[POSTER] Augmented Reality for User-Friendly Intra-Oral Scanning

Janine Thoma[1*]   Michal Havlena[1†]   Severin Stalder[2‡]   Luc Van Gool[1,3§]

[1] Computer Vision Laboratory, ETH Zurich, Switzerland   [2] MHT Optic Research AG, Switzerland   [3] PSI, KU Leuven, Belgium

ABSTRACT

Digital impressions of teeth, obtained through intra-oral scanning, allow for more efficient and cost effective treatments of many dental indications. Current state-of-the-art intra-oral impression acquisition systems make use of a separate monitor to show the scanning progress, forcing the dentist to divert attention away from the scanner and the patient. In this paper, we present an augmented reality based solution to this problem. During the scanning process, an optical see-through head-mounted display is used to show an online overlay of the dynamic dental model onto the patient's teeth. The dentist can then fully focus on the patient and the scanner, while still being able to keep track of the current state of the model. This type of novel application, which fundamentally changes the human-computer interaction of intra-oral scanning systems, requires a fast and accurate registration of a dynamically growing model onto a glossy, partially occluded surface at a very small scale. To meet this demand, we propose application tailored algorithms for indirect high accuracy online 3D teeth tracking and optical see-through head-mounted display calibration. Experimental results indicate that our system does have a potential to noticeably facilitate intra-oral scanning in the future.

Keywords: Intra-oral scanning, augmented reality, HMD calibration, implicit pose estimation.

Index Terms: H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Artificial, augmented, and virtual realities

1 INTRODUCTION

With the increasing availability and continuously improving performance of optical see-through head-mounted displays (OST-HMDs), the potential of augmented reality is ever increasing. In this paper, we introduce one possible future application of augmented reality targeted at facilitating intra-oral scanning. The goal of intra-oral scanning, as shown in Figure 1, is to acquire digital impressions of patients' mouths. Such digital impressions of teeth are used in prosthetics, implantology, and orthodontics, where they build a basis for restorations and surgical guides. In combination with chair side milling, i.e. the manufacturing of fillings and restorations in small milling units directly at the dentist's practice, indications which formerly required dentist visits spread over several days are now treatable more efficiently.

A crucial drawback of current state-of-the-art intra-oral scanning systems is the need for the dentist to monitor the scanning process on a separate screen. Such monitoring is necessary because reflectance, steep surfaces, and fast movements of the patient, the scanner, or the dentist himself may lead to holes and missing parts in the scanned impression. This necessity to divide the dentist's attention between a separate monitor on one side, and the patient

---

*e-mail:jthoma@vision.ee.ethz.ch

†e-mail:michal.havlena@vision.ee.ethz.ch

‡e-mail:severin.stalder@mht.ch

§e-mail:vangool@vision.ee.ethz.ch

Figure 1: Intra-oral scanning process with a separate screen (upper image of CEREC Omnicam by Dentsply Sirona) and our proposed system with augmented reality glasses (lower image) for visualizing the dynamically growing teeth model.

and the scanner on the other, renders the scanning process mentally challenging and unintuitive. The system presented in this paper resolves this problem by bringing those two sides together. By incorporating an OST-HMD, the proposed system directly overlays the dynamically growing scanned model onto the patient's teeth, eliminating the need for monitoring the process on an additional screen. This results in a more user-friendly intra-oral scanning system. The improved usability has the potential to shorten scanning times and thus make the treatments more comfortable and cost effective for patients.

A useful overlay of the impression during scanning requires fast and accurate registration of a dynamically growing model onto a glossy, slightly translucent, partially occluded surface. We address this problem by leveraging on the model pose information obtained by the intra-oral dental scanner. Given the relative pose of the teeth and the scanner, it becomes possible to estimate the pose of the teeth with respect to the AR glasses by solving the easier task of estimating the pose of the dental scanner w.r.t. the AR glasses instead.

In addition to correct teeth pose estimation, a correct display of

IEEE computer society

the model also requires an accurate calibration of the OST-HMD. The calibration process should be fast, user-friendly, and, to avoid additional increase of the cost of the system, not involve any additional active hardware. To this end, our proposed system employs Single Point Active Alignment [16] calibration with a vision-based tracker.

The main contributions of this paper lie both in proposing a novel application of augmented reality, which fundamentally changes human-computer interaction in intra-oral scanning systems, as well as in providing the necessary problem tailored algorithms to achieve the desired online pose estimation and required calibration accuracies for our proposed system.

The remainder of this paper is structured as follows: A review of the existing work on the vision aspects of our system is given in Section 2. Section 3 outlines the different components and algorithms used in our proposed system. Experimental evaluation and conclusions are located in Sections 4 and 5.

2 RELATED WORK

Augmented Reality and Intra-Oral Scanning   With regard to dentistry in general, video augmentation was used to provide restoration previews after intra-oral scanning [1] in the past. For this purpose, a preview of the proposed improvements was superimposed onto a video of the patient's mouth in place of the original teeth. Unlike our proposed application, the work [1] neither uses OST-HMDs, nor does it aim at facilitating the process of intra-oral scanning for dentists. To the best of our knowledge, the presented project is the first one aiming at using augmented reality for user-friendly intra-oral scanning.

3D Tracking   The teeth tracking approach introduced in this paper requires an online six degree of freedom (DoF) pose estimation of the rigid intra-oral scanner based on the image sequence acquired by the monocular camera of the AR glasses. There are numerous techniques and approaches addressing this problem, e.g., [7]. An overview of tracking paradigms applied in augmented reality can be found in [21]. Unlike many recent works on tracking for augmented reality, which use hybrid visual-inertial tracking approaches [12] or multiple cameras, our framework is purely vision-based. To facilitate tracking, even in situations when some of the frames do not provide any trackable features due to motion blur, we restrict ourselves to estimating the pose of the current frame only. The teeth themselves are being tracked using a method similar to [17].

Calibration   OST-HMD calibration is a crucial aspect of research in augmented reality. A popular method for head-mounted display calibration is the Single-Point Active Alignment Method (SPAAM) [16]. The main advantage of the SPAAM approach is that the head of a user is allowed to move during calibration. A sizable number of calibration approaches [5, 2, 15, 8, 11] are based on SPAAM. Evaluations of different SPAAM versions [15, 9] suggest that Stereo SPAAM [4] is more accurate than monocular SPAAM [16]. However, the results in [15] also indicate that monocular SPAAM is faster. The SPAAM variant used in this work is similar to Depth-SