Video-based augmented reality combining CT-scan and instrument position data to microscope view in middle ear surgery

Raabid Hussain1,2✉, Alain Lalande1,2, Roberto Marroquin3, Caroline Guigou1,3 & Alexis Bozorg Grayeli1,3

The aim of the study was to develop and assess the performance of a video-based augmented reality system, combining preoperative computed tomography (CT) and real-time microscopic video, as the first crucial step to keyhole middle ear procedures through a tympanic membrane puncture. Six different artificial human temporal bones were included in this prospective study. Six stainless steel fiducial markers were glued on the periphery of the eardrum, and a high-resolution CT-scan of the temporal bone was obtained. Virtual endoscopy of the middle ear based on this CT-scan was conducted on Osirix software. Virtual endoscopy image was registered to the microscope-based video of the intact tympanic membrane based on fiducial markers and a homography transformation was applied during microscope movements. These movements were tracked using Speeded-Up Robust Features (SURF) method. Simultaneously, a micro-surgical instrument was identified and tracked using a Kalman filter. The 3D position of the instrument was extracted by solving a three-point perspective framework. For evaluation, the instrument was introduced through the tympanic membrane and ink droplets were injected on three middle ear structures. An average initial registration accuracy of 0.21 ± 0.10 mm (n = 3) was achieved with a slow propagation error during tracking (0.04 ± 0.07 mm). The estimated surgical instrument tip position error was 0.33 ± 0.22 mm. The target structures’ localization accuracy was 0.52 ± 0.15 mm. The submillimetric accuracy of our system without tracker is compatible with ear surgery.

Middle ear surgery involves manipulation of small, delicate and complex structures inside a very confined workspace. Critical nerves and blood vessels are in close proximity of these structures making submillimetric accuracy, a prerequisite for surgical procedures to be carried out safely1. Conventional image guidance systems used in ear surgery have a limited use due to their insufficient precision and poor ergonomics1.

In conventional surgery, middle ear contents are approached through external auditory canal after tympanomeatal flap elevation. This approach entails bleeding into the middle ear, risk of tympanic membrane perforation or laterisation, injury to ossicles or corda tympani and postoperative care for several days2. Alternatively, transtympanic procedures have been designed to access middle ear cleft structures through a small puncture in the tympanic membrane which would spontaneously heal as during a grommet insertion. This route has been employed for different indications such as ossicular chain repair, drug administration and labyrinthine fistula diagnosis3–5. The procedure offers several potential advantages over traditional surgery: faster route, tympanic membrane preservation, reduced bleeding, simpler and less painful post-operative care. However, manipulation of fragile ossicles through this keyhole approach will probably require a robot-based technique6,7. Moreover, visualization of middle ear content and surgical instruments behind the closed tympanic membrane is essential. This goal may be achieved through combination of middle ear CT-scan and real-time microscopic video in an augmented reality (AR) framework.

Different works have been proposed on AR based surgical systems mainly targeting orthopedics, hepatobiliary and neurologic surgeries8. However, only few studies have targeted cranial base and otolaryngology domains

1ImViA Laboratory, University of Burgundy Franche Comte, Dijon, France. 2Medical Imaging Department, Dijon University Hospital, Dijon, France. 3Otolaryngology Department, Dijon University Hospital, Dijon, France. ✉e-mail: raabid.hussain@u-bourgogne.fr
As a proof of concept, it has been shown that AR technology can be used to perform transtympanic procedures with submillimetric precision. An AR system was designed to register the real-time middle ear OCT images onto the microscope view. The system includes a dedicated otologic robotic system such as Robotol (Collin Medical SAS, France) and an external tracking system like optical or electromagnetic trackers to track the motion between the patient and the microscope. This setup allows for real-time registration of the OCT images with the microscope view, providing the surgeon with a virtual endoscopy function. The system was tested on both human cadaveric temporal bones and artificial temporal bone phantoms.

### Table 1. Experimental conditions.

| Experiment Number | 1 | 2 | 3 | 4 | 5 | 6 |
|-------------------|---|---|---|---|---|---|
| Phantom model     | TF-ba | TF-ka | TF-bm | TF-ec | TF-dc | TF-bm |
| Corresponding patient age | 55 | 1 | 55 | 6 | 2 | 55 |
| Number of markers used | 5 | 6 | 5 | 5 | 6 | 5 |
| X, Y translations | Yes | Yes | Yes | Yes | Yes | Yes |
| Z translations | Yes | Yes | Yes | No | Yes | Yes |
| Rotation | Yes | Yes | Yes | Yes | No | No |
| Movement speed | Low | High | High | High | Low | High |

### Material and methods

#### Experimental setup.
Six artificial human temporal bone specimens (Phacon Inc., Leipzig, Germany), with variations in corresponding age, size, and anatomy, were included in this prospective study. Ethical approval and informed consent were not required for this study. Five or six fiducial markers (0.5 mm diameter and 1 mm long stainless-steel wire) were glued to the periphery of the tympanic membrane, evenly distributed on its perimeter (Table 1). To optimize image to object registration, the markers were placed far apart in a non-linear configuration with their combined centre coinciding with the projection of the target on the plane defined by the markers.

All specimens underwent pre-operative CT-scan (0.6 × 0.6 × 0.3 mm³ voxel size, General Electric Medical Systems, Buc, France). 3D reconstruction, based on DICOM data, of middle ear cleft was carried out using OsiriX virtual endoscopy function (Pixmeo SARL, Bernex, Switzerland). The 3D reconstruction was obtained by placing the virtual endoscope in the external auditory canal facing the umbo 10 mm outside the tympanic membrane. This image of the middle ear cleft structures was used as reference to warp around the microscope video. In parallel, otoendoscopy was performed for all temporal bone specimens with a surgical microscope (Zoom Pro 10.76, 115 mm working distance, 8–50x zoom, Perfix, Escalquens, France) connected to a high definition camera (xIQ MQ013CG-ON, Ximea Gmbh, Munster, Germany) to visualize the tympanic membrane.

### References

- Wisotzky et al.
- Liu et al.
- Lee et al.
- DaVinci surgical system
- Robotol system
- Phacon Inc.

[www.phacon.de/en/hno/felsenbein](https://www.phacon.de/en/hno/felsenbein)
microneedle was introduced into the middle ear through a puncture hole in the tympanic membrane. It was controlled by a micromanipulator (DC3314R, World Precision Instruments, Sarasota, FL, USA) with 3 degrees of freedom and a $37 \times 20 \times 20$ mm$^3$ workspace with 0.1 mm precision.

AR was implemented by combining real-time video images of the external auditory canal and the tympanic membrane to the 3D CT-scan reconstruction of the middle ear cavity (Fig. 1). The software was developed using OpenCV, Eigen libraries and Ximea API in XCode (C++). The program was run on an iMac computer (2.9 GHz Intel Core i5 processor, 8 GB 1600 MHz DDR3 RAM, NVIDIA GeForce GT 750 M 1024MB graphic card, OSX Yosemite 10.10.5 operating system). The system involved 3 main processes: initial registration, microscope movement tracking and instrument identification (Fig. 2).

The inputs of the system were the following:

1) The video was the real-time film of the tympanic membrane acquired through the microscope (Fig. 3a).
2) The camera matrix represented extrinsic and intrinsic properties of the camera. The focal length of the camera, used in 3D pose estimation, was determined using Zhang’s algorithm$^{32}$.
3) The reconstructed middle ear image (behind the tympanic membrane) was obtained from the preoperative CT scan through Osirix’s 3D virtual endoscopy function (Fig. 3c).

Initial registration. This first step consisted of registering the reconstructed CT-scan image to the real-time microscopic image of the tympanic membrane extracted from the video. From the CT-scan image, fiducial markers were extracted using contrast enhancement and thresholding (Fig. 3(d)). Marker centre points, obtained by detecting blob-like contours in the image using topological structural analysis$^{33}$, were highlighted on the reconstructed image for assistance (Fig. 4(a)). Their corresponding points were manually selected in the microscopic image. The reconstructed image was then warped onto the microscopic image using RANdom SAmple Consensus (RANSAC) based homography$^{34,35}$.
where $P_i$ is the registration matrix, $P_i'$ are the detected marker points in CT image and $P_i$ are their corresponding points in microscope image determined using a RANSAC approach. This algorithm finds the best possible correspondence between points in a small neighbouring window of the selected fiducial points to minimize the registration error. From Eq. (1), $H_R$ can be determined by minimizing the error function. An ellipse shaped mask (used to filter out non planar features in microscope tracking) was also extracted using these corresponding marker points.

A blend operator was also integrated into the system to allow the user to control the opacity of the registered CT image over the microscope video during surgery (as per requirements):

$$I_{AR} = \beta I_M + (1 - \beta) I_{CT}$$

where $I_{AR}$ is the augmented reality output, $I_M$ is the microscope image, $I_{CT}$ is the registered CT image and $\beta \in [0,1]$ is the blend factor.

**Microscope movement tracking.** A robust estimation of the operative microscope movements solely based on image features was developed in order to maintain correspondence with the CT-scan image\textsuperscript{22}. A tracking scheme, comprising of RANSAC and nearest neighbour based Speeded-Up Robust Features (SURF) matching process was employed to determine transformation between consecutive frames\textsuperscript{36–38}. SURF is an algorithm which uses different mathematical formulations to extract information about key points in an image (e.g. corners and edges). The feature-matching algorithm compared all the key points between consecutive frames using random sampling based on feature distance. Any key point that had more than one close matches was not considered for determining the transformation. The ellipse-shaped mask, generated in the initial registration step using fiducial markers, was used to mask out non-planar features in the microscope tracking.

**Figure 2.** Flow diagram of the proposed methodology. Sub-processes belonging to each main step are grouped together with similar box styles. See text for details.
marker centre points, was used to further refine the transformation by filtering out any non-planar features present outside the eardrum\(^2\). A chained homography framework (cumulating the transformations between all the previous frames:

\[
H = H_T H
\]  

where \(H\) is the cumulative homography and \(H_T\) is the transformation between current and previous frames) was used to warp the registered reconstructed image onto the current microscopic frame.

**Surgical instrument tracking.** Three collinear colour markers were painted on the surgical instrument. The proposed instrument identification approach assumed that no instrument was present in the first microscope frame. The first frame underwent a transformation based on the homography \(H\) and was subtracted from the current frame to obtain an approximation of the area occupied by the instrument. A pruning step was carried out to eliminate any false positive regions (due to discrepancies in \(H\)). If only a small area was obtained (less than a threshold), it indicated that no instrument was present in the current frame. Otherwise, the instrument entry point was then searched in the approximated instrument region (on the frame boundary points only). The collinear markers were extracted from the image using colour thresholding followed by pruning. However, due to small focus range of the microscope, the extraction was not perfect, and the centre points were extracted using blob detection\(^1\). The tool entry point was then used to associate the marker centres to marker labels \(B\), \(C\) and \(D\) where \(B\) is closest to the instrument tip \(A\), and \(D\) is closest to the tool entry point. A Kalman filter was used to
refine the marker centre points, eliminating any residual degradation caused by the blurring effect. This filter is a mathematical algorithm which estimates the state of a system from its dynamic model and a series of partial or distorted observed measurements over time. The instrument tip location can then be deduced as:

\[
    a = \frac{1}{3} \left( b + c + d + \frac{AB}{CD} (c - d) + \frac{AC}{BD} (b - d) + \frac{AD}{BC} (b - c) \right)
\]

where \( a \) is the projection of the instrument tip \( A \) on the 2D image frame, \( b, c \) and \( d \) are projections of the markers and alphabet pairs represent physical distances between corresponding markers. Three-point perspective framework was used to estimate 3D pose of the instrument. By setting the focal length of the camera as the \( z \) coordinate of the image projection points \( (b, c \) and \( d) \) and measuring the physical distance between markers \((AB, BC, CD)\), the position of the instrument tip could be estimated using Eq. (4), by fitting the physical geometry \((3D)\) of the tool onto the projected lines \(Ob, Oc\) and \(Od\), where \( O \) is the origin of the camera axis.

**Evaluation.** A quantitative analysis was performed to evaluate the registration process which involved computing the distance between the positions of fiducial points input by the user and their corresponding points (after registration). The root mean square error was used to quantify the results.

The tracking accuracy of the system was evaluated every 30 seconds for 2 minutes. The microscopic view was translated, rotated, zoomed-in and out with an approximate speed of 5–10 mm/s during this period (Table 1). The distances between the “real” positions of the markers and their estimated positions (by the system) were checked. The word “real” was used because these points were computed using template matching algorithm. This algorithm consisted of taking the neighbourhood of the corresponding point as a template and estimating its location in the current frame based on different transformations (scaling, translation and rotation).

For evaluation of the surgical instrument tracking, the position provided by the robotic manipulator attached to the instrument was used as the reference: Pre-known displacements of 2, 4 and 6 mm were applied, independently in each optical axis, using the micromanipulator. Their corresponding displacements detected by the system were measured and the instrument tracking error was computed as the root mean square error of the difference. Averages of 50 samples per individual displacement were recorded for analysis. Additionally, a statistical one-way ANOVA test was carried out to compare the inter-axis pose estimation performance in each individual axis. A p-value <0.05 was considered as significant.

A second series of experiments were performed to further assess the performance of the system. In these five experiments, the movement tracking was assessed for a longer time period (8 minutes) with different experimental conditions to further approach surgical conditions. A surgical microscope (Zeiss OPMI MDO SS Microscope, Zeiss, US) was employed for these experiments. The experiments were performed in different lighting conditions and liquid red ink was introduced to simulate hemorrhage. Movements similar to previous set of experiments were applied to the microscope and the tracking accuracy was measured accordingly. Experimental conditions for these experiments are listed in Table 2. One-way ANOVA test was carried out to compare the tracking results between different experiments. A p-value <0.05 was considered as significant.

In three additional experiments (phantom models: TF-bm, TF-ba, TF-dc), the instrument was firstly placed on the umbo and then introduced into the middle ear space through a small puncture in the tympanic membrane. The tip was placed on the extremity of the long process of incus and finally the highest point on the round window niche. Micro-droplets of ink were injected at these target points. The tympanic membrane was then removed and locations of the droplets were verified by computing the distance between their actual locations and the expected ones.

**Results**

The system remained stable in all cases throughout the experiments (see Supplementary Video S1). Different stages of the experimental study are depicted in Fig. 4 and the augmented reality display window is depicted in Fig. 1. A global mean image refresh rate of 12 ± 1 frames per second (fps) was obtained.

| Experiment Number | A1 | A2 | A3 | A4 | A5 |
|-------------------|----|----|----|----|----|
| Phantom model     | TF-ba | TF-dc | TF-bm | TF-bm | TF-bm |
| Corresponding patient age | 55 | 2 | 55 | 55 | 55 |
| Number of markers used | 5 | 5 | 5 | 5 | 5 |
| Movement speed    | High | High | Low | Low | High |
| Ambient lighting  | High | Medium | Medium | Medium | High |
| Microscope light  | High | Medium | Medium | High | High |
| Comments          | 1. External object (wire) was introduced between 3 and 4 minutes. A big jerk was applied at 2.15 minutes requiring re-registration | — | — | 1. External object (needle) was introduced between 3 and 6 minutes. Liquid ink was introduced at 5 minutes | Liquid ink was present throughout the experiment |

Table 2. Additional experimental conditions. Additional experimental conditions. The experiments were performed for a duration of eight minutes each, using fiducial markers as reference points for evaluation. During experiments, translations, rotations and pitches were applied to the microscope. Further details on temporal bone phantoms are available at https://www.phacon.de/en/hno/felsenbein.
Fiducial marker based initial registration eased and speeded up the process of corresponding point selection. A mean registration error of $0.21 \pm 0.10$ mm ($n=6$) with a mean registration time of $5.57 \pm 2.65$ seconds (range: 1.2–8.3 seconds) was noted.

The microscope tracking process also yielded a sub-millimetric drift ($0.04 \pm 0.07$ mm at 120 seconds), suggesting a very slow propagation error (Fig. 5). Similarly, in additional experiments (with simulated surgical conditions), an average drift of $0.04 \pm 0.11$ mm at 8 minutes was obtained (Fig. 6). The system maintained synchronization in all the experiments. No significant difference in the performance was observed between experiments (p-value, non-significant). During experiment A1, a sudden jerk was applied at 2.15 minutes to the microscope in order to check the limitations of the system. Consequently, a re-registration was required as the system could not comprehend extreme movements (such as jerks). In experiment A4, liquid ink covered two of the registration points which were used for determining the registration and tracking errors. Thus, the performance could only be evaluated qualitatively. The experiment with the introduction of liquid red ink can be seen in Fig. 4(c,d) and Supplementary Video S1.

**Figure 5.** Mean initial registration and tracking errors during the 2 minute tracking of microscope movements. Values represent mean ± standard deviation ($n=6$).

**Figure 6.** Tracking errors for additional experiments during the 8 minute tracking in different experimental conditions using a surgical microscope.
The surgical instrument was also accurately identified. In different experiments, a small oscillatory instrument movement was observed. Different displacements were applied in each individual axis and the 3D pose was estimated (Table 3). A mean instrument tip position error of $0.19 \pm 0.05$ mm ($n = 150$) in X-axis, $0.19 \pm 0.02$ mm ($n = 150$) in Y-axis and $0.55 \pm 0.46$ mm ($n = 150$) in Z-axis was observed. Tukey HSD post-hoc test revealed that the pose estimation in X and Y axes was acceptable and significantly better than the estimation in Z axis ($p$-value $< 0.05$ when compared with both X and Y axes) as small deviation in instrument identification constitutes a large deviation in Z axis pose estimation. No significant difference was observed between X and Y axes pose estimations. The mean pose estimation error $\sqrt{X^2 + Y^2 + Z^2}$ was $0.33 \pm 0.22$ mm ($n = 450$).

Similarly, the target structures were accurately reached with mean localization errors ($n = 3$) of $0.56 \pm 0.14$ mm, $0.54 \pm 0.16$ mm and $0.46 \pm 0.19$ mm for umbo, incus tip and round window niche, respectively (Fig. 7). The mean target error was $0.52 \pm 0.15$ mm ($n = 9$).

**Discussion**

In this study, we showed that a marker-based AR system combining preoperative CT image with real-time 2D video from the operative microscope, based only on computer vision and without any tracking system techniques, is possible. A 3D reconstructed view of the CT-scan was registered to the microscopic view based on homography transformation. The system employed algorithms from different domains (e.g. image processing, visual perception, endoscopy, radiology and autonomous navigation). It provided additional visual information on the middle ear structures and the surgical instrument with submillimetric precision, compatible for middle ear surgery.

In previous works, the performance of the motion tracking with manual initial registration was analysed on phantom and cadaver subjects. Similar results were obtained for both types of subjects. This study, provides crucial steps toward the applicability of AR on middle ear in the operating room by enhancing the registration step, enabling the system to process the video in real-time and to automatically detect instruments. The CT to video registration appears to be crucial, since errors during this step will propagate throughout the process. Indeed, in most computer-assisted surgical systems, image registration plays paramount role in the overall performance of the system. In endoscope-CT registration, combinations of different intensity-based schemes such as cross-correlation, squared intensity difference, pattern intensity, normalised and gradient mutual information have shown promising results. Similarly, feature-based schemes involving natural landmarks, contour based feature points, iterative closest point and k-means clustering have also been exploited. The main challenge of our system was the low similarity between the multi-modal images. To overcome this, artificial markers that increase visibility and lower the perturbation were introduced. Indeed, very few natural landmarks are visible on both CT and during otoscopy around the tympanic membrane, making introduction of markers beneficial in

| Movement Direction | Actual Displacement |
|--------------------|---------------------|
|                    | 2 mm    | 4 mm    | 6 mm    |
| X (mm)             | $2.13 \pm 0.01$ | $3.75 \pm 0.04$ | $6.18 \pm 0.10$ |
| Y (mm)             | $1.95 \pm 0.04$ | $4.07 \pm 0.01$ | $5.54 \pm 0.02$ |
| Z (mm)             | $2.35 \pm 0.56$ | $3.65 \pm 0.21$ | $6.96 \pm 0.61$ |

Table 3. Accuracy of surgical instrument tracking. 3D pose estimation of the microneedle after predetermined displacements by a micromanipulator. Data represent mean ± standard deviation of 50 samples for each direction and displacement.

Figure 7. Qualitative analysis of the localisation of the injected droplets. A right temporal bone (TF-ba) is shown in operative position. The black marker dots (arrow heads) represent the targets where the ink droplets were injected. CT: Corda tympani, Inc.: Incus, M: malleus, RW: round window, St: Stapes.
terms of both registration time (10–15 seconds vs 55–80 seconds) and accuracy (0.21 mm vs 0.25 mm) as compared to manual registration35. In the current setup, the markers were placed arbitrarily around the periphery of the tympanic membrane which was assumed to represent a near-planar surface. As a future step for clinical trials, patient specific custom rings in contact with the tympanic membrane can be designed to house the markers in an ergonomic and robust manner. Even with millimetric markers, finding the exact corresponding points is practically infeasible. This limitation was overcome by implementing a RANSAC algorithm in the registration process. This mathematical algorithm allows estimating the parameters of a model with iterative measures and possible aberrant values. In our model, the algorithm took into account similar points in a small neighbouring window.

