15]. However, the extracted features are spread across the entire image plane comprising of the TM and auditory canal. The ellipse generated in previous step is used to filter out features that do not lie on TM (assumed planar). A robust feature matching scheme based on RANSAC and nearest neighbour (FLANN [17]) frameworks is used to determine the homography transformation H_T between consecutive frames. A chained homography framework is then used to warp the registered reconstructed image onto the endoscopic frame:

$$H^{i+1} = H_T * H^i \ , \tag{1}$$

where H^0 is set as identity. H^i can then be multiplied with H_R to transform the original reconstructed image to the current time step. A linear blend operator is used for warping reconstructed image onto the current endoscopic frame.

2.3 3D Pose Estimation of Surgical Instrument

In surgical microscopes, small depth of focus leads to a degradation of gradient and colour information. Thus popular approaches for instrument identification do not perform well. Consequently, three collinear ink markers were attached to the instrument (Fig. 3). The three marker regions are extracted using thresholding. Since, discrepancies are present, a pruning step followed by blob detection is carried out to extract centres of the largest three regions. A linear Kalman filter is used to refine the marker centre points to eliminate any residual degradation.

Since, the instrument may enter from any direction and protrude indefinitely, geometric priors are not valid. The proposed approach assumes that no instrument is present in the first endoscopic frame. The first frame undergoes a transformation based on H^i. Background subtraction followed by pruning (owing to discrepancies in H^i) is used to extract the tool entry point in the frame boundary. The tool entry point is then used to associate the marker centres to marker labels B, C and D. The instrument tip location can then be obtained using a set of perception based equations that lead to:

$$a = \frac{1}{3}(b + c + d + \frac{AB}{CD}(c - d) + \frac{AC}{BD}(b - d) + \frac{AD}{BC}(b - c)) \ , \tag{2}$$

where a is projection of the surgical instrument tip A on the 2D image frame, b, c and d are projections of the markers and alphabet pairs represent the physical distance between the markers.

A three point perspective framework is then used to estimate the 3D pose of the instrument. Given focal length of the camera, known physical distance between 3 markers and their projected 2D coordinates, the position of the instrument tip can be estimated by fitting the physical geometry of the tool onto the projected lines Ob, Oc and Od (Fig. 3) [13].

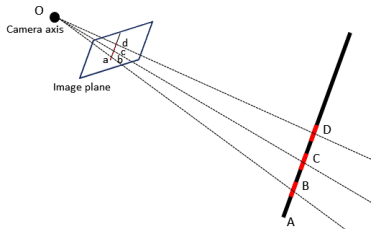
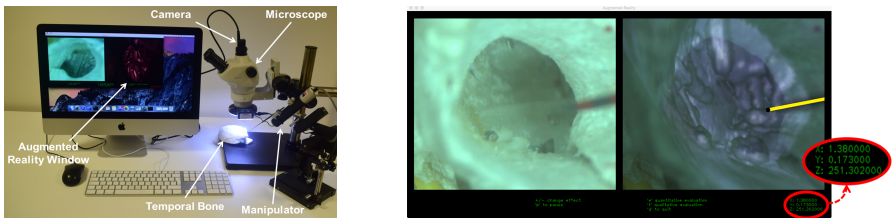


Fig. 3: Three collinear point framework for 3D pose estimation.

3 Experimental Setup and Results

The proposed system was initially evaluated on five temporal bone phantoms (corresponding patient ages: 1-55 years). All specimens underwent a preoperative

CT scan (Light speed, 64 detector rows, 0.6 x 0.6 x 0.3 mm^3 voxel size, General Electric Medical Systems, France). Six 1 mm fiducial markers, were attached around TM in a non-linear configuration with their combined centre coinciding with the target [18]. Real-time video was acquired using a microscope lens. Small movements were applied to the microscope in order to test the robustness of the system. The experimental setup and the augmented reality output are shown in Fig. 4. Processing speed of 12 frames per second (fps) was realized.



(a) Experimental setup. (b) Augmented Reality Window.

