

# PROCEEDINGS OF SPIE

[SPIDigitalLibrary.org/conference-proceedings-of-spie](https://SPIDigitalLibrary.org/conference-proceedings-of-spie)

## Image-based measurement by instrument tip tracking for tympanoplasty using digital surgical microscopy

Gard, Niklas, Rosenthal, Jean-Claude, Jurk, Silvio, Schneider, Armin, Eisert, Peter

Niklas Gard, Jean-Claude Rosenthal, Silvio Jurk, Armin Schneider, Peter Eisert, "Image-based measurement by instrument tip tracking for tympanoplasty using digital surgical microscopy," Proc. SPIE 10951, Medical Imaging 2019: Image-Guided Procedures, Robotic Interventions, and Modeling, 1095119 (8 March 2019); doi: 10.1117/12.2512415

**SPIE.**

Event: SPIE Medical Imaging, 2019, San Diego, California, United States

# Image-based measurement by instrument tip tracking for tympanoplasty using digital surgical microscopy

Niklas Gard<sup>a,b</sup>, Jean-Claude Rosenthal<sup>a</sup>, Silvio Jurk<sup>a</sup>, Armin Schneider<sup>c</sup>, and Peter Eisert<sup>a,b</sup>

<sup>a</sup>Fraunhofer Heinrich-Hertz Institute, Berlin, Germany

<sup>b</sup>Humboldt University of Berlin, Berlin, Germany

<sup>c</sup>ARRI, Arnold & Richter Cine Technik GmbH & Co. Betriebs KG, München, Germany

## ABSTRACT

We propose a new method to support tympanoplasty operations for ear, nose and throat (ENT) surgery, namely the reconstruction of the eardrum. This intervention needs precise distance and contour measurements to allow a successful operation. Currently, surgeons only have limited tools to measure patient specific anatomy and rely on their experience and visual judgement to estimate anatomical dimensions. Therefore, we provide an image-based augmented reality (AR) measuring tool using a complete digital processing chain, giving instant feedback about anatomical structures with high metric accuracy.

Our method detects the center of gravity of a marked spherical instrument tip in the stereoscopic image pair of a digital surgical microscope and triangulates points in 3D space of the calibrated stereo system. We track the tip using a self-updating template-matching algorithm. An accurate subpixel refinement of the tip center prevents drift and guarantees highly accurate stereo correspondences. GPU implementation and a color-based pre-detection allows real-time tracking in high-resolution images. Reconstructed trajectories form a true-scale virtual stencil, which is projected directly into the surgeons field of view as a precise registered AR overlay. This overlay supports the surgeon while cutting a patient specific shape from autologous tissue.

Measurement accuracy and real-time tracking performance are evaluated using a depth-of-field test body and a temporal bone model, where the obtained 3D path-reconstruction is compared to a CT scan. Our approach provides great potential to improve state-of-the-art surgical workflows by reducing operating times and facilitating intraoperative decisions.

**Keywords:** surgical tracking tool, ENT surgery, tympanoplasty, 3D imaging, augmented reality, digital surgical microscopy, image-based measurement

## 1. INTRODUCTION

Digital surgical microscopes offer a complete digital processing chain, enabling new options of seamless integrated image processing algorithms and information visualization for image-guided interventions and therapies. The digital binocular represents an enhanced augmented reality (AR) device, as overlays can be perfectly registered to captured images if an accurate stereo calibration is provided.

Tympanoplasty is a common and widely used treatment in ear, nose, throat (ENT) surgery, e.g. over 21000 interventions have been performed 2017 in Germany.<sup>1</sup> During the intervention, mostly autologous tissue needs to be cut to fit the patients' specific anatomy so that it can be used as an eardrum replacement. Therefore, additional methods have been proposed to reduce the cutting time<sup>2</sup> or to measure the size of the tympanic membrane contour to achieve a better outcome for the patient.<sup>3,4</sup> In this work and in contrast to mentioned methods, we describe an entirely image-based measuring system for a tympanoplasty operation to support the replacement of the tympanic membrane. The procedure itself is a time-consuming tedious iterating back-and-forth process to see if the cut shape fits the patient's eardrum dimensions. Our method allows the surgeon to facilitate this process and to optimize the fit size. A virtual eardrum contour is obtained by moving a marked spherical instrument tip along the anatomical contour (Fig.1(a)). The 3D shape is reconstructed on-the-fly by

---

Further author information:

Niklas Gard: E-mail: niklas.gard@hhi.fraunhofer.de

Medical Imaging 2019: Image-Guided Procedures, Robotic Interventions, and Modeling,  
edited by Baowei Fei, Cristian A. Linte, Proc. of SPIE Vol. 10951, 1095119 · © 2019 SPIE  
CCC code: 1605-7422/19/\$18 · doi: 10.1117/12.2512415

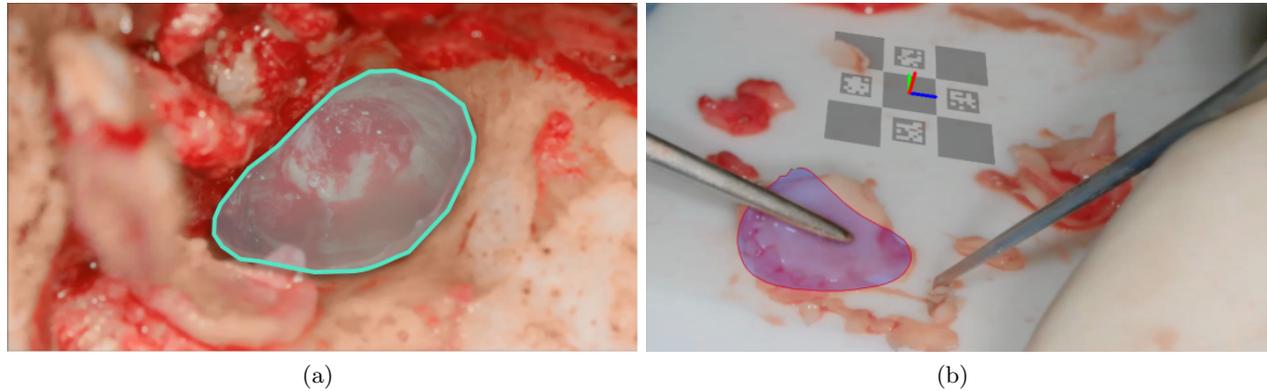


Figure 1. (a) Conceptual visualization showing a highlighted contour in the ear canal, which could be selected with the instrument tip. (b) The selected contour is projected into the surgeon's field of view, next to the cutting area.