Another challenge was to maintain correspondence between CT and video by tracking the microscope movements. In routine applications, the microscope will be quasi-static. However, in order to validate the robustness of the system, various movements were applied to the microscope. With the use of SURF algorithm, a minimal propagation error was observed during tracking, even when intricate motion was applied, allowing lengthy surgical procedures. Since the tracking was based on image features, the fiducial markers do not need to be visible in the surgical video after registration step has been successfully carried out and this represents a potential advantage in terms of ergonomics. Virtual objects introduce additional unwanted occlusions leading to loss of internal organ information in an AR system. The blend operator allows the surgeon to turn off or decrease the opacity of virtual image when it is not required. Moreover, the raw video from the microscope is also available to the surgeon next to the augmented reality display. Furthermore, to address the loss of information on internal organs, a combination of transtympanic endoscopy and AR may be utilized in the operating room, in order to validate the AR information and to explore details that are less visible on CT-scan data such as adhesions. Transtympanic endoscopy has already been evaluated in similar key-hole procedures.

Keyhole surgery cannot be performed without instrument depth information inside the middle ear and behind the tympanic membrane. Under operative microscope, conventional computer vision approaches exploiting natural features like gradient information or greyish nature of the surgical instruments are bound to fail as the perception range of microscopes is limited. In addition, since the instrument may enter from any direction and protrude indefinitely, so geometric priors may not be valid. Our proposed method, using colour markers, took into account such specifications of the otologic surgery. Although the markers can be placed anywhere on the instrument, the segment containing the markers needs to remain in the video frame for accurate pose estimation. However, this method cannot determine instrument pose angle in the optical axis, without introduction of additional priors e.g. coplanar markers.

The accuracy of the system on 3 middle ear target structures was submillimetric and this level of precision is essential for otologic procedures. This is the most important performance factor as it encompasses all different aspects of the system: precision of CT reconstruction, registration, motion and instrument tracking. This performance may be improved by integrating additional 3D information about target structures. Moreover, since the tracking was based on image features, the fiducial markers do not need to be visible in the surgical video after registration step has been successfully carried out and this represents a potential advantage in terms of ergonomics. Virtual objects introduce additional unwanted occlusions leading to loss of internal organ information in an AR system. The blend operator allows the surgeon to turn off or decrease the opacity of virtual image when it is not required. Moreover, the raw video from the microscope is also available to the surgeon next to the augmented reality display. Furthermore, to address the loss of information on internal organs, a combination of transtympanic endoscopy and AR may be utilized in the operating room, in order to validate the AR information and to explore details that are less visible on CT-scan data such as adhesions. Transtympanic endoscopy has already been evaluated in similar key-hole procedures.

Conclusion

In conclusion, the proposed AR system based only on computer vision techniques provided a precise vision of the middle ear contents and the surgical instrument behind the closed tympanic membrane in real-time with a high image fresh rate. The system maintained correspondence between CT-scan and video during microscope movements. This technique opens insights to different transtympanic procedures such as drug administration, labyrinthine fistula repair and ossicular chain reconstruction by a transtympanic keyhole approach.

Data availability

The data i.e. the phantoms that support the findings of this study are available from Phacon Inc., Leipzig, Germany (https://www.phacon.de/en/hno/felsenbein).

Received: 3 October 2019; Accepted: 26 March 2020; Published online: 21 April 2020

