Pattern Recognition and Mixed Reality for Computer-Aided Maxillofacial Surgery and Oncological Assessment

Antonio Pepe
Institute of Computer Graphics and Vision
Graz University of Technology
Graz, Austria
antonio.pepe@tugraz.at

Gianpaolo Francesco Trotta
Department of Mechanics, Mathematics and Management
Polytechnic University of Bari
Bari, Italy
gianpaolofrancesco.trotta@poliba.it

Antonio Pepe, Gianpaolo Francesco Trotta, Christina Gsaxner, Jürgen Wallner, Jan Egger
Computer Algorithms for Medicine Laboratory
Graz, Austria

Christina Gsaxner, Jürgen Wallner, Jan Egger
Department of Oral and Maxillofacial Surgery
Medical University of Graz
Graz, Austria

Christina Gsaxner, Dieter Schmalstieg, Jan Egger
Institute of Computer Graphics and Vision
Graz University of Technology
Graz, Austria
egger@tugraz.at

Vitoantonio Bevilacqua
Department of Electrical and Information Engineering
Polytechnic University of Bari
Bari, Italy
vitoantonio.bevilacqua@poliba.it

***Abstract*— Maxillofacial tumor resection is often dependent on the expertise of the surgeon performing the operation. This study wants to be a first exploration of the role that commercial mixed reality headsets could have in this field. With this purpose in mind, a mixed reality Head Mounted Display (HMD) application is proposed to ease this task and provide a training tool. Due to the invasiveness of the operation, a marker-less registration has been considered as advantageous and therefore a pattern recognition algorithm has been adopted for properly placing a segmented PET-CT scan over the target face. To document the validity and appreciation rate of such a system, groups of physicians and engineers were asked to evaluate and assess the resulting prototype according to the standard ISO-9241/110. The application showed a noticeable accuracy of millimeters, consistent with other biomedical studies, due to intrinsic limitations of the device. Nonetheless, the remarkably positive feedbacks collected from both groups suggest high interest in further work.**

***Index Terms*— Mixed Reality, Human-Computer Interaction, Maxillofacial Surgery, Head and Neck Cancer, Medical Imaging, Microsoft HoloLens, PET-CT.**

I. Introduction

Head and neck cancer is the sixth most diffused type of cancer, accounting for 6% of all cases, with evident peaks in areas characterized by higher alcohol and tobacco consumptions [1]. A correct oncological assessment and resection of a tumor in the maxillofacial area is of crucial importance to reduce the chances of a relapse. However, to the best of our knowledge, surgeons have currently no way to precisely outline the affected areas by simply referring to the available data, such as a PET-CT scan of the patient. They rely on their personal understanding by means of palpation and operating microscopes, and on local examinations of the extracted tissues performed by pathologists [2][3][4]. The difficulty of the overall task varies from case to case and can lead to a partial misunderstanding of the resection margins, increasing the chances of a recurrence [5]. To support the specialist during this task, the introduction of a mixed reality (MR) application has been evaluated, both as a surgical support tool as well as for training purpose. Previous studies have already shown high interest for MR-aided surgery: a noteworthy example has evaluated a Microsoft HoloLens application for breast cancer resection [6]. Some investigations have been conducted also for maxillofacial surgery, using video see-through HMDs [7]. These devices are not being considered in this study, due to their reduced field of view and higher visualization lag. Furthermore, all the examined studies used physical markers for the registration of the 3D models in the scene. These markers are difficult to apply as this kind of surgery is often highly invasive, depending on the exact location of the malign masses. To avoid the introduction of such elements and still be able to estimate the exact position of the patient, a pattern recognition algorithm has been adopted to detect facial landmarks and use them as markers [8]. These biometric markers are then used to properly register the volumetric rendering of the tumor masses which were previously generated from a segmented PET-CT scan of the patient. In addition, practical sessions with physicians and technical experts have been organized to evaluate how this tool is assessed from both parts.