precise and constant tracking and stereo triangulation of the refined detection point in the left and the right microscope image. We plan to display the resulting 3D-stencil directly into the surgeon's field of view, next to the cutting area during the intervention as shown in Fig. 1(b). This provides a hint to cut and verify a well-fitted eardrum replacement on the operation table. Tedious cutting combined with back and forth testing will be avoided.

The microscope used for this work is the ARRISCOPE.<sup>5</sup> It is to our knowledge the only medical approved entirely digital surgical microscope showing the microscope image on a video see-through OLED binocular. Seen as an AR device, this makes the microscope come close to the *ultimate display* as it has been described by AR-pioneer Ivan Sutherland in 1965, because the image is seen from the users natural perspective.<sup>6</sup> Hence, augmentation can be presented in an immersive environment and can be integrated into a seamless workflow.

## 2. RELATED WORK

In surgical microscopy and laparoscopic surgery, a surgeon works with additional operational instruments, so using them as input device to perform measurements is a very natural form of interaction. During the last years instrument tracking for minimally invasive surgery has been a frequently discussed topic. A detailed literature review about different visual methods and applications is given by Bouget et al.<sup>7</sup> Although a lot of articles focus on tracking in laparoscopic images<sup>8-11</sup> or retinal microsurgery<sup>12,13</sup> visual instrument tracking for ENT surgery has not been discussed, to our knowledge, yet. In laparoscopic surgery, a main research focus is the ability to detect arbitrary instruments by using machine learning (e.g. Random Forest Segmentation, SVM) methods.<sup>8,9</sup> In contrast, in our case the instrument is known and we clearly focus on precise and accurate 3D measurement tasks. Furthermore 3D measurement tasks for laparoscopy as described in<sup>14</sup> naturally work on a different scale than in microscopy and rather do not need sub-millimeter accuracy as it necessary for our application.

Rosenthal et al. use image-based stereo measurement in the context of ENT surgery to facilitate the selection of an appropriately-sized stapes prosthesis by evaluating manually selected points in dense disparity maps.<sup>15</sup> A three dimensional analysis of the human ear-drum in order to register its shape is given by Karygiann et al.<sup>16</sup> The authors capture the shape information of the tympanic membrane with a special Light Field Oscope.<sup>17</sup>

As we address a digital processing chain, we also see a great potential in combining the abilities of AR and microscopy as described e.g. by Aschke et al.<sup>18</sup> In the field of medical applications, AR is usually based on passive markers. External cameras are used to register the patient and pointing devices with the viewing pose.<sup>19</sup> Up to today the augmentation in such a system is possible, but the additional information is only visible on external screens. Because no additional cameras or trackers are needed in our system, the synchronization effort is reduced to a minimum. A related research project dealing with AR in the context of stereo-video see-through microscopy is *Scope+*.<sup>20</sup>

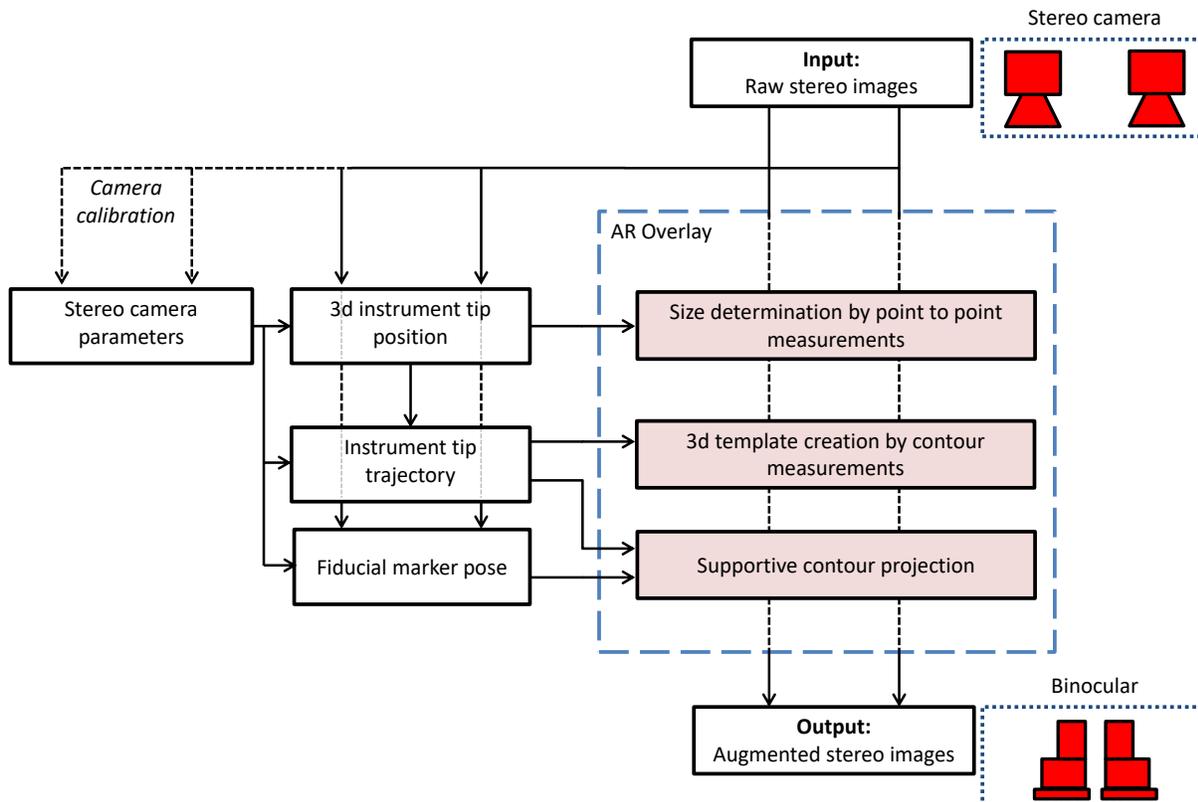


Figure 2. Workflow for stereo-based measurement and image augmentation in an ENT scenario. Only one of the color filled boxes is processed at a time.

### 3. METHOD

Fig. 2 depicts the complete measurement workflow proposed in this work. The position of the measurement tip is reconstructed in 3D space using the exact detection position in the left and the right image and the stereo camera parameters. The detection in two arbitrary image pairs allows a point to point size measurement, tracking the tip for a number of successive frames leads to a 3D contour, which can be displayed as a support, while cutting tissue.

#### 3.1 Description of the measuring instrument

We use a custom-made instrument with a green spherical tip having a radius of 0.5mm (Fig. 3(a)) as a tactile input device to mark the examined contour. Green was chosen because it is inexistent in anatomy and contrasting with the reddish surgical situs. The idea to use such an instrument instead of other methods for user interaction is because of the intuitiveness and the similarity to the surgical instruments used in the context of ENT surgery. The spherical shape allows an accurate determination of the tip center from different perspectives, assuming an approximately elliptical projection. When working in maximum zoom level (e.g.6×) a radius of 0.25mm is needed for precise positioning. A double sided tool having spheres with 2 radii for different zoom levels would be a reasonable choice. Manipulators having the required shape exist e.g. in dental medicine. A ball-shaped plugger coated with a colored layer at its tip fulfills our requirements.

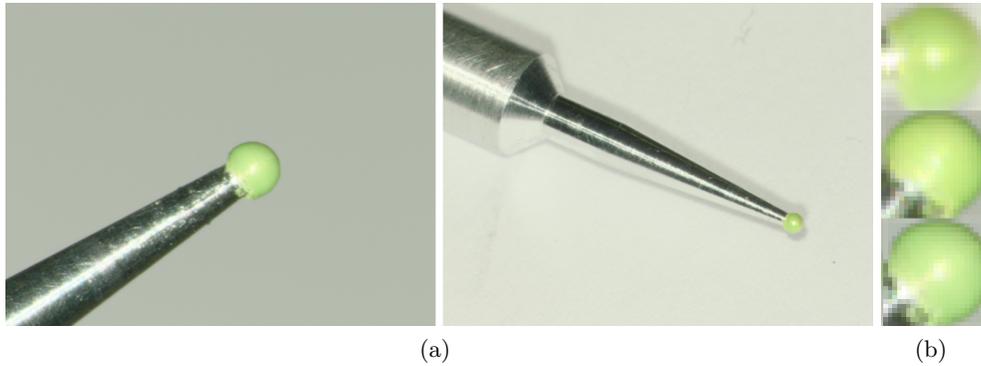


Figure 3. (a) Prototypical spherical instrument tip. (b) An example for an initial template set which is down sampled to the required size of  $w \times w$  pixel.

### 3.2 Calibration and preparation

The image-based measurement method makes use of the stereo triangulation principle. To provide precise measurements, the intrinsic and extrinsic parameters of the stereo system need to be determined by camera calibration. We calibrate the stereo camera integrated in the microscope head with a special checkerboard pattern for close range calibration<sup>21</sup> using long focal lengths. The model-based approach estimates an initial orientation of the calibration pattern from two known reference points and optimizes the camera parameters of a virtual camera-pair in a way that the synthetically generated images of the pattern matches with the observed pattern in the real image. To achieve accurate 3D reconstruction results, the different zoom levels of the microscope are calibrated separately. For the evaluation phase, a fixed focus and zoom level setting is used in minimum and maximum magnification. The used surgical microscope magnifies between  $1.6\times$  (minimum) and  $9.6\times$  (maximum), resulting in an  $6\times$  optical zoom.

When using a measurement tip for the first time, a set of  $n$  image patches  $\{T_1, \dots, T_n\}$  with a fixed size of  $w \times w$  pixel has to be created from an offline sequence. Each patch is cut closely around the perimeter of the sphere in the image to include as little background as possible. We use single images shot in the middle zoom position, cut patches manually and provide some variation, regarding pose and lightning conditions. In our controlled test environment three to five patches are sufficient (Fig. 3(b)). Since the shape of the tip is very simple, we found small patches with  $w = 12$  to be sufficient for image sizes of  $960 \times 540$  pixels.

### 3.3 Instrument tip detection

The detection algorithm can be divided in three stages. First, a binary mask of the region of the tip in the image is generated from a color-based pre-segmentation. A template matching with a set of pregenerated patches is then used to find the approximate location of the center of the instrument tip. Afterwards, a gradient-based circle fitting refines the center position on a subpixel level and one template is updated with information from the image.

#### 3.3.1 Presegmentation

Based on the color distribution of the image patches showing the tip, a binary mask of possible tip locations is created. To get a distinctive representation of the tip color, the RGB template is transformed into the HSV color space and a histogram for the three channels is generated. Assuming that the main part of the patch shows the tip and a smaller part shows the background, the tip color is estimated as the color range between a lower and upper quantile around the median of the histogram (Fig. 4) channels. The quantiles allow a deviation of one third of the standard deviation for the hue channel and two times the standard deviation for the saturation and value channel. By selecting the pixels having a hue value inside this interval, a segmented tip is found inside the patch.

This segmented region is then used to create an oriented bounding ellipsoid inside the HSV colorspace with a principal component analysis (PCA). An intersection check of observed color values with the ellipsoid (Fig. 5),

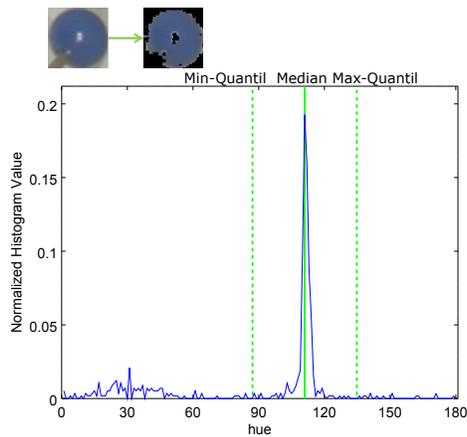


Figure 4. Hue channel histogram with median and min/max quantiles marked. In this example, a different blue colored tip is used.

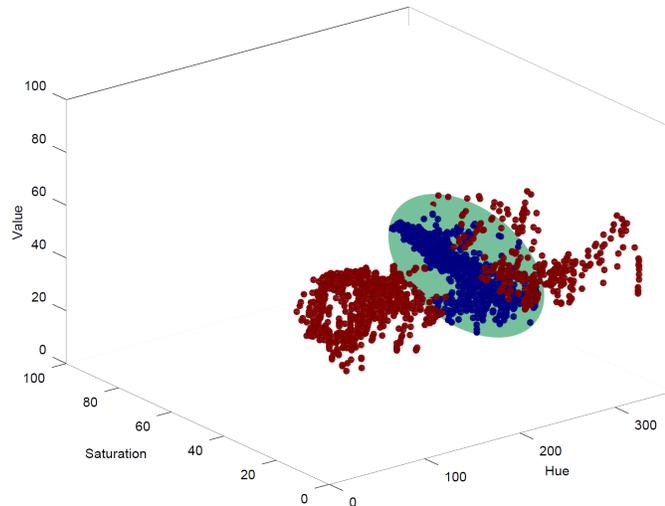


Figure 5. All dots describe HSV values of an initial template with blue dots lying inside and red dots lying outside of the green bounding ellipsoid. The bounding ellipsoid is used for the intersection test with image pixels to generate the binary mask.

implemented with a CUDA kernel, indicates a potential relation of a pixel and the tip. The spatial orientation of the ellipsoid is determined by the center of gravity and the three eigenvectors obtained by PCA. From these values, a rotation matrix is generated. After a transformation to a centered coordinate system using the rotation matrix and the center point, we are able to determine the three elliptical radii of an axis aligned bounding ellipsoid. Later, observed input pixel values are transformed with the same transformation and the ellipsoid is used for the intersection check. By this, the whole input image can be scanned for the signaled tip and a set of possible locations is obtained. Further on, this output is processed with a morphological operator to remove noise and to close holes in the mask.

### 3.3.2 Initial detection

To detect the instrument tip, a GPU-accelerated template matching approach with an arbitrary number of manually created square shaped template patches is used. We implemented a complete parallel and sub-pixel accurate processing chain in CUDA. By using the GPU texture memory, intensity values can be read out on a subpixel level. This allows a coupling of the templates, which have a fixed size, with a free scale factor and to calculate the points of comparison between template and image accordingly. The scale factor  $s$  is determined from the diameter of the instrument tip, the size of the template in pixels and the current zoom and focus level of the microscope. It is varied by a small offset during the matching process to compensate size changes of the tip in the working area. As cost function during the matching, we use the normalized cross correlation (NCC), because of its invariance to global brightness changes. Only pixels which are marked as valid by the mask described in the previous chapter are checked.

For each template, a response candidate is received, which is the output pixel with the highest response value. This template specific response is needed during the update process. To get the matching position, a combined response map is calculated by summing up the template-specific response maps. The pixel position  $\mathbf{p}_{TM}$  having the maximum value in the output map is selected as a first estimation of the sphere center.

### 3.3.3 Match refinement and template adaption

The initial detection results in an integer pixel coordinate, describing the average position of the best matches for each template. Still the quality of 3D reconstruction can be improved by refining the match to subpixel coordinates. Furthermore, when updating the template small inaccuracies will accumulate and can introduce a

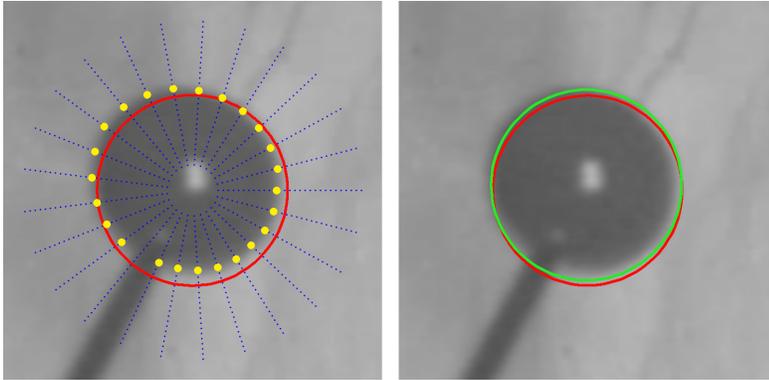


Figure 6. Positional refinement of instrument tip. Starting from the initial estimate (red circle), a new estimate (green circle) is fit to the maximum gradients (yellow dots).

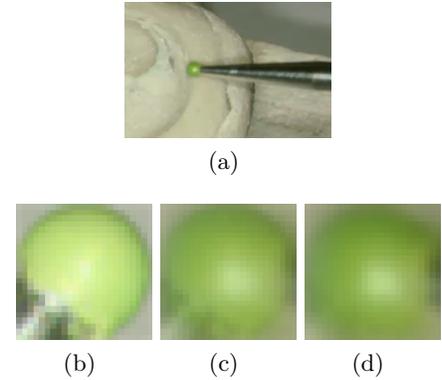


Figure 7. Adaptation to target appearance: (a) Instrument in test environment (b) Initial template (c) Template after 25 frames (d) Template after 50 frames

drift, a known problem for updated template matching.<sup>22</sup> Therefore, an iterative circle fitting algorithm is used to calculate a subpixel coordinate and to match the known geometric shape of the tip with the image.

On the circle (Fig. 6, red circle) around  $\mathbf{p}_{TM}$  with a radius of  $r = 0.5(s \cdot w)$ ,  $l$  points are sampled, each being the center of a scanline through the normal of circle at that point. On each line, intensity values for  $m$  equidistant points (Fig. 6, blue points) are read from the image on a subpixel grid by bi-linear interpolation and form a gray scale profile.

A zero-crossing in the second derivative<sup>23</sup> of the one dimensional grayscale profile is assumed to signalize the point of transition between sphere and background. Furthermore, to reduce the probability of finding zero crossings resulting from image noise, the first derivative has to exceed a predefined threshold. A list of all points found (Fig. 6, yellow points) is fed to a Gauss-Newton optimization, using an M-estimator to achieve robustness to outliers,<sup>24</sup> aiming to find the circle which fits best to the list of hypotheses points (Fig. 6, green circle). If no circle fit can be found or if it is not possible to find a hypothesis on at least half the scanlines a false positive match is recognized and the detection is discarded. Also, if the optimized radius differs too much from the initial input radius  $r$  the detection is discarded.

To make the matching robust to local brightness changes and ensure high match responses independent from the scene, an update mechanism is integrated. Let  $\mathbf{p}_c$  be the corrected 2D match position and  $\{\mathbf{p}_{TM1}, \dots, \mathbf{p}_{TMn}\}$  be the position of the best match for each individual template with  $\{v_1, \dots, v_n\} \in [-1, 1]$  as responses. The template  $T_j$  which is selected to be updated is the one with the shortest spatial distance between  $\mathbf{p}_c$  and  $\mathbf{p}_{TMj}$ . Based on the refined radius, a patch  $T'_j$  is cut from the input image and scaled to the size of  $T_j$ . Let  $w$  be the weighting with which  $T'_j$  can influence the result. The updated patch  $T_{j_{new}}$ , which replaces  $T_j$  is calculated with

$$T_{j_{new}} = (1 - w)T_j + w\left(T_j \frac{1 - v_j}{2} + T'_j \frac{v_j + 1}{2}\right) \quad \text{with } w \in [0, 1]. \quad (1)$$

We always update the best matching template to avoid the forward-propagation of errors. Templates which are not considered for updating over several consecutive frames will be used as a fallback if an error is evolving in the template which has been updated before. We use an empirically found weighting factor of  $w = 0.8$  and the parameters  $l = 30$  and  $m = 1.8r_{TM}$  during our tests. Fig. 7 shows the ability of the update and matching process to adapt to a different tip appearance. While tracking over 50 frames the original template fades into the tip appearance of the target scene, including the tip being slightly out of focus.

### 3.4 Trajectory estimation

After successful detection of the instrument tip,  $\mathbf{p}_c^L$  and  $\mathbf{p}_c^R$  define the estimated 2D subpixel coordinates of the sphere center in the left and the right image. The 3D position can be calculated by stereo triangulation. A rectification of the stereo images is not needed. If lens distortion is present in the images the point coordinates

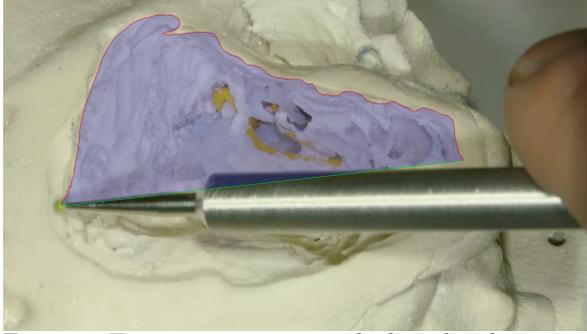


Figure 8. Trajectory creation with the colored instrument tip. The path is visualized as an overlay inside the binocular.

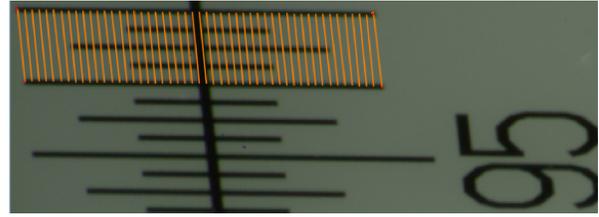


Figure 9. Depth-of-field test body used for accuracy evaluations seen in maximum zoom level. For a target distance of 1 mm 50 point-to-point measurements have been carried out.

have to be undistorted though. The calculated 3D point is located on top of the spherical surface. To find the center of the sphere, it is moved on the line connecting the camera center and the point by the radius of the sphere.

For the first frame of a tracking sequence or if only one single point measurement is needed, template matching and correction are done for both images independently. The vertical offset between the two 2D points may not exceed a given threshold to ensure a correct correspondence. The threshold depends on the adjustment quality of the stereo setup. In our case, an offset of two percent of the image height is allowed as the two native images are nearly horizontally aligned.

The disparity between the detected instrument tip centers in the left and right image at time  $t_n$  is given by  $\mathbf{d}(t_n) = \mathbf{p}_c^L(t_n) - \mathbf{p}_c^R(t_n)$ . It is assumed that the disparity does not change significantly between two consecutive frames. While tracking sequences this assumption allows us to avoid template matching for one image by using  $\mathbf{p}_c^L(t_{n+1}) + \mathbf{d}(t_n)$  as starting point for the optimization algorithm in the other image. We further increase the stability of the tracking by subsequently interchanging  $\mathbf{p}_c^L$  and  $\mathbf{p}_c^R$  in the above calculations for every frame.

When creating a trajectory, all reconstructed 3D points are appended to a vector. Especially if the microscope is in high zoom level, a natural jittering of the manually operated instrument tip will be visible. Thus, a smoothing filter is applied on the 3D point set to compensate for unsteady instrument movements and to create a smooth path.

### 3.5 Live Feedback

A live feedback of the extracted contour and the detected position is displayed on the screen inside the digital binocular. The tip detection is emphasized by a surrounding colored circle, trajectories are drawn as a thin line and the surrounded area can be highlighted by a semitransparent color (Fig. 8). Later, during the cutting process of the eardrum replacement on the mayo stand, the microscope inclination is registered to a fiducial marker. This marker is placed on the cutting board, so that the reconstructed shape can be augmented in a true-scale manner to the captured image of the cutting board (Fig. 1(b)) and displayed stereoscopically in the digital binocular or on additional displays. The so called ChArUco diamond marker<sup>25</sup> is highly robust to partial occlusion which is likely to occur in surgical environments. Accuracy evaluations for comparable markers targeting a medical context have been carried out before.<sup>26</sup>

## 4. EXPERIMENTAL RESULTS

So far our proposed system is mainly a proof-of-concept based on first experiments and simulations with a temporal bone model and a new prototypical instrument. No experiments have been done inside the human body. We want to show that our system is accurate enough to create a helpful 3D trajectory and that it is fast enough to be considered as a real-time solution for a digital microscope. Currently, the described processing chain is implemented as a plugin usable on the development version of the surgical microscope.

TARGET (mm)	ACTUAL (mm)		
	Mean	MeanAbsErr	StdDev
5	4.82	0.18	0.009
10	9.56	0.44	0.135
15	14.35	0.65	0.062
20	19.02	0.98	0.105
25	24.06	0.94	0.061
30	28.71	1.29	0.074
35	33.80	1.19	0.101
40	38.69	1.31	0.087

Table 1. Results of the accuracy evaluation in minimum zoom level. 50 measurements have been taken for every target distance.

TARGET (mm)	ACTUAL (mm)		
	Mean	MeanAbsErr	StdDev
1	0.97	0.04	0.034
2	1.92	0.08	0.020
3	2.94	0.05	0.014
4	3.92	0.08	0.058
5	4.92	0.08	0.015

Table 2. Results of the accuracy evaluation in maximum zoom level. 50 measurements have been taken for every target distance.

Initially, the accuracy of the close-range camera calibration and the obtained 3D reconstruction was evaluated by comparing image-based distance measurements on a depth-of-field test body captured in different orientations to ground truth values. Measurements for several target distances in minimum and maximum zoom positions have been carried out by selecting specific stereo correspondence points between the left and right image manually.

Table 1 and 2 show these results for multiple point-to-point measurement tasks. Comparison of target versus actual distances are evaluated by mean, mean absolute error and standard deviation values. 50 measurements have been recalled for each target (see Fig. 9). This results in a total of 400 measurements for minimum zoom level and 250 measurements for maximum zoom level. Achieved accuracies range from  $0.2mm$  to  $1.3mm$  in minimum zoom and from  $0.04mm$  to  $0.08mm$  in maximum zoom. It can be concluded, that the achieved accuracy is precise enough for the described ENT tasks. Furthermore, three main implications can be drawn from this evaluation: First, higher accuracies can be reached in maximum zoom level than in minimum zoom level. Second, higher accuracies can be reached for shorter distance measurements for both zoom levels. Third, the measurement range is limited to the microscope's depth-of-field.

Further experiments have been focused on the challenge to track the signalized instrument tip and reconstruct the path trajectory. A sequence of 300 stereo frames, in which the tip is outlining the complete outer temporal

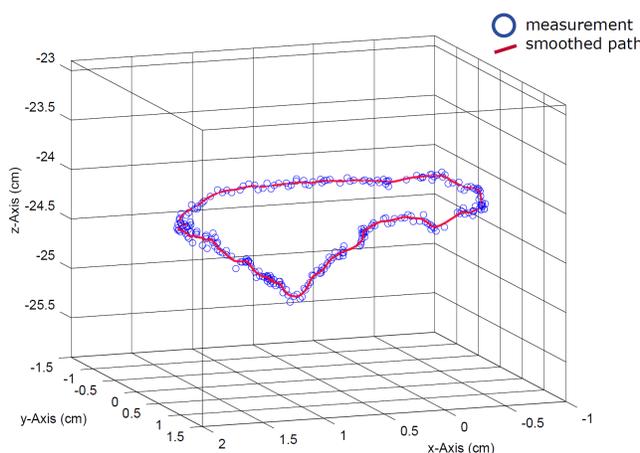


Figure 10. Reconstruction of the path in 3D space (red line). Blue dots are representing the raw measurements.