References

1. Schwam, Z. G., Kaul, V. Z., Cosetti, M. K. & Wanna, G. B. The utility of intraoperative navigation of the temporal bone for otolaryngology resident training. Laryngoscope, https://doi.org/10.1002/lary.28219 (2019).
2. Schwager, K. Acute complications during middle ear surgery: part 1: Problems during tympanoplasty—what to do? HNO. 55(4), 307–315 (2007).
3. 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).
4. Kakehata, S. Transtympanic endoscopy for diagnosis of middle ear pathology. Otolaryngol. Clin. North Am. 46(2), 227–232 (2013).
5. 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).
6. Bozzato, A., Bozzato, V., Al Kadah, B. & Schick, B. A novel multipurpose modular mini-endoscope for otology. Eur. Arch. Otorhinolaryngol. 271(12), 3341–3348 (2014).
7. Aukstakalnis, S. Practical Augmented Reality: A Guide to the Technologies, Applications, and Human Factors for AR and VR. (Addison-Wesley Professional, 2016).
8. Vävra, P. et al. Recent development of augmented reality in surgery: a review. J. Healthc. Eng.; https://doi.org/10.1155/2017/4574172 (2017).
9. Wong, K., Yee, H. M., Xavier, B. A. & Grillone, G. A. Applications of augmented reality in otolaryngology: A systematic review. Otolaryngol. Head Neck Surg. 159(6), 956–967 (2018).
10. Russain, R., Lalander, A., Guigo, C. & Bozorg, A. G. Contribution of Augmented Reality to Minimally Invasive Computer-Assisted Cranial Base Surgery. IEEE J. Biomed. Health Inform.; https://doi.org/10.1109/JBHI.2019.2954003 (2019).
11. Lee, J. et al. Clinical Utility of Intraoperative Tympanomastoectomy Assessment Using a Surgical Microscope Integrated with an Optical Coherence Tomography. Sci. Rep. 8(1), 17432 (2018).
12. Wisotzky, E. L. et al. Interactive and Multimodal-based Augmented Reality for Remote Assistance using a Digital Surgical Microscope. IEEE Conference on Virtual Reality and 3D User Interfaces (VR), 1477–1484, https://doi.org/10.1109/VR.2019.8797682 (2019).

13. Liu, W. P. et al. Cadaveric feasibility study of da vinci si–assisted cochlear implant with augmented visual navigation for otologic surgery. JAMA Otolaryngol. Head Neck Surg. 140(3), 208–214 (2014).

14. Miroir, M. et al. RobOtol: from design to evaluation of a robot for middle ear surgery. IEEE/RSJ International Conference on Intelligent Robots and Systems, 850–856, https://doi.org/10.1109/IROS.2010.5650390 (2010).

15. Hussain, R. et al. Real-time augmented reality for ear surgery. Med. Image Comput. Comput. Assist. Interv. 11073, 324–331, https://doi.org/10.1007/978-3-030-00937-3_38 (2018).

16. Bong, J. H. et al. Endoscopic navigation system with extended field of view using augmented reality technology. Int. J. Med. Robot. 14(2), 1886 (2018).

17. Citardi, M. J., Yao, W. & Luong, A. Next-generation surgical navigation systems in sinus and skull base surgery. Otolaryngol. Clin. North Am. 50(3), 617–632 (2017).

18. Hata, N. et al. Image guided microscopic surgery system using mutual-information based registration. International Conference on Visualization in Biomedical Computing, 317–326, https://doi.org/10.1007/978-0-0406969 (1996).

19. Chu, Y. et al. Registration and fusion quantification of augmented reality based nasal endoscopic surgery. Med. Image Anal. 42, 241–256 (2017).

20. Katli, D. et al. A system for context-aware intraoperative augmented reality in dental implant surgery. Int. J. Comput. Assist. Radiol. Surg. 10(1), 101–108 (2015).

21. Lapeer, R. J. et al. Using a passive coordinate measurement arm for motion tracking of a rigid endoscope for augmented-reality image-guided surgery. Int. J. Med. Robot. 10(1), 65–77 (2014).

22. Marroquin, R., Lalande, A., Hussain, R., Guigou, C. & Grayell, A. B. Augmented reality combining otoendoscopy and high resolution temporal bone CT scan. Otol. Neurotol. 39(8), 931–939 (2018).

23. Murugesan, Y. P., Alsadoon, A., Manoranjan, P. & Prasad, P. W. A novel rotational matrix and translation vector algorithm: geometric accuracy for augmented reality in oral and maxillofacial surgery. Int. J. Med. Robot. 14(3), 1889 (2018).

24. Wang, J., Suegura, H., Yang, L., Kobayashi, E. & Sakuma, I. Video see-through augmented reality for oral and maxillofacial surgery. Int. J. Med. Robot. 13(2), 1754 (2017).

25. Climent, J. & Mars, P. Automatic Instrument Localization in Laparoscopic Surgery. In Progress in computer vision and image analysis 73 (eds. Bunke, H., Villanueva, J. J., Sanchez, G. & Otafu, X.) 123–136 (World Scientific, 2009).