978-1-5386-5724-9/17/$31.00 ©2018 IEEE

II. Experimental framework

A. General software architecture

The necessary steps used to generate the 3D models are managed within a desktop application based on MeVisLab 3.0, using a PET-CT scan of the patient as input [9][10]. The output models are then imported within an augmented reality (AR) scene designed in Unity3D 2017.3, using the Mixed Reality Framework and Visual Studio 2017 [11][12][13][14][15].
This configuration allows to build applications which target the so-called Universal Windows Platform and therefore also compatible with Microsoft HoloLens.
Figure 1 gives an overview of the overall architecture, which is described in sections II.B and II.C.

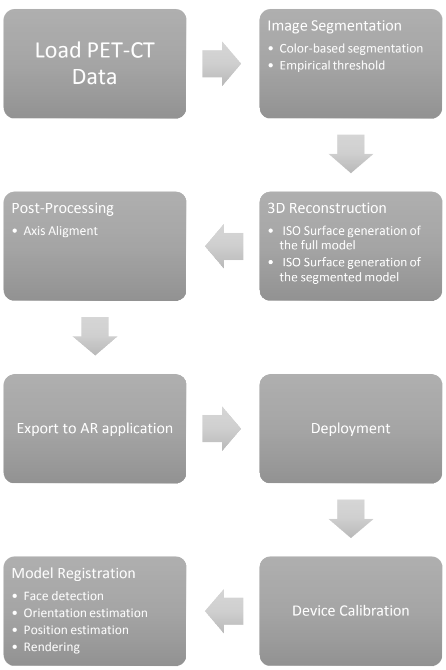

Figure 1 – The system workflow

B. MeVisLab application

The desktop application segments the slices of a given PET-CT scan according to a defined threshold. This provided satisfactory results with most of the available cases. These segmentations have been checked by a physician of the Medical University of Graz, whose knowledge and ability were considered as alloyed gold standard test. Given the non-central role of segmentation within this project, only the subset of segmented scans was considered.
To reduce the computations and the chance of false positives, the application allows the user to select a so-called Volume of Interest, which corresponded to the maxillofacial area: in this way the doctor can visualize also bigger volumes, in case a patient has tumors not only in this area. In the last step the application calculates the Iso-Surfaces of the segmented volumes and generates the 3D meshes, as shown in Figure 2. These steps are executed also without the segmentation step, so to have an Iso-Surface representation of the patient's face. Further transforms are manually applied to the model, to guarantee a fixed orientation parallel to the gaze, and a coincidence between the origin of the axis and the center of mass. For this last step, the scripts provided by MeshLab are used, which can also be integrated in a custom MeVisLab module [16].

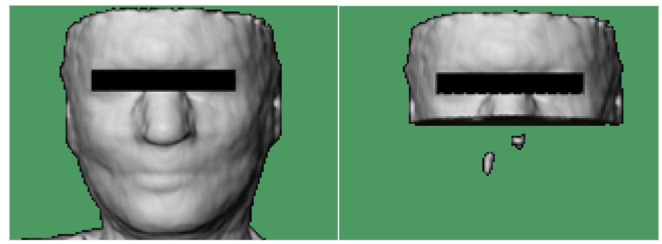

Figure 2 – 3D models generated with the desktop applications. Left: IsoSurface of the non-segmented scan. Right: IsoSurface of the segmented scan. The image includes also the forehead for a better understanding.

Eventually, the meshes are exported in a Wavefront file (.obj), as this format is natively supported by both MeVisLab and Unity3D. The lack of compression of the format did not represent a problem as the tumor masses were generally simple and small object.

C. Mixed-Reality application

To provide a fully automated registration process, a pattern-based approach has been chosen. The HMD provides a built-in RGB camera with well-known intrinsic parameters. The frames generated by this camera are processed using a pattern recognition algorithm. Although different methods are already available[17][18][19], the study focuses on the one proposed by Kazemi and Sullivan based on a cascade of regressors, as recommended also in other studies [8][20][21]. The implementation provided by DLib was used [22].

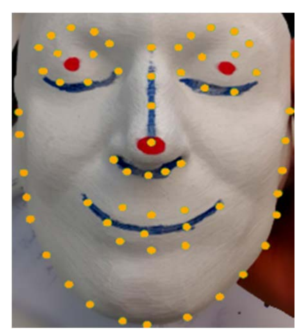

Figure 3 – Pattern recognition algorithm applied on a picture of the printed 3D model. In orange the landmark points.

