Interactive and Multimodal-based Augmented Reality for Remote Assistance using a Digital Surgical Microscope

Eric L. Wisotzky *
Fraunhofer HHI & HU Berlin

Jean-Claude Rosenthal†
Fraunhofer HHI

Peter Eisert
Fraunhofer HHI & HU Berlin

Anna Hilsmann
Fraunhofer HHI

Falko Schmid
Ubimax GmbH

Michael Bauer
pripares GmbH

Armin Schneider
ARRI Medical GmbH

Florian C. Uecker
Charité Universitätsmedizin Berlin

Abstract

We present an interactive and multimodal-based augmented reality system for computer-assisted surgery in the context of ear, nose and throat (ENT) treatment. The proposed processing pipeline uses fully digital stereoscopic imaging devices, which support multispectral and white light imaging to generate high resolution image data, and consists of five modules. Input/output data handling, a hybrid multimodal image analysis and a bi-directional interactive augmented reality (AR) and mixed reality (MR) interface for local and remote surgical assistance are of high relevance for the complete framework. The hybrid multimodal 3D scene analysis module uses different wavelengths to classify tissue structures and combines this spectral data with metric 3D information. Additionally, we propose a zoom-independent intraoperative tool for virtual ossicular prosthesis insertion (e.g. stapedectomy) guaranteeing very high metric accuracy in sub-millimeter range (1/10 mm). A bi-directional interactive AR/MR communication module guarantees low latency, while consisting surgical information and avoiding informational overload. Display agnostic AR/MR visualization can show our analyzed data synchronized inside the digital binocular, the 3D display or any connected head-mounted-display (HMD). In addition, the analyzed data can be enriched with annotations by involving external clinical experts using AR/MR and furthermore an accurate registration of preoperative data. The benefits of such a collaborative surgical system are manifold and will lead to a highly improved patient outcome through an easier tissue classification and reduced surgery risk.

Index Terms: H.5.1 [Information Interfaces And Presentation]: Multimedia Information Systems—Artificial, augmented, and virtual realities; I.2.1 [Artificial Intelligence]: Applications and Expert Systems—Medicine and science; I.2.10 [Artificial Intelligence]: Vision and Scene Understanding—3D/stereo scene analysis; I.4.8 [Image Processing And Computer Vision]: Scene Analysis—Stereo; J.3 [Computer Applications]: Life And Medical Sciences—Medical information systems;

*e-mail:eric.wisotzky@hhi.fraunhofer.de
†e-mail:jean-claude.rosenthal@hhi.fraunhofer.de
Both authors contributed equally to this work as first author

1 Introduction

During surgery, a surgeon differentiates between healthy tissue structures, which have to be maintained, and abnormal or damaged tissue, which has to be removed, replaced or reconnected. This continuous differentiation is based on his experience and knowledge only and entails great risk because injuring important structures, as nerves, can cause permanent damage to the patient's health. Nowadays, optical devices, like magnifying glasses, surgical microscopes and endoscopes, are used to support the surgeon in more than every second surgery [17]. In some medical fields the number increases up to 80% usage [24], as a three dimensional optical magnification of the operating field allows more complicated and complex surgeries. The general working principle of surgical microscopes has been consistent during the last decades, distinguishing between the Greenough principle using two separate objectives and the Common Main Objective (CMO) / telescope principle using one shared main objective (Fig. 1). Such microscopic systems have always been using analogous optical imaging systems and are well understood.

Nonetheless, a simple analog and purely optical magnification does not give information about the accurate scale of the tissue structures and tissue characteristics, as such systems show several drawbacks as soon as modern computer vision algorithms or medical augmented reality (AR)/ mixed reality (MR) applications shall be applied. First, a beam splitter is obligatory to digitize the analog input signal, resulting in lower image quality in terms of contrast and resolution. Additionally, the captured perspective differs from the surgeon's field-of-view. Further, system calibration and preoperative data registration is complicated and suffers from low spatial accuracy. Besides these limiting imaging fac-

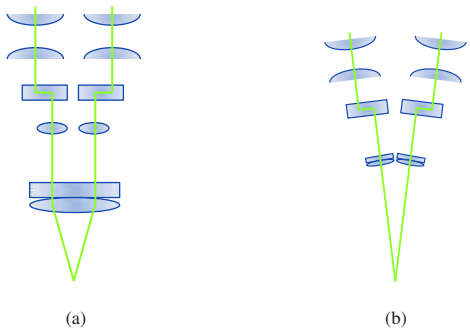


Figure 1: Microscope optical path - simplified visualization: (a) CMO principle (b) Greenough principle

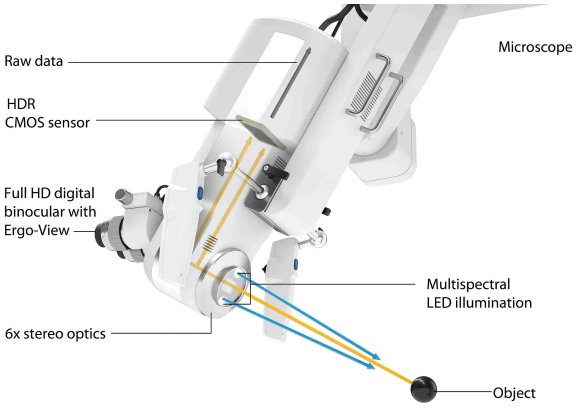


Figure 2: Digital binocular and high resolution imaging unit of microscopic head.

tors, current medical AR systems rely on external tracking hardware, e.g. electro-magnetic tracking (EMT) or optical tracking systems based on infrared light using fiducial markers. These systems hold further challenges, as EMT can suffer from signal interference and optical tracking systems need a line-of-sight to work properly [40, 19, 7]. The configuration of such a system is time-consuming, complicated, error-prone and can easily interrupt the ongoing surgical procedure.

As the possibilities resulting from digitization are of increasing importance, digital microscopes and endoscopes are increasingly being used. Such fully digital devices consist of a complete digital processing chain enabling new forms of integrated image processing algorithms, intraoperative assistance and surgical aware AR/MR information visualization. The display technology is chosen depending on the intended surgical use. In the presented application, we use the digital binocular as the primary display for visualization, nonetheless augmented data can be distributed to any external 2D/3D display or remote VR/AR visualization unit e.g. AR glasses, HMD. Thus, consulting external experts using AR/MR communication during surgery becomes more feasible. A detailed overview for endoscopic and microscopic AR applications can be found in [25, 23, 9, 20].

In this interdisciplinary project, we outline an entirely AR/MR image-based multimodal processing chain using white light and multispectral imaging in the context of ear, nose and throat (ENT) surgery. This includes stereoscopic 3D reconstruction of the surgical area to measure the correct scale as well as multispectral scene analysis to capture optical tissue behavior not visible under white light illumination in the RGB-space. In detail, 3D point clouds of specific anatomical structures at varying time points are generated. 3D measurements and the documentation of tumor dimensions or any other tissue-of-interest (TOI) allow true-scale comparison to preoperative data, while multispectral tissue analysis allows accurate tissue classification. These metric and spectral measurements as well as anatomical 3D representations can be used for intraoperative assistance by augmenting anatomical structures with enriched surgical data. In addition, such surgical data sets are valuable for the documentation of interventions and for the creation of annotated videos for surgical training [28, 16].

The advanced multimodal imaging and analysis unit is extended by an interactive collaboration interface offering bi-directional communication to support local and remote surgical assistance for collaboration between internal and remote experts. Additionally, the proposed pipeline is source agnostic as it is applicable to 3D endoscopy, 3D microscopy and other stereoscopy imaging devices.

2 Exposition

Our proposed system consists of five major modules. Fig. 3 depicts the complete system architecture. Two in-/output modules (I/O) handle the multimodal image input (Fig. 3 yellow parts, see Sec. 2.1) and output (Fig. 3 green parts, see Sec. 2.5) and its AR/MR visualization in the context of remote surgical assistance. The third module implements a bi-directional communication interface (Fig. 3 red parts) which connects the I/O modules (see Sec. 2.4).

The two remaining modules are part of a multimodal-based image analysis interface (Fig. 3 purple parts): 3D reconstruction (zoom-independent stereo calibration, image-based measurement, MR-assisted stapedectomy, see Sec. 2.2) and tissue classification using multispectral imaging (see Sec. 2.3), feeding enriched data into visualization unit within the bi-directional communication module. In the following subsection the specific challenges of all modules are outlined.

2.1 Imaging Input

2.1.1 Microscopic Input

The microscope used in this work[1] (Fig. 2) is the first digital surgical microscope which has received medical approval for ENT surgery in the European Union (EU) [1]. The microscopic specifications are presented in Tab. 1. The images are captured by a high resolution digital camera sensor ($2 \times 1920 \times 1080$) with 60 fps. The optical system is motor controlled allowing an accurate and precise positioning of specific focal lengths and focus positions which is crucial for a high accuracy zoom-independent calibration, see Sec. 2.2.1. Captured images are displayed inside a digital binocular using a high-resolution OLED display. As a consequence, these images can be analyzed before being shown to the surgeon without any additional temporal latency or spatial registration errors. Therefore, analyzed data, like classified tissue regions or resulting image-based 3D measurements, can be augmented with high spatial accuracy directly into the surgeon's field of view (FoV). The digital binocular allows layered overlays including preoperative data (e.g. CT or MRT images), vital signs of the patient and augmentations of the scene analysis (Sec. 2.2 and Sec. 2.3) or of an external expert into the surgeon's view. This results in a direct intraoperative surgical assistance without interfering with existing surgical workflows and resulting in lower surgical risk as well as reduced procedure times [31]. In contrast to the complex system setup of existing AR solutions, our proposed system only requires a one-time calibration, configuration of additional tracking hardware for surgical navigation is not necessary.

The camera sensor consists of three channels (red - R, green - G and blue - B). Fig. 5 shows the sensitivity of all three channels. To achieve a balanced white light illumination, the illumination unit

[1] ARRISCOPE, ARRI Medical GmbH, Munich, Germany

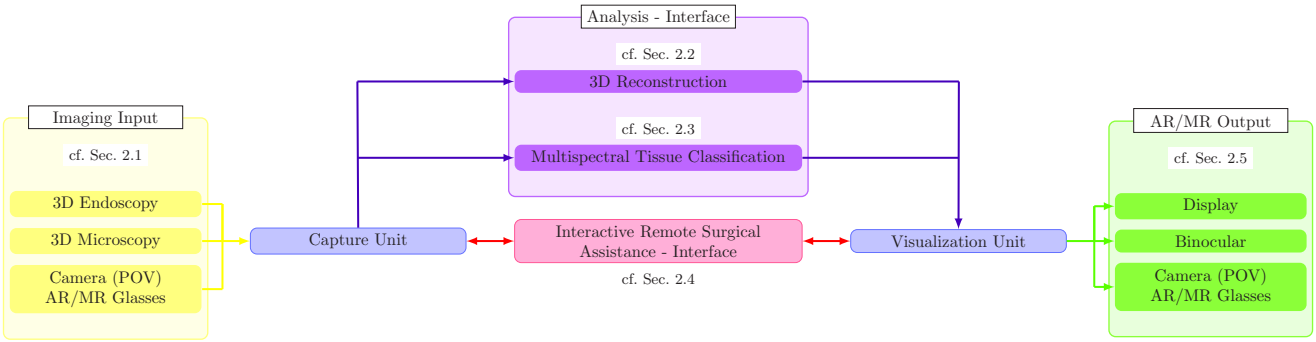


Figure 3: Pipeline for a multimodal AR/MR framework using multispectral and white light imaging.

Table 1: Microscope specifications

Sensor specifications		
CMOS sensor		3392×2200 px
Output resolution	$2\times$ Full HD	$2 \times 1920 \times 1080$ px
Mechanical dimension		23.76×13.365 mm^2
Diagonal dimension		29.26 mm (1.7")
Frame rate		60 fps
Camera optics specifications		
Beam angle	Wide-angle	16.52°
	Zoom	6.26°
Focal length	Mean	58.237 mm
	Wide-angle	64.523 mm
	Zoom	65.220 mm
Field of View (FoV)	max.	27 mm x 48 mm
at 210 mm WD	min.	13 mm x 23 mm
Magnification		$1.6\times$ up to $9.6\times$

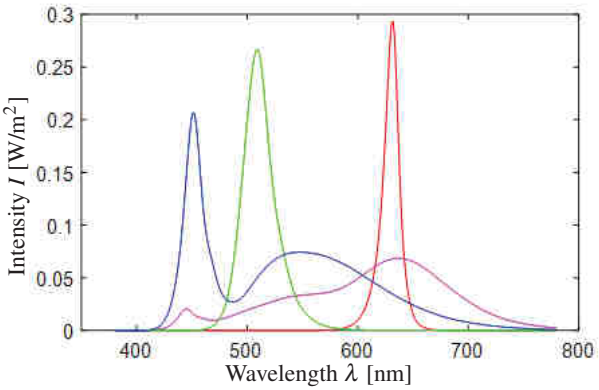


Figure 4: This plot shows the energy spectra of the four channel LED illumination unit of the microscope, used in this project.

of the microscope uses the spectral combination of four different light-emitting diodes (LED), a red LED, a warm white (yellowish) LED, a green LED and a cold white LED. Each LED can be adjusted independently in terms of intensity to mix the emitted light for a uniform white impression. The energy spectra of the four LEDs are presented in Fig. 4. Due to the fact, that the illumination unit uses four independently controlled LEDs, multispectral imaging becomes possible, see Sec. 2.3. Using different illumination spectra together with the responses of the three different spectral RGB channels of the sensor allows spectral reconstruction of the captured object.

2.1.2 Endoscopic Input

In contrast to digital microscopes, 3D endoscopes have to deal with many limitation and restrictions as they are used for non open procedures, namely minimally-invasive surgery (MIS). The dimension of a 3D endoscopy tube is limited by the anatomy of the related body cavities, like the nostril and auditory canal for ENT or the size of the trocar for abdominal surgery. Therefore, optical parts need to be heavily miniaturized. This influences directly the sensor size resulting in noisy and low contrast images. In addition, endoscopic images suffer from difficult light conditions, soiled lens (blood, water, smoke) and motion blur.

Tab. 2 contains relevant selected technical values from different manufacturers for 3D laparoscopes currently used in MIS and rele-

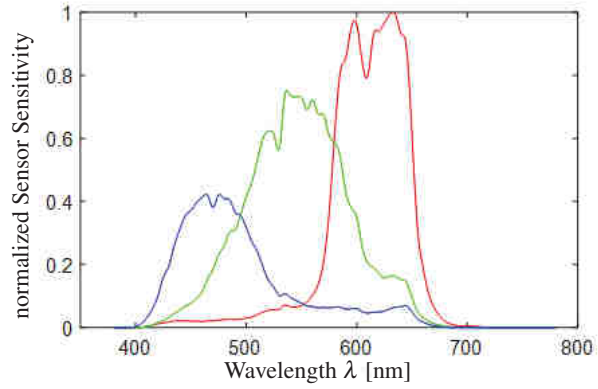


Figure 5: This plot shows the sensitivity of the CMOS sensor of the RGB channels.

Table 2: Selected 3D endoscope specifications

Output resolutions	interlaced / progressive	up to 1920×1080 px
Frame rate		25 or 30 fps
Interaxial distance		approx. 3.5 - 4.5 mm
Endoscope diameter		approx. 8.0 - 9.5 mm

vant for this project. Within this project a 3D laparoscope[2] is used for preliminary endoscopic analysis.

2.2 3D Reconstruction

For 3D reconstruction and surgical AR/MR applications a calibration of the stereoscopic system is crucial, to get a real world representation of the surgical scene and the underlying optical system. Therefore, a zoom-independent calibration scheme is defined, to allow image-based measurements resulting in metric and 3D data representation in Sec. 2.2.1. This 3D data serves as a basis for high-level surgical AR/MR assistance functions as described for stapedectomy in Sec. 2.2.2 and for tissue classification in Sec. 2.3.

2.2.1 Zoom-independent Calibration & 3D Measurement

The calibration of optical 3D systems for photogrammetric applications and stereo image processing algorithms are highly coupled. Both topics will be discussed in the subsequent paragraphs. The first paragraph will focus on the principle consideration when it comes to the calibration of very long focal lengths. The second paragraph outlines a real-time stereo processing pipeline for image-based 3D measurements which extends the calibration topic by outlining the benefit of a quasi auto calibration for stereoscopic systems using zoom lenses.

Zoom-independent Calibration

Camera calibration techniques using checkerboards have been intensively discussed for many applications including medical scenarios in [36, 43, 2, 12, 29] and are widely accepted within the computer vision community. Nonetheless, these methods do not address the problem of calibrating optical imaging systems which have varying focus and zoom settings. Especially, the calibration of very long focal lengths is still a difficult problem. The works of Stamatopoulos et al. [34, 33] are the most recent works which actually investigate zoom lenses for photogrammetry tasks and give valuable insights when it comes to the calibration of long focal lengths. His observations are based on a single DSLR camera, which uses a macro lens of $105mm$. This is comparable to some extent to our lens specification. In our case the maximum zoom level has a focal length of $348mm$. The usage of different calibration patterns for different zoom settings is quite cumbersome because of a small field-of-view and a decreasing depth-of-field for high magnification factors. For these practical reasons, we use a hybrid calibration target, which can be used for any zoom/focus combination. Fig. 6 shows this hybrid target captured with varying zoom levels, guaranteeing crisp and high-resolution checkerboard features. The calibration method follows a model-based approach using synthesized images and applying an image registration via gradient-descent as described in [6]. The combination of the hybrid synthesized image registration and the motorized control unit, to adjust specific zoom/focus settings, facilitates significantly the crucial calibration process. Fig. 7 shows a possible calibration strategy consisting of 21 defined calibration positions.

For every calibration point 10 images are captured, resulting in 210 calibration images. This suggests to be a good trade-off between the amount of calibration data acquisition and expected calibration quality. Based on these calibration images, we derive independently the intrinsic and extrinsic parameters of the microscopic system for each lens left and right, namely: the focal length f_x, f_y, three radial distortion parameters r^2, r^4, r^6 and the principal point C_x, C_y. Afterwards, we calibrate the stereo system by calculating the orientation of both lenses to each other based on their extrinsic pose in relation to the calibration target, generating a rotation matrix R_{3x3} and a translation vector $\vec{t}_3$. Hence, we will derive 21 projection matrices P for known zoom/focus positions. For intermediate zoom/focus positions, we interpolate our optical parameters using a look-up-table holding calibrated system values and the current motor position.

The overall evaluation of different zoom/focus settings and the automation of the presented calibration strategy will be performed, as the overall goal is to develop a calibration toolchain which can be easily integrated and maintained into a demanding environment like the operating room.

3D Measurement

A successful calibrated system enables us to perform image-based measurements in any zoom level with varying focus settings if we can determine reliable sub-pixel correspondence of left points $p_l(x,y)$ and right points $p_r(x,y)$ in stereoscopic views, namely the left image I_L and the right image I_R, respectively. Therefore, a real-time capable stereo image processing pipeline is applied [30], which has been successfully used for simple point-to-point distance measurements inside the tympanic cavity. This basic measurement tool consists of two steps: First, it applies a quasi auto calibration by a scene dependent rectification of image pairs via detection of robust feature points (e.g. comparable to BRIEF features [4]). During this auto calibration the authors constantly calculate a $3x3$ homography matrix $\widetilde{H}$ from the feature point correspondences. In this way, zoom setting changes can be detected when the homography changes noticeably. This is of high importance in our scenario as this feature functions as an online stereoscopic geometry quality control tool. However, if a zoom change occurs, the following stereoscopic misalignments caused by left/right zoom offset, vertical, tilt, keystone or rotation errors are corrected. Hence, the authors obtain a rectified image pair $\widehat{P}\,(\widehat{I}_L, \widehat{I}_R)$ by warping only I_R and keeping I_L unchanged as follows:

$$\widehat{I}_L = I_L * \widetilde{I}_{3x3} \quad \text{with } \widetilde{I} = \text{identity matrix} \tag{1}$$

$$\widehat{I}_R = I_R * \widetilde{H}_{3x3} \tag{2}$$

Second, a highly parallelized sub-pixel dense disparity estimation (comparable to SGBM [15]) on a GPU is performed. The disparity estimator is feed with $\widehat{P}\,(\widehat{I}_L, \widehat{I}_R)$ as input. Now, the correspondence problem can be reduced to a one dimensional search resulting in a fast and robust stereo estimation of correspondences. Within the pipeline, this measurement tool serves as a core processing unit for additional measurement modules, which directly relates to the outlined high-level surgical measurement task, cf. Sec. 1.

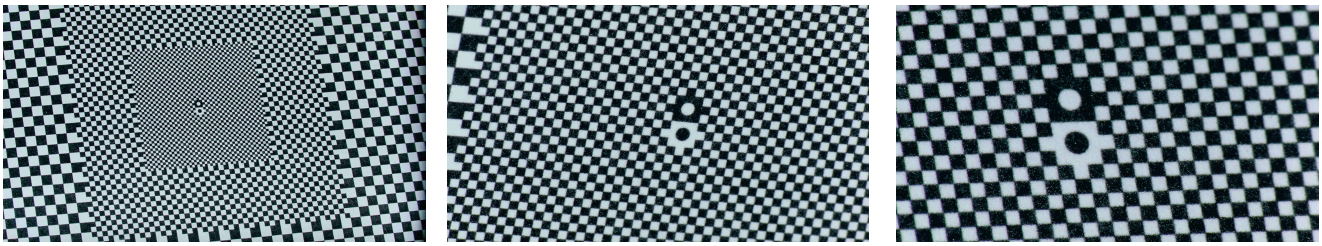
As next steps, a registration of CT/MRI data to the live view as well as the tracking of anatomical features or landmarks related to high-level surgical tasks will be build upon the existing stereo core processing pipeline.

2.2.2 MR-assisted Stapedectomy

The aim of a stapedectomy treatment is to recover lost hearing ability caused by otosclerosis. Therefore, damaged parts of the stapes within the ossicular chain are replaced with a micro prosthesis. For reference, Fig. 9a shows the complete middle ear anatomy also indicating the stapes location and Fig. 9b shows a snapshot of an actual procedure of a prosthesis insertion.

During the intervention, the surgeon needs to choose a patient specific appropriate-sized prosthesis which ranges from 2.5 mm to

[2] SCHÖLLY FIBEROPTIC GMBH, Denzlingen, Germany



(a) (b) (c)

Figure 6: Hybrid calibration pattern: (a) Minimum zoom level (b) Mid zoom level (c) Maximum zoom level

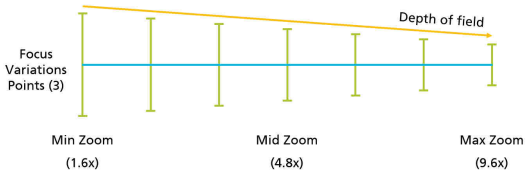


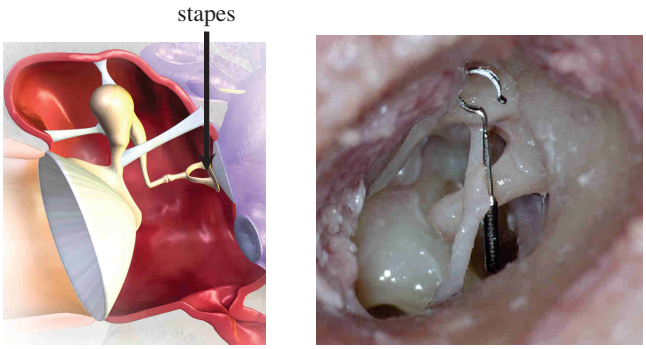
Figure 7: Overview of defined calibration points ranging from minimum zoom to maximum zoom consisting of 7 zoom positions whereas every zoom position considers 3 steps to model focus variations and decreased depth-of-field effect.

Figure 8: Dual view of a 3D model stapes prosthesis which will be used for virtual augmentation to derive the correct prosthesis size.

10.0 mm with a given stepsize of 0.25 mm. An accurate measurement of the incus-footplate distance can improve the hearing outcome significantly as reported in [31]. But current surgical measurement solutions are limited in the following ways: Instruments are inaccurate (stepsize > 0.5 mm) and unhandy, plastic dummy prosthesis are deformable and in general the surgeon runs always the risk of unwanted and harmful tissue contact. Therefore, our proposed idea for an MR-stapedectomy procedure will focus on the true-scale insertion of a virtual 3D prosthesis model into the situs to simulate the replacement of the damaged stapes within the ossicular chain. The interaction between the real world and the positioning of the virtual prosthesis will be implemented using two joysticks integrated in the microscopic handle. Fig. 8 shows a 3D model of such a prosthesis.

Based on this true-scale prosthesis representation the surgeon will be able to choose the best possible prosthesis size without touching any tissue before the final insertion. From this approach two main benefits arise: (1) the risk of complications caused by harmful tissue contact is reduced and (2) a patient-tailored stapes prosthesis can be inserted to recover his hearing capability in the best possible way.

Future work will evaluate the true scale representation of the virtual prosthesis taking into account the calibration results in terms of accuracy and precision. In addition, a seamless integrated joystick interface will allow the MR interaction the between the real



(a) (b)

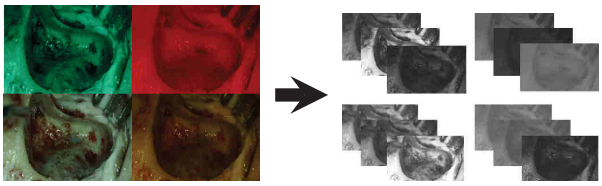
Figure 9: (a) Drawing of middle ear anatomy and the ossicular chain. Source: Blausen.com staff (2014). "Medical gallery of Blausen Medical 2014". WikiJournal of Medicine 1 (2). DOI:10.15347/wjm/2014.010. ISSN 2002-4436. (b) Snapshot of a stapedectomy intervention showing a prosthesis near the ossicular chain.

world and the virtual prosthesis. It is also noteworthy, that the same technique can be applied for other types of ear prosthesis such as partial/total ossicular chain replacement prostheses (PORP/TORP) and can be useful for drilling the cavity for cochlear implant processors.

2.3 Multispectral Tissue Analysis

Multispectral imaging (MSI) is used in biomedicine and food science for cell segmentation, skin and meat analysis [21, 26, 18]. For intraoperative medical therapy, few authors have shown the feasibility [45, 44, 42, 41], but MSI is not established [21]. In principle, different techniques exists to acquire the multispectral data of a scene: hyperspectral cameras can be used with linescan or snapshot mode to acquire the dataset in a short amount of time but with small resolution [13, 22, 41]. Alternatively, a filter-wheel in front of the sensor or after the light source can step through the hyperspectral space [11, 3]. However, this technique is very time-consuming in terms of acquisition and image postprocessing [42]. Therefore, these techniques are not eligible for the targeted application due to, e.g. no real-time capability and expensive postprocessing steps.

This project follows the idea to extract multispectral tissue data combining wide-band LEDs as illumination unit and a RGB-camera as three-channel acquisition device. The image scene will be illuminated using all available LEDs consecutively and acquire images synchronized to the illumination sequence. This illumination modality covers the visual spectrum (approx. 400 nm to 780 nm).



(a) (b)

Figure 10: Concept of a multispectral imaging. (a) The four RGB camera responses acquired under different light conditions. (b) The twelve extracted spectral bands from the multispectral imaging.

2.3.1 Spectral Calibration

To make the sensor data usable for multispectral analysis, spectral calibration is required. As the LED illumination changes rapidly and only takes 200 ms for a complete sequence, it is supposed that all RGB responses include exactly the same view.

The denoising and conversion of raw image data I_{raw} to reflectance I_{res} is achieved using

$$I_{res} = \frac{I_{raw} - I_{dark}}{I_{white} - I_{dark}}, \tag{3}$$

where I_{raw} is the measured pixel information and I_{dark} contains the information of a dark reference image. This includes white balance alignment and dark current correction as well. This method requires a white reference calibration board to be scanned in all illumination settings to acquire I_{white} for white balance alignment. This step is a one point calibration method, where only one reference target is measured [27].

Further, the distances δ_1 between illumination source and investigated surface as well as δ_2 between surface and camera sensor have an impact on the sensor response [41]. Due to the system design, both distances show always the same ratio, i.e. a microscope and an endoscope hold the sensor and illumination source in the same mounting head. Therefore, one proportionality constant α can describe the relation between irradiation intensity and sensor signal. Remains the distance between mounting head and object constant, a one-time calibration would be sufficient. However, in clinical routine, α changes several times during a single surgery. Therefore, an adaptive α adjustment is needed. This will be solved using the described 3D measurements, since it can determine the object to camera distances δ, c.f. Sec. 2.2.1.

2.3.2 Spectral Reconstruction

The reconstruction of in-situ reflectivity functions (high-dimensionally) from a low-dimensional digital camera response is a challenging requirement, as it is generally an ill-posed problem [38, 39]. In literature, various spectral reconstruction methods are presented to manage this problem. A common strategy of overcome the ill-posed character of reflectance estimation is to utilize additional information of the underlying acquisition process to obtain reasonable results. One of the major and mostly used reconstruction methods is the Wiener filter [14, 32, 35, 38, 39]. The main problem using the Wiener filter is, usually more than three camera responses (R, G and B) are needed for an accurate spectral reconstruction [5]. This problem can be reduced if several images of the same view using different illumination settings (see Fig. 4) are acquired. Thus, the dimension of the input data is increased to allow a more accurate reconstruction [37].

The usage of different illumination settings I_1, I_2, I_3, I_4 of the same scene (see Fig. 10a) will allow to calculate twelve spectral bands (see Fig. 10b). The individual optical tissue behaviors are extracted using these twelve bands.

2.3.3 Classification and Visualization

The different reflection spectra of the different tissue types are not visually recognized, as the differences of the reflections are marginal if the complete visual range is considered. Therefore, the challenge is to classify different reconstructed spectra (many inputs over the complete visual range) into several tissue categories (few outputs). For large data and the absence of classification rules, as in our case, machine learning is a potential method for modeling the data. Therefore, one aim of the project is to build up a tissue database including the optical behavior of several tissue types and a large patient cohort. This will facilitate the classification problem.

Further, it will be ~~is~~ investigated, whether it would not be simpler and more reasonable to drop classification while simply adapt AR/MR visualization. As, one obstacle will be AR/MR visualization of the high dimensional tissue data, it seems possible to modify the color space representation in a way that the small reflection differences between the tissue types emerge prominently. Using the high dimensional spectral data, a normal RGB-image with high color-fidelity has to be extracted and by modifying the color representation, additional tissue information can highlighted without loss of other important image information. In this approach, the surgeon arrives on the classification on his own, which will simplify ethical and legal impacts.

2.4 Interactive Remote Surgical Assistance

For remote surgical assistance, it needs to be assured that all surgeons can rely and work on consistent data. Furthermore, the entire pipeline needs to run in real-time without any noticeable latency. Meaning, the low-latency communication software needs to be able to handle interactive scribblings, large 3D data sets, preoperative data and vital parameters of the patient fulfilling these constraints.

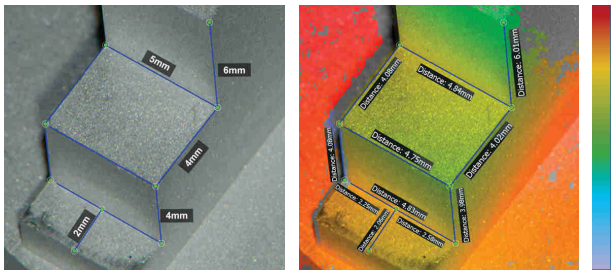
Therefore, a technology currently used for a major European football league will be adapted and extended for the surgical use case. So far, this system transmits de-centralized live data of game statistics (e.g. game score, game highlights, football player statistics, etc.) into a centralized information news ticker in real-time. The communication interface targets three network architectures: (1) a wired local are network (LAN) inside the hospital, (2) a wireless local are network (WiFi) in the operating room and (3) a wide area network (WAN) which offers prioritized and secured telematic infrastructure for physicians in that European country.

For any of these network scenarios, we address a low-latency interactive bi-directional communication user interface, which allows medicolegal consultations between surgical experts or detailed patient counseling before the intervention.

2.5 AR/MR Output & Visualization

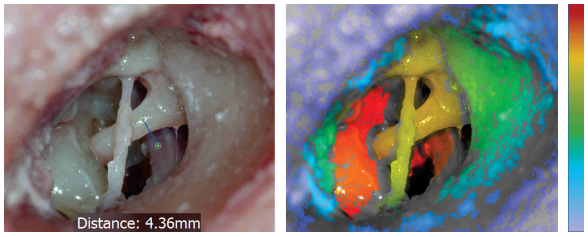
For the output module in Fig. 3, it has to be investigated to what extent it is possible to calculate an accurate high-resolution image by use of correlations between position, time and spectral range on all visualization devices. Nonetheless, the format agnostic imaging pipeline allows us an adaptive presentation of the underlying mutual information. This might be an external display showing overall information related to the ongoing procedure, the binocular and a remote HMD allowing interaction with specific virtual objects for communication between surgeons or video streaming including audio comments and annotation into a lecture room for trainee surgeons or medical students.

Additionally, the expansion of classical image analysis methods like image registration, tracking of multispectral imaging data, and inclusion of preoperative images results in a high degree of robustness and stability. Nonetheless, current results for image-based measurements and the extraction of high-level features using for intraoperative assistance look promising. First AR/MR visualization ideas from 3D reconstruction results are shown in Fig. 11 and



(a) (b) (c)

Figure 11: (a) Results of image-based measurement for a test specimen with known dimensions. The magnification factor is 8.1× (b) Sub-pixel registered AR overlay of related color-encoded depth information. (c) Depth scale ranging from near (blueish) to far (reddish).



(a) (b) (c)

Figure 12: (a) Results of in-vivo image-based measurement (b) Sub-pixel registered AR overlay of related color-encoded depth information. (c) Depth scale ranging from near (blueish) to far (reddish).

Fig. 12 for a test specimen and an in-vivo example with large magnification. In addition, Fig. 13 shows augmentation results of classified blood vessels as classical AR color overlay. As introduced in Sec. 2.3.3, further visualization options, e.g. adapted color space, are possible for the presentation of the spectral analysis.

3 Conclusion

We proposed a unique and integrated multimodal AR/MR framework using a fully digital surgical microscope for ENT surgery supporting local and remote assistance for intraoperative decision support.

The technological and market potential is very large for a product development based on our idea. Surgeries are the primary source (up to 50%) of errors in the health care sector [8]. These errors and related harm to the patient causes follow-up surgeries and treatments, which lead to additional costs [10]. A volumetric tissue differentiation and visualization tool, which can highlight and separate important anatomical structures in combination with an interactive remote assistance interface, will have a promising potential to decrease related expenses and number of accidental organ damage for the best patient outcome and safety.

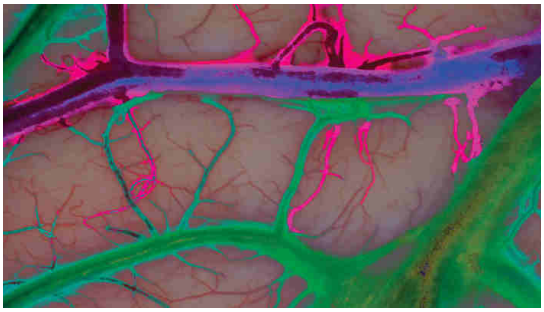


Figure 13: In this visualization example classified blood vessels (vein in green and arteria in red) are presented as AR overlay.

Acknowledgements

The described project MultiARC is funded under the grant number 16SV8061 by the German Federal Ministry of Education and Research (BMBF).

References

[1] ARRI Medical. Arriscope - the next step in all digital 3d surgical microscopy. http://www.arrimedical.com, Accessed, 09 January 2019.

[2] J.-Y. Bouguet. Camera calibration toolbox for Matlab, 2008. http://www.vision.caltech.edu/bouguetj/calib_doc/index.htm, Accessed, 09 January 2019.

[3] J. Brauers, N. Schulte, and T. Aach. Multispectral filter-wheel cameras: Geometric distortion model and compensation algorithms. IEEE Transactions on Image Processing, 17(12):2368–2380, 2008. [doi:10.1109/TIP.2008.2006605].

[4] M. Calonder, V. Lepetit, M. Ozuysal, T. Trzcinski, C. Strecha, and P. Fua. BRIEF: Computing a local binary descriptor very fast. IEEE Transactions on Pattern Analysis and Machine Intelligence, 34(7):1281–1298, 2012.

[5] S. Chen and Q. Liu. Modified Wiener estimation of diffuse reflectance spectra from RGB values by the synthesis of new colors for tissue measurements. J. Biomed. Opt., 17(3):030501, 2012.

[6] P. Eisert. Model-based Camera Calibration Using Analysis by Synthesis Techniques. In Proc. International Workshop on Vision, Modeling, and Visualization, pages 307–314, Erlangen, 2002.

[7] M. Feuerstein. Augmented reality in laparoscopic surgery. New concepts for intraoperative multimodal imaging. TU München, 2007.

[8] Frost and Sullivan. 9837-54, 2014.

[9] J. J. Fuertes, F. López-Mir, V. Naranjo, M. Ortega, E. Villanueva, and M. Alcañiz. Augmented reality system for keyhole surgery - performance and accuracy validation. In Proceedings of the International Conference on Computer Graphics Theory and Applications - Volume 1: GRAPP, (VISIGRAPP 2011), pages 273–279. INSTICC, SciTePress, 2011.

[10] D. G. für Orthopädie und orthopädische Chirurgie. Faktencheck gesundheit, 2013.

[11] Y. Garini, I. T. Young, and G. McNamara. Spectral Imaging: Principles and Applications. Cytometry Part A, 69(A):735–747, 2006.

[12] S. Garrido-Jurado, R. Muñoz-Salinas, F. J. Madrid-Cuevas, and M. J. Marín-Jiménez. Automatic generation and detection of highly reliable fiducial markers under occlusion. Pattern Recognition, 47:2280–2292, 2014.

[13] B. Geelen, C. Blanch, P. Gonzalez, N. Tack, and A. Lambrechts. A tiny VIS-NIR snapshot multispectral camera. In Advanced Fabrication Technologies for Micro/Nano Optics Photonics VIII, volume 9374 of Proc. SPIE, pages 937414–1, 2015.

[14] H. Haneishi, T. Hasegawa, A. Hosoi, Y. Yokoyama, N. Tsumura, and Y. Miyake. System design for accurately estimating the spectral reflectance of art paintings. Appl. Opt., 39(35):6621–6632, 2000.

[15] H. Hirschmuller. Accurate and efficient stereo processing by semi-global matching and mutual information. In 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), volume 2, pages 807–814 vol. 2, June 2005.

[16] A. J. Hung, J. Chen, A. Shah, and I. S. Gill. Telementoring and Telesurgery for Minimally Invasive Procedures. The Journal of Urology, 199(2):355–369, 2018.

[17] J. Ilgner, J. J.-H. Park, and M. Westhofen. Praktische Aspekte zur hochauflösenden Stereovideo-Dokumentation intraoperativer Befunde in der Hals-, Nasen- und Ohrenheilkunde. Bildverarbeitung für die Medizin, 14(1):450–454, 2010.

[18] M. Kamruzzaman, D. F. Barbin, G. ElMasry, D.-W. Sun, and P. Allen. Potential of hyperspectral imaging and pattern recognition for categorization and authentication of red meat. Innovative Food Science and Emerging Technologies, 16:316–325, 2012.

[19] W. S. Khor, B. Baker, K. Amin, A. Chan, K. Patel, and J. Wong. Augmented and virtual reality in surgery - the digital surgical environment: applications, limitations and legal pitfalls. Annals of Translational Medicine, 4(23):454–454, 2016.

[20] F. López-Mir, V. Naranjo, J. J. Fuertes, M. Alcañiz, J. Bueno, and E. Pareja. Design and validation of an augmented reality system for laparoscopic surgery in a real environment. BioMed Research International, 2013, 2013.

[21] G. Lu and B. Fei. Medical hyperspectral imaging: a review. Journal of Biomedical Optics, 19(1):10901, 2014. [doi:10.1117/1.JBO.19.1.010901].

[22] A. S. Luthman, S. Dumitru, I. Quiros-Gonzalez, J. Joseph, and S. E. Bohndiek. Fluorescence hyperspectral imaging (fHSI) using a spectrally resolved detector array. Journal of Biophotonics, 10(6):840–853, 2017.

[23] L. Maier-Hein. Optical techniques for 3d surface reconstruction in computer-assisted laparoscopic surgery. Medical Image Analysis, 17(974), 2013.

[24] MediClin. Struktur und Leistungen der Klinik für Chirurgie. Technical report, MediClin Müritz-Klinikum, Waren (Müritz), 2007.

[25] A. Meola, F. Cutolo, M. Carbone, F. Cagnazzo, M. Ferrari, and V. Ferrari. Augmented reality in neurosurgery: a systematic review. Neurosurgical Review, 40(4):537–548, 2017.

[26] J. Pichette, A. Laurence, L. Angulo, F. Lesage, A. Bouthillier, D. K. Nguyen, and F. Leblond. Intraoperative video-rate hemodynamic response assessment in human cortex using snapshot hyperspectral optical imaging. Neurophotonics, 3(4):045003, 2016. [doi:10.1117/1.NPh.3.4.045003].

[27] G. Polder, G. W. van der Heijden, L. P. Keizer, and I. T. Young. Calibration and characterisation of imaging spectrographs. Journal of near Infrared Spectroscopy, 11(3):193–210, 2003.

[28] T. A. Ponsky, M. Schwachter, J. Parry, S. Rothenberg, and K. M. Augestad. Telementoring: the surgical tool of the future. European Journal of Pediatric Surgery, 24(04):287–294, 2014.

[29] J.-C. Rosenthal, N. Gard, A. Schneider, and P. Eisert. Kalibrierung stereoskopischer Systeme für medizinische Messaufgaben. In 16th Annual CURAC Conference. PZH-Verlag, 2017.

[30] J.-C. Rosenthal, N. Gard, A. Schneider, and P. Eisert. Microscopic image-based determination of stapes prosthesis length. In Conference Proceedings 32nd Int. Congress of Computer Assisted Radiology and Surgery, CARS, pages 59–60, June 2018.

[31] I. Saliba and C.Marchica. The Relationship between Stapes Prosthesis Length and Rate of Stapedectomy Success. Clin Med Insights Ear Nose Throat, 8(23), 2015.

[32] N. Shimano, K. Terai, and M. Hironaga. Recovery of spectral reflectances of objects being imaged by multispectral cameras. J. Opt. Soc. Am. A, 24(10):3211–3219, 2007.

[33] C. Stamatopoulos and C. Fraser. Calibration of long focal length cameras in close range photogrammetry. The Photogrammetric Record, 26:339–360, 2011.

[34] C. Stamatopoulos, C. Fraser, and S. Cronk. On the self-calibration of long focal length lenses. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 38(Part 5), 2010.

[35] P. T. Stigell, K. Miyata, and M. Hauta-Kasari. Wiener estimation method in estimating of spectral reflectance from RGB images. Pattern Recognit. Image Anal., 17(2):233–242, 2007.

[36] R. Tsai. A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf TV cameras and lenses. IEEE Journal on Robotics and Automation, 3(4):323–344, Aug. 1987.

[37] P. Urban, M. Desch, K. Happel, and D. Spiehl. Recovering camera sensitivities using target-based reflectances captured under multiple led-illuminations. In Proc. of Workshop on Color Image Processing, pages 9–16, 2010.

[38] P. Urban, M. R. Rosen, and R. S. Berns. A spatially adaptive wiener filter for reflectance estimation. In Proc. IS T/SID Color Imaging Conf., volume 16, pages 279–284, 2008.

[39] P. Urban, M. R. Rosen, and R. S. Berns. Spectral image reconstruction using an edge preserving spatio-spectral Wiener estimation. J. Opt. Soc. Am. A, 26(8):1865–1875, 2009.

[40] P. Vávra, J. Roman, P. Zonča, P. Ihnát, M. Němec, J. Kumar, N. Habib, and A. El-Gendi. Recent Development of Augmented Reality in Surgery: A Review. Journal of Healthcare Engineering, pages 1–9, 2017.

[41] E. L. Wisotzky, B. Kossack, F. C. Uecker, P. Arens, S. Dommerich, A. Hilsmann, and P. Eisert. Validation of two techniques for intraoperative hyperspectral human tissue determination. In Proceedings of SPIE, volume 10951, page 1095171, 2019.

[42] E. L. Wisotzky, F. C. Uecker, P. Arens, S. Dommerich, A. Hilsmann, and P. Eisert. Intraoperative hyperspectral determination of human tissue properties. Journal of Biomedical Optics, 23(9):091409, 2018.

[43] Z. Zhang. Flexible camera calibration by viewing a plane from unknown orientations. In Proceedings of the Seventh IEEE International Conference on Computer Vision, Kerkyra, 1999.

[44] K. J. Zuzak, R. P. Francis, E. F. Wehner, J. Smith, M. Litorja, D. W. Allen, C. Tracy, J. Cadeddu, and E. Livingston. Hyperspectral imaging utilizing lctf and dlp technology for surgical and clinical applications. In Design and Quality for Biomedical Technologies II, volume 7170, page 71700C. International Society for Optics and Photonics, 2009.

[45] K. J. Zuzak, M. D. Schaeberle, E. N. Lewis, and I. W. Levin. Visible reflectance hyperspectral imaging: characterization of a noninvasive, in vivo system for determining tissue perfusion. Analytical chemistry, 74(9):2021–2028, 2002.