26. Doignon, C., Graebeling, P. & De Mathelin, M. Real-time segmentation of surgical instruments inside the abdominal cavity using a joint hue saturation color feature. Real-Time Imaging 11(5), 429–442 (2015).

27. Allan, M. et al. Toward detection and localization of instruments in minimally invasive surgery. IEEE Trans. Biomed. Eng. 60(4), 1050–1058 (2013).

28. Nageotte, F., Zanne, P., Doignon, C. & De Mathelin, M. Visual servoing-based endoscopic path following for robot-assisted laparoscopic surgery. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2364–2369, https://doi.org/10.1109/IROS.2006.282647 (2006).

29. Labadie, R. F. In vitro assessment of image-guided otologic surgery: submillimeter accuracy within the region of the temporal bone. Otolaryngol. Head Neck Surg. 132(3), 435–442 (2005).

30. Shin, S. et al. A single camera tracking system for 3D position, grasper angle, and rolling angle of laparoscopic instruments. Int. J. Precis. Eng. Man. 15(10), 2153–2160 (2014).

31. West, J. B. et al. Fiducial point placement and the accuracy of point-based, rigid body registration. Neurosurgery 48(4), 810–817 (2001).

32. Zhang, Z. A flexible new technique for camera calibration. IEEE Trans. Pattern Anal. Mach. Intell. 22(11), 1330–1334 (2000).

33. Hartley, R. & Zisserman A. Multiple view geometry in computer vision 2nd ed. (Cambridge University Press, 2003).

34. Suzuki, S. Topological structural analysis of digitized binary images by border following. Comput. Vis. Image Underst. 33(2), 185–193 (1986).

35. Fischler, M. A. & Bolles, R. C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM. 24(6), 381–395 (1981).

36. Muja, M. & Lowe, D. G. Fast approximate nearest neighbors with automatic algorithm configuration. International Conference in Computer Vision Theory and Applications (VISAPP) 12(2), 331–340, https://doi.org/10.5220/0001787803310340 (2009).

37. Bay, H., Ess, A., Tuytelars, T. & Van Gool, L. Speeded-up robust features (SURF). Comput. Vis. Image Underst. 110(3), 346–359 (2008).

38. Kalman, R. E. A new approach to linear filtering and prediction problems. J. Basic Eng. 82(1), 35–45 (1960).

39. 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).

40. 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).

41. Yin, Y., Wakid, M., Kirmizibayrak, C., Bielanowicz, S. & Hahn, J. 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).

42. Osaka, T. et al. Rendering-based video-CT registration with physical constraints for image-guided endoscopic sinus surgery. Proc. SPIE Int. Opt. Eng. 9415, https://doi.org/10.1117/12.2081753 (2015).

43. Jun, G. X., Li, H. & Yi, N. Feature points based image registration between endoscope image and the CT image. IEEE International Conference on Electric Information and Control Engineering, 2190–2193, https://doi.org/10.1109/ICEICE.2011.5778261 (2011).

44. Wengert, C., Cattin, P., Du, J. M., Baur, C. & Szekely, G. Markerless endoscopic registration and referencing. Med. Image Comput. Comput. Assist. Interv. 14190, 816–823, https://doi.org/10.1109/ISBI.2016.7595852 (2016).

45. Habermehl, D. et al. Evaluation of different fiducial markers for image-guided radiotherapy and particle therapy. J. Radiat. Res. 54(suppl_1), 161–168 (2013).

Acknowledgements
The author would like to thank Oticon Medical, Vallauris, France for their financial support. Authors would also like to thank Mr. David Fofi for his technical advice.

Author contributions
R.H.: writing manuscript, developing software (registration and instrument tracking), experiment design, performing experiment. A.L.: initial idea, writing manuscript, resource management, data analysis. R.M.: developing software (motion tracking and GUI), experiment design. C.G.: performing experiments, sample preparation and CT scans, medical expertise. A.B.G.: initial idea, experiment design, resource management, medical expertise. All authors reviewed the manuscript.
Competing interests
The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-63839-2.
Correspondence and requests for materials should be addressed to R.H.
Reprints and permissions information is available at www.nature.com/reprints.
Publisher’s note Springer Nature remains neutral with regard to jurisdical claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2020