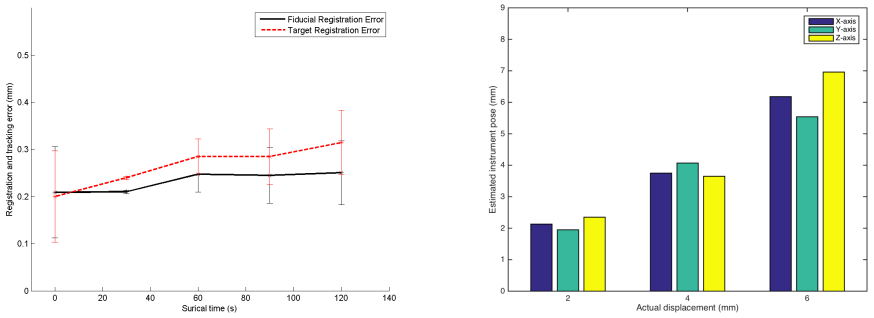
Fig. 4: (a) Augmented reality system. (b) Real-time video from the microscope (left) and augmented reality window (right).

The fiducial marker points in the reconstructed image were automatically detected and displayed on the screen and the user selected their corresponding fiducial points on the microscopic frame. Mean fiducial registration error (physical distance between estimated positions in microscopic and transformed reconstructed images) of 0.21 mm was observed.

During surgery, microscope will remain quasi-static. However to validate robustness of the system, combinations of translation, rotation and scaling with a speed of 0-10 mm/s were applied to the microscope. The system, evaluated at 30 second intervals, maintained synchronisation with a slow propagation error of 0.02 mm/min (Fig. 5a). Fiducial markers were used as reference points for evaluation. Template matching was used to automatically detect the fiducial points in the current frame. These were compared with the fiducial points in transformed reconstructed image to compute the tracking error.

The system was also evaluated on pre-chosen surgical target structures (incus and round window niche). TM of four temporal bone cadavers was removed and the above experiments were repeated. Similarly, a mean target registration error (computation similar to fiducial registration error) of 0.20 mm was observed with a slow propagation error of 0.05 mm/min (Fig. 5a).

For instrument pose estimation, pre-known displacements were applied in each axis and a total of 50 samples per displacement were recorded. Mean pose estimation errors of 0.20, 0.18 and 0.60 mm were observed in X, Y and Z axes respectively (Fig. 5b). The pose estimation in X and Y axes was better than in Z axis because any small deviation in instrument identification constitutes a relatively large deviation in the Z pose estimation.



(a) Registration and tracking accuracy. (b) 3D pose estimation accuracy.

Fig. 5: Experimental results. (a) Registration and tracking accuracy of the AR system evaluated at fiducial and surgical targets. (b) Displacement accuracy assessment of 3D pose estimation process (displayed statistics are for 50 samples).

4 Conclusion

An AR based robotic assistance system for transtympanic procedures was presented. A preoperative CT scan image of the middle ear cleft was combined with the real-time microscope video of TM using 6 fiducial markers as point correspondences in a semi-automatic RANSAC based homography framework. The system is independent of marker placement technique and is capable of functioning with endoscopes and mono/stereo microscopes. Initial registration is the most crucial stage as any error introduced during this stage will propagate throughout the procedure. Mean registration error of 0.21 mm was observed. To keep synchronisation, the relative microscope-target movements were then tracked using a SURF based robust feature matching framework. A microscopic propagation error was observed. Simultaneously, 3D pose of a needle instrument, upto 0.33 mm mean precision, was provided for assistance to the surgeon using a monovision based perspective framework. Additional geometric priors can be incorporated to compute pose of angled instruments. Initial experiments have shown promising results, achieving sub-millimetric precision, and opening new perspectives to the application of minmally invasive procedures in otology.

References

1. Gurr, A., Sudhoff, H., Hildmann, H.: Approaches to the middle ear. In: Hildmann H., Sudhoff H. (eds.) Middle Ear Surgery. pp. 19-23. Springer (2006). doi: 10.1007/978-3-540-47671-9
2. Marroquin, R., Lalande, A., Hussain, R., Guigou, C., Grayeli, A.B.: Augmented reality of the middle ear combining otoendoscopy and temporal bone computed tomography. Accepted in Otol. Neurotol. (2018).