The output of this step, as shown in Figure 3, determines the position of 68 landmark points with a well-defined meaning. The robustness of the algorithm permits to determine these points not only on the real person, but also on the printed phantom and on virtual images of the 3D model. A subset of these points is used to determine the direction to be followed to reach the target point, starting from the camera position and orientation in the AR scene. These vectors are estimated introducing the so-called projection and camera-view matrices, which are provided by the manufacturer.

Using the spatial reconstruction capabilities of the Spatial Mapping module, a ray-casting algorithm is used to determine the points along these directions which are closest to the camera. The distance between these points and the center of mass is calculated on the 3D model, so that the object can then be placed in the augmented environment, once the landmarks points are known. Furthermore, the triangle similarity principle is introduced as a correction factor to increase the accuracy. As the distance between facial landmarks is known a priori from the PET-CT scan, this information is used to estimate an overall distance of the patient's head. Eventually, an estimation of the face orientation is obtained solving a perspective-n-point problem over the two sets of points also using the Levenberg-Marquardt optimization [23][24].

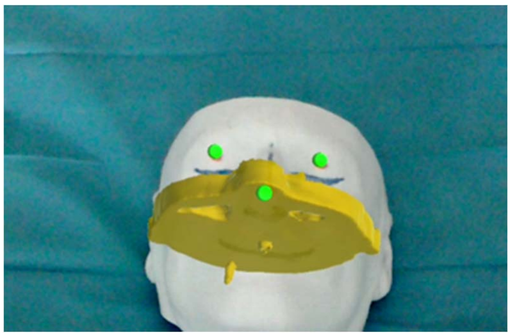

Figure 4 – An example of a registered model. The horizontal yellow line as well as the three green points were used as a reference. The two yellow masses (bottom) represent the tumors.

D. Feedbacks evaluation

To quantify the specialists' satisfaction with the prototype, each of them was asked to answer a questionnaire written according to the standard ISO-9241/110, as performed in other showcases of biomedical applications [25][26][27]. The answers were structured following a 6-points Likert scale, ranging from 0 to 5 i.e. from complete disagreement to complete agreement. The list of the questions is provided in Table 1, followed by the possible answers in Table 2.

Table 1 - ISO-9241/110-based questionnaire

1. The software is easy to use
2. The software is easy to understand without prior training
3. The software is easy to understand after an initial training
4. The software offers all the necessary functionalities (the minimum requirements are met)
5. The software successfully automates repetitive tasks (minimal manual input)
6. The way of interaction is uniform through the operation cycle
7. The introduction of the software considerably reduces the overall operation time (or manual tasks)
8. The introduction of the software could considerably increase the quality of the operation (e.g. less risk of failures, ...)
9. The introduction of a similar software in a surgery room would be beneficial
10. The software would be helpful for educational purposes
11. The software would be helpful in an ambulatory and/or doctor's office
12. I am an expert in the medical field
13. I am an expert in the field of human-computer interaction

Table 2 – List of possible answers for each question. Each of them is treated as a number 0-5 for averaging.

0. Completely disagree
1. Disagree
2. Somewhat disagree
3. Somewhat agree
4. Agree
5. Completely disagree

E. Test scenario

For testing and presentation purposes, given the difficulty of reserving a surgery room on a regular basis, the standard scenario was reconstructed in a room of the University-Hospital. The real patient was replaced with a 3D-printed head, reconstructed from a PET-CT scan. This allowed us to ignore the scale factor and any derived error when mapping the 3D model on the target. The phantom was placed over a bed with a blue-green background, to better resemble the surgery room and thus help the physicians in correctly picturing the real scenario, as shown in Figure 5.

Figure 5 – An image of the testbed.

Face traits and reference markers were also added on the printed model. The traits, i.e. lip-line and eyelashes, were helpful to both increase the detection rate as well as provide a better topological understanding of masses. The facial features extracted around these areas were not considered for our purpose, but only these on the nose and around the jaw, as they are well-distinguishable also on the 3D print. Red reference markers were placed on the nose tip and above the eyes to help have a tangible measure of the registration error. The same markers were added also on the virtual model, but in a green color.