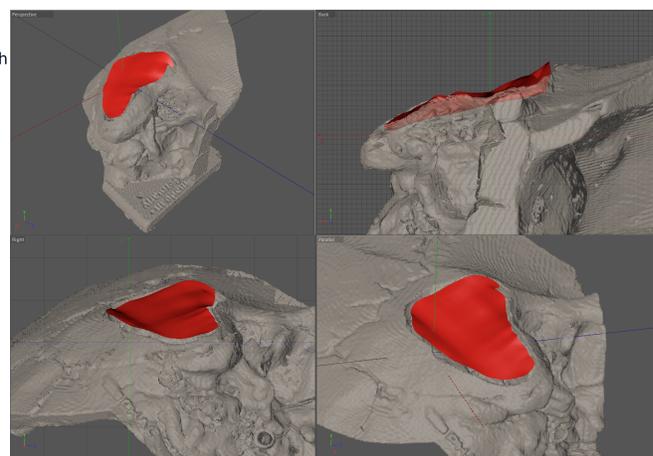


Figure 11. Reconstruction of the temporal bone model based on data from a CT scan registered with the surface generated from our 3D path trajectory.

bone shape, has been used for this test. Taking a closer look at the raw 3D measurements in Fig. 10, shaky contours caused by unsteady movements of the hand are clearly visible. The noise/jitter is removed by applying a Kalman filter before adding points to the path. This way, we obtain a smooth path accurately approximating the shape. This 3D path is stored and will be used as a stencil in the course of the surgery as shown in Fig. 1(b). The final result of the test is visualized by the red line in Fig. 10.

For further evaluation we took a computed tomography (CT) scan of the temporal bone model and use a three dimensional reconstruction generated from this scan as a ground truth. For the test shown in Fig. 11 a new sequence is generated where the bone contour is traced with the tip again. The instrument tip position can be estimated consistently for the complete sequence in the left and the right frame. A closed surface is tessellated from the path points to give a better optical impression of the shape. Then, we register the surface manually to the model created from the CT scan. The reconstructed surface fits the geometric structure of the contour in the model and its dimensions also match with the ground truth. In this case, for evaluation, not the position of the eardrum was traced, instead a more superficial contour, in which the precision can be demonstrated more comprehensible. Nevertheless, the test contour is comparable in terms of aspect ratio and shape of a real eardrum contour.

Using the current instrument tip with a radius of 0.5mm with the maximum magnification level proved to be difficult, since the relative size of the tip in the image is too large. This aggravates a precise positioning of the tip and increases the error of the stereo correspondence for the approximated circle center. Thus, further tests are planned where the measurement tip is having only half the diameter than used in this study. Since with more magnification the motion blur also increases, the parameters for circle fitting have to be tuned to react on sharp edges and by this blurred frames are filtered out to avoid inaccurate measurements.

The sequence shown in the previous section, consisting of about 1700 stereo images, is used to test the tracking performance. A continuous tracking and reconstruction is possible with more than 60 fps if the image resolution is  $960 \times 540$  and the template resolution is  $12 \times 12$  pixel. Even if we use a resolution of  $1920 \times 1080$  and templates of size  $24 \times 24$  pixel we still can process 29 fps. The evaluation system uses a *Nvidia GeForce GTX TITAN* graphics card and an *Intel(R) Xeon(R) Processor E5-2687W v3* with 3.10GHz.

## 5. CONCLUSION

We propose a new tool for digital surgical microscopes that enables distance measurements and 3D contour reconstruction by tracking a special marked instrument tip in ENT surgery without external tracking hardware. With this procedure the microscope can be extended to an interactive AR device allowing a direct interaction in the surgical area. Multiple measurement tasks have been evaluated with a ground truth test body to show that accurate measurements in microscope images are possible by using stereo triangulation. An optically tracked 3D path trajectory of the temporal bone model has been successfully validated with data from a CT scan. Gained information, like path trajectories and measured distances, are visualized as an additional image layer directly inside the binocular. The tool runs as a real-time plugin on the digital microscope. Its usefulness is not restricted to tympanoplasty and a transfer to other medical domains where surgical microscopes are used seems to be feasible.

In future work, we will validate our approach with clinical image data from ENT surgery to see how well it can cope with challenges as blood on the instrument tip, as well as reflections and de-focused images. Up to now, we assume that the measurement will take place during a controlled period in the course of the operation where these interfering factors can be minimized. In addition, we plan to use a smaller measurement tip to improve the performance with a large magnification. Also an integration of deep learning based detection strategies is imaginable to replace the template matching and to make the tracking more robust and flexible. Besides improving our algorithm, we want to compare the benefits of our method with the current workflow of a tympanoplasty in terms of efficiency and usability.

## ACKNOWLEDGMENTS

This work is supported by the German Federal Ministry for Economic Affairs and Energy (3DInMed project, grant no. 01MT15001A).

## REFERENCES

- [1] “Fallpauschalenbezogene Krankenhausstatistik (DRG-Statistik) Diagnosen, Prozeduren, Fallpauschalen und Case Mix der vollstationären Patientinnen und Patienten in Krankenhäusern.” Statistisches Bundesamt, 08 October 2018 [https://www.destatis.de/DE/Publikationen/Thematisch/Gesundheit/Krankenhaeuser/OperationenProzeduren5231401177014.pdf?\\_\\_blob=publicationFile](https://www.destatis.de/DE/Publikationen/Thematisch/Gesundheit/Krankenhaeuser/OperationenProzeduren5231401177014.pdf?__blob=publicationFile). (Accessed: 01 November 2018).
- [2] Seta, E. D., Covelli, E., Seta, D. D., and Filipo, R., “Cartilage tympanoplasty: How to reduce surgery time,” *The Journal of Laryngology & Otology* **124**(7), 784–785 (2010).
- [3] Eren, S. B., Tugrul, S., Ozucer, B., Dogan, R., and Ozturan, O., “A randomized prospective trial of a novel device for measuring perforation size during inlay butterfly myringoplasty,” *American Journal of Otolaryngology* **35**(3), 305–308 (2014).
- [4] Solís, S. M., del S. Hernández-Montes, M., and Santoyo, F. M., “Tympanic membrane contour measurement with two source positions in digital holographic interferometry,” *Biomed. Opt. Express* **3**(12), 3203–3210 (2012).
- [5] ARRI Medical, “ARRISCOPE See the difference.” 2018 [http://www.arrimedical.com/assets/pdf/Brochure\\_ARRISCOPE\\_EV02\\_English.pdf](http://www.arrimedical.com/assets/pdf/Brochure_ARRISCOPE_EV02_English.pdf). (Accessed: 02 November 2018).
- [6] Bichlmeier, C., “Medical products as a platform for augmented reality.” Medical Augmented Reality Blog, December 2015 <http://www.medicalaugmentedreality.com/2015/12/medical-products-as-a-platform-for-augmented-reality>. (Accessed: 10 October 2017).
- [7] Bouget, D., Allan, M., Stoyanov, D., and Jannin, P., “Vision-based and marker-less surgical tool detection and tracking: a review of the literature,” *Medical Image Analysis* **35**, 633–654 (2017).
- [8] Allan, M., Ourselin, S., Thompson, S., Hawkes, D. J., Kelly, J., and Stoyanov, D., “Toward detection and localization of instruments in minimally invasive surgery,” *IEEE Transactions on Biomedical Engineering* **60**(4), 1050–1058 (2013).
- [9] Bouget, D., Benenson, R., Omran, M., Riffaud, L., Schiele, B., and Jannin, P., “Detecting surgical tools by modelling local appearance and global shape,” *IEEE Transactions on Medical Imaging* **34**(12), 2603–2617 (2015).
- [10] Du, X., Allan, M., Dore, A., Ourselin, S., Hawkes, D. J., Kelly, J., and Stoyanov, D., “Combined 2d and 3d tracking of surgical instruments for minimally invasive and robotic-assisted surgery,” in [*International Conference on Information Processing in Computer-Assisted Interventions*], (2016).
- [11] Speidel, S., E. Kuhn, S. B., S. Röhl, H. K., Müller-Stich, B., and Dillmann, R., “Visual tracking of da vinci instruments for laparoscopic surgery,” in [*Proc. SPIE 9036, Medical Imaging 2014: Image-Guided Procedures, Robotic Interventions, and Modeling*], (March 2014).
- [12] Sznitman, R., Ali, K., Richa, R., Taylor, R. H., Hager, G. D., and Fua, P., “Data-driven visual tracking in retinal microsurgery,” in [*MICCAI (2)*], *Lecture Notes in Computer Science* **7511**, 568–575, Springer (2012).
- [13] Rieke, N., Tan, D. J., di San Filippo, C. A., Tombari, F., Alsheakhali, M., Belagiannis, V., Eslami, A., and Navab, N., “Real-time localization of articulated surgical instruments in retinal microsurgery,” *Medical Image Analysis* **34**, 82–100 (2016).
- [14] Bodenstedt, S., Wagner, M., Mayer, B., Stemmer, K., Kenngott, H., Müller-Stich, B. P., Dillmann, R., and Speidel, S., “Image-based laparoscopic bowel measurement,” *Int. J. Computer Assisted Radiology and Surgery* **11**(3), 407–419 (2016).
- [15] Rosenthal, J.-C., Gard, N., Schneider, A., and Eisert, P., “Microscopic image-based determination of stapes prosthesis length,” in [*CARS 2018-Computer Assisted Radiology and Surgery Proceedings of the 32st International Congress and Exhibition*], 59–60 (2018).
- [16] Karygianni, S., Martinello, M., Spinoulas, L., Frossard, P., and Tomic, I., “Automated eardrum registration from light-field data,” in [*ICIP*], 1243–1247, IEEE (2018).
- [17] Bedard, N., Shope, T., Hoberman, A., Haralam, M. A., Shaikh, N., Kovačević, J., Balram, N., and Tošić, I., “Light field otoscope design for 3d in vivo imaging of the middle ear,” *Biomed. Opt. Express* **8**, 260–272 (Jan 2017).

- [18] Aschke, M., Wirtz, C. R., Raczkowski, J., Wörn, H., and Kunze, S., “Augmented reality in operating microscopes for neurosurgical interventions,” *Proceedings 1st International IEEE EMBS Conference on Neural Engineering*, 652–655 (2003).
- [19] Fischer, J., Neff, M., Freudenstein, D., and Bartz, D., “Medical augmented reality based on commercial image guided surgery,” in [*Proceedings Eurographics Workshop on Virtual Environments*], The Eurographics Association (2004).
- [20] Huang, Y.-H., Yu, T.-C., Tsai, P.-H., Wang, Y.-X., Yang, W.-l., and Ouhyoung, M., “Scope+: A stereoscopic video see-through augmented reality microscope,” in [*Adjunct Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*], 33–34 (2015).
- [21] Eisert, P., “Model-based camera calibration using analysis by synthesis techniques,” in [*International Workshop on Vision, Modeling, and Visualization*], 307–314 (Nov. 2002).
- [22] Matthews, I., Ishikawa, T., and Baker, S., “The template update problem,” *IEEE Transactions on Pattern Analysis and Machine Intelligence* **26**, 810–815 (June 2004).
- [23] Haralick, R. M., “Digital step edges from zero crossing of second directional derivatives,” *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**(1), 58–68 (1984).
- [24] Vacchetti, L., Lepetit, V., and Fua, P., “Combining edge and texture information for real-time accurate 3d camera tracking,” in [*Proceedings of the 3rd IEEE/ACM International Symposium on Mixed and Augmented Reality, ISMAR '04*], 48–57 (2004).
- [25] Garrido-Jurado, S., Muñoz-Salinas, R., Madrid-Cuevas, F., and Marín-Jiménez, M., “Automatic generation and detection of highly reliable fiducial markers under occlusion,” *Pattern Recognition* **47**(6), 2280 – 2292 (2014).
- [26] Belhaoua, A., Kornmann, A., and Radoux, J.-P., “Accuracy analysis of an augmented reality system,” in [*Signal Processing (ICSP), 2014 12th International Conference on*], 1169–1174, IEEE (2014).