3. Dean, M., Chao, W.C., Poe, D.: Eustachian Tube Dilation via a Transtympanic Approach in 6 Cadaver Heads: A Feasibility Study. Otolaryngol. Head Neck Surg. 155(4), 654-656 (2016). doi: 10.1177/0194599816655096
4. Mood, Z.A., Daniel, S.J.: Use of a microendoscope for transtympanic drug delivery to the round window membrane in chinchillas. Otol. Neurotol. 33(8), 1292-1296 (2012). doi: 10.1097/MAO.0b013e318263d33e
5. Viergever, M.A., Maintz, J.A., Klein, S., Murphy, K., Staring, M., Pluim, J.P.: A survey of medical image registration under review. Med. Image Anal. 33, 140-144 (2016). doi: 10.1016/j.media.2016.06.030
6. Hummel, J., Figl, M., Bax, M., Bergmann, H., Birkfellner, W.: 2D/3D registration of endoscopic ultrasound to CT volume data. Phys. Med. Biol. 53(16), 4303 (2008). doi: 10.1088/0031-9155/53/16/006
7. Yim, Y., Wakid, M., Kirmizibayrak, C.: Registration of 3D CT data to 2D endoscopic image using a gradient mutual information based viewpoint matching for image-guided medialization laryngoplasty. J. Comput. Sci. Eng. 4(4), 368-387 (2010). doi: 10.5626/JCSE.2010.4.4.368
8. Jun, G.X., Li-li, H., Yi, N.: Feature points based image registration between endoscope image and the CT image. In: IEEE International Conference on Electric Information and Control Engineering, pp. 2190-2193. IEEE Press (2011). doi: 10.1109/ICEICE.2011.5778261
9. Wengert, C., Cattin, P., Du, J.M., Baur, C., Szekely, G.: Markerless endoscopic registration and referencing. In: International Conference on Medical Image Computing and Computer Assisted Intervention, pp. 816-823. Springer (2006). doi: 10.1007/11866565_100
10. Haase, S., Wasza, J., Kilgus, T., Hornegger, J.: Laparoscopic instrument localization using a 3D time of flight/RGB endoscope. In: IEEE Workshop on Applications of Computer Vision, pp. 449-454. IEEE Press (2013). doi: 10.1109/WACV.2013.6475053
11. Allan, M., Ourselin, S., Thompson, S., Hawkes, D.J., Kelly, J., Stoyanov, D.: Toward detection and localization of instruments in minimally invasive surgery. IEEE Trans. Biomed. Eng. 60(4), 1050-1058 (2013). doi: 10.1109/TBME.2012.2229278
12. Nageotte, F., Zanne, P., Doignon, C., Mathelin, M.D.: Visual servoing based endoscopic path following for robot-assisted laparoscopic surgery. In: IEEE/RSJ International conference on intelligent robots and systems, pp. 2364-2369, IEEE Press (2006). doi: 10.1109/IROS.2006.282647
13. Liu, S.G., Peng, K., Huang, F.S., Zhang, G.X., Li, P.: A portable 3d vision coordinate measurement system using a light pen. Key Eng. Mater. 295, 331-336 (2005). doi: 10.4028/www.scientific.net/KEM.295-296.331
14. Zhang, Z.: A flexible new technique for camera calibration. IEEE Trans. Pattern Anal. Mach. Intell. 22(11), 1330-1334 (2000). doi: 10.1109/34.888718
15. Hartley, R. and Zisserman, A. Multiple view geometry in computer vision. Cambridge University Press (2003)
16. Bay, A., Tuytelaars, T., Gool, L.V.: Surf: Speeded up robust features. In: 9th European Conference on Computer Vision, pp. 404-417. Springer (2006). doi: 10.1007/11744023_32
17. Muja, M., Lowe, D.G.: Fast approximate nearest neighbors with automatic algorithm configuration. In: 4th International Conference on Computer Vision Theory and Applications, pp. 331-340, Springer (2009). doi: 10.5220/0001787803310340
18. West, J.B., Fitzpatrick, J.M., Toms, S.A., Maurer Jr, C.R, Maciunas, R.J.: Fiducial point placement and the accuracy of point-based, rigid body registration. Neurosurgery. 48(4), 810-817 (2001). doi: 10.1097/0006123-200104000-00023