III. Results

Practical tests have been conducted with the participation of six physicians and nurses from the Department of Oral and Maxillofacial Surgery of the Medical University of Graz and seven AR specialists from the Graz University of Technology. Each participant got the chance to try the application in the reconstructed test scenario and was then asked to fill a form containing the questions from Table 1. The forms were collected anonymously and separated in two clusters, physicians and technicians, respectively reported in Figure 6 and Figure 7. The main reason behind this is to be able to properly identify the medical needs without any possible influence introduced by the rest and vice versa.

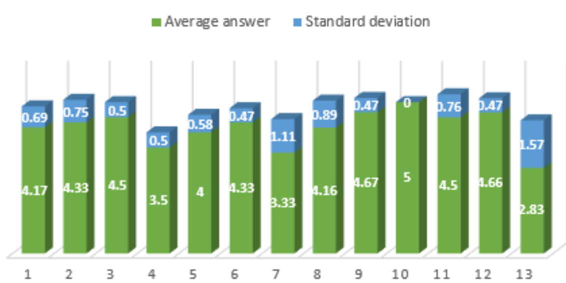

Figure 6 - Ensemble reporting of the answers provided by the medical group. The absolute values of the standard deviation (blue) are added on top of the average values (green).

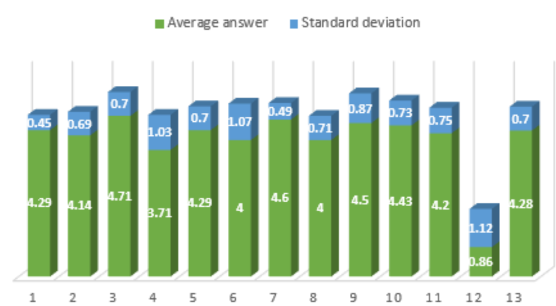

Figure 7 – Ensemble reporting of the answers provided by the technical group. The absolute values of the standard deviation (blue) are added on top of the average values (green).

To be able to quantitatively define the goodness of this approach compared to the one in the reference study [5], the overall positioning error of the system has also been measured. The results are shown in Table 3.

Table 3 – Overall error comparison between the current results and the reference study. The reference axes are parallel to those perceived by the user at registration time.

Error	Proposed approach (μ ± σ)	Reference approach (μ ± σ)
Up-Down dimension	(-4.5 ± 2.9) mm	(-1.1 ± 2.00) mm
Right-Left dimension	(3.3 ± 2.3) mm	(0.1 ± 1.2) mm
Back-front dimension	(-9.3 ± 6.1) mm	not specified

IV. Discussions

The application of mixed reality in this field shows promising results. Despite it being only an initial attempt, the work has been positively assessed from both groups. The medical group showed more criticism with questions 4 and 7, where they all provided similar reasons. Particularly, they did not see the application as a tool for reducing the overall time of the operation, but rather for increasing the overall quality and success rate. They also addressed further information to be visualized in the AR scene, including bones and main vessels, which will be included in the next studies.

Comparing the errors measured in both studies, it is possible to extract a pattern from the measurements. Despite the absence of clear physical markers, the error saw an increment limited to circa 3.3mm on average. Future work might aim at error reduction where the main constraint currently is given by the depth perception, due to intrinsic limitations of Microsoft HoloLens. Further approaches need to be evaluated to this regard, which may include the introduction of further sensors, like Microsoft Kinect and Intel Realsense, as well as further image-based approaches which may contribute to the error detection [28][29][30].

V. Conclusions

In this work, a method for the automatic registration of segmented PET-CT scans in mixed reality environments has been examined. The method uses a pattern recognition algorithm to detect facial landmarks on a patient as a starting point to estimate the real position and orientation of the target. To create a test-bed, a 3D model of a patient's head has been printed based on an Iso-Surface reconstruction of the PET-CT. The proposed method has shown good results compared to the current state of the art, although further work is needed to make the technique ready for a live surgery scenario. Nonetheless, given that this was only a first exploratory trial in this field, very positive feedback has been collected about its introduction in surgical trainings and preoperative planning of tumor resections in the maxillofacial area. Further work will investigate the application of numerical methods for the reduction of the positioning error, as well as the introduction of tracking methods which would allow to track the position of the tumor also during the complete duration of the operation.

Acknowledgment

This work received funding from the *Austrian Science Fund (FWF) KLI 678-B31*: "enFaced: Virtual and Augmented Reality Training and Navigation Module for 3D-Printed Facial Defect

Reconstructions" (PIs J. Wallner and J. Egger), from *POR PUGLIA FESR* – FSE 2014-2020 European Regional Development Fund – *Act 1.6 Project NR. 9757YR7 SOS* – Smart Operating Shelter, from *TU Graz Lead Project* (Mechanics, Modeling and Simulation of Aortic Dissection) and CAMed (COMET K-Project 871132) which is funded by the Austrian Federal Ministry of Transport, Innovation and Technology (BMVIT) and the Austrian Federal Ministry for Digital and Economic Affairs (BMDW) and the Styrian Business Promotion Agency (SFG).

References

[1] A. Argiris, M.K. Karamouzis, D. Raben, R.L. Ferris, Head and neck cancer, The Lancet, vol. 371, pp. 1695-1709, 2008.

[2] J. L. Connolly, et al. Role of the surgical pathologist in the diagnosis and management of the cancer patient, 6th ed., Holland-Frei Cancer medicine, 2003.

[3] National Institutes of Health, Technologies enhance tumor surgery: helping surgeons spot and remove cancer, News In Health, 2016.

[4] D.A. Ross, S. Ariyan, R. Restifo, C.T. Sasaki, Use of the operating microscope and loupes for head and neck free microvascular tissue transfer: a retrospective comparison, Arch Otolaryngol Head Neck Surg., 2003, 129(2):189-93.

[5] M. Amit, S. Na'ara, L. Leider-Trejo, S. Akrish, J.T. Cohen, S. Billan, Z. Gil, Improving the rate of negative margins after surgery for oral cavity squamous cell carcinoma: A prospective randomized controlled study, Head&Neck, vol. 38, pp. E1803-E1809, 2015.

[6] S. Perkins, et al. A mixed-reality system for breast surgical planning, IEEE ISMAR Adjunct Proceedings, Nantes, 2017.

[7] J. Wang, et al. Video see-through augmented reality for oral and maxillofacial surgery, Int J Med Robot, 2017, 13(2).

[8] V. Kazemi and J. Sullivan, One millisecond face alignment with an ensemble of regression trees, IEEE CVPR, 2014.

[9] MeVisLab, MeVis Medical Solutions AG. [Online]. https://www.mevislab.de/. [Accessed 27-Aug-2018]

[10] J. Egger, et al. HTC Vive MeVisLab integration via OpenVR for medical applications, PLoS ONE, vol. 12, 2017.

[11] Unity 3D, Unity Technologies. [Online] Available at https://www.unity3d.com/ [Accessed 27-Aug-2018]

[12] Mixed Reality Toolkit, Microsoft. [Online] Available at https://github.com/Microsoft/MixedRealityToolkit-Unity [Accessed 27-Aug-2018]

[13] Visual Studio, Microsoft. [Online] Available at https://visualstudio.microsoft.com/ [Accessed 27-Aug-2018]

[14] V. Bevilacqua, A. Brunetti, G. Trigiante, G.F. Trotta, M. Fiorentino, V.M. Manghisi, A.E. Uva, Design and Development of a Forearm Rehabilitation System Based on an Augmented Reality Serious Game. In Italian Workshop on Artificial Life and Evolutionary Computation, pp. 127-136, CCIS, vol. 587. Springer, Cham, 2015 doi: https://doi.org/10.1007/978-3-319-32695-5_12

[15] A.E. Uva, M. Fiorentino, M. Gattullo, M. Colaprico, M.F. de Ruvo, F. Marino, G.F. Trotta, V.M. Manghisi, A. Boccaccio, V. Bevilacqua, G. Monno, Design of a projective AR workbench for manual working stations. In International Conference on Augmented Reality, Virtual Reality and Computer Graphics (pp. 358-367). Springer, Cham, 2016 doi: https://doi.org/10.1007/978-3-319-40621-3_25

[16] MeshLab, CNR Visual Computing Lab. [Online] Available at http://www.meshlab.net/ [Accessed 27-Aug-2018]

[17] V. Bevilacqua, P. Casorio, G. Mastronardi, Extending Hough Transform to a Points' Cloud for 3D-Face Nose-Tip Detection, International Conference on Intelligent Computing (ICIC2008), pp. 1200-1209, vol. 5227, Springer, 2008 doi: https://doi.org/10.1007/978-3-540-85984-0_144

[18] V. Bevilacqua, L., Cariello, D., Columbo, D., Daleno, M., Dellisanti Fabiano, M., Giannini, G., Mastronardi, M., Castellano, Retinal Fundus Biometric Analysis for Personal Identifications, International Conference on Intelligent Computing (ICIC2008), pp. 1229-1237 vol. 5227, 2008 doi: https://doi.org/10.1007/978-3-540-85984-0_147

[19] V. Bevilacqua, G. Filograno, G. Mastronardi, Face detection by means of skin detection, International Conference on Intelligent Computing (ICIC2008), pp. 1210-1220, vol. 5227, Springer, 2008 doi: https://doi.org/10.1007/978-3-540-85984-0_145

[20] V. Bevilacqua, D. D'Ambruoso, G. Mandolino, M. Suma, A new tool to support diagnosis of neurological disorders by means of facial expressions, IEEE International Symposium on Medical Measurements and Applications, Medical Measurements and Applications Proceedings (MeMeA2011), pp. 544-549, 2011 doi: https://ieeexplore.ieee.org/document/5966766/

[21] V. Bevilacqua, A.E. Uva, M. Fiorentino, G.F. Trotta, M. Dimatteo, E. Nasca, A.N. Nocera, G.D. Cascarano, A. Brunetti, N. Caporusso, R. Pellicciari, G. Defazio. A comprehensive method for assessing the blepharospasm cases severity, International Conference on Recent Trends in Image Processing and Pattern Recognition (RTIP2R2016), pp. 369-381, vol. 709 Springer, 2016 doi: https://doi.org/10.1007/978-981-10-4859-3_33

[22] Dlib, http://dlib.net/. [Online] [Accessed 27-Aug-2018]

[23] Y. Zheng, Y. Kuang, S. Sugimoto, K. Åström, M. Okutomi, Revisiting the PnP Problem: A Fast, General and Optimal Solution, ICCV, 2013.

[24] D. W. Marquardt, An Algorithm for the Least-Squares Estimation of Nonlinear Parameters, SIAM Journal of Applied Mathematics, vol. 11, no. 2, pp. 431-441, 1963.

[25] J. Egger, et al. Interactive reconstructions of cranial 3D implants under MeVisLab as an alternative to commercial planning software, in PLoS One, vol. 12, 2017.

[26] J. Egger, et al. Computer-aided position planning of miniplates to treat facial bone defects, in PLoS One, vol. 12, 2017.

[27] S. Ziebarth, A. Kizina, H. U. Hoppe, L. Dini, A Serious Game for Training Patient-Centered Medical Interviews, IEEE ICALT, 2014.

[28] V. Bevilacqua, L. Biasi, A. Pepe, G. Mastronardi, N. Caporusso, A Computer Vision Method for the Italian Finger Spelling Recognition, International Conference on Intelligent Computing (ICIC2015) pp. 264-274, LNCS vol. 9227, Springer, 2015 doi: https://doi.org/10.1007/978-3-319-22053-6_28

[29] M. Garon, P.O. Boulet, J.P. Doironz, L. Beaulieu, J.F. Lalonde, Real-Time High-Resolution 3D Data on the Hololens, IEEE ISMAR, 2016.

[30] R. Radkowski, J. Ingebrand, HoloLens for Assembly Assistance – A Focus Group Report, International Conference on Virtual, Augmented and Mixed Reality, LNCS vol. 10280, Springer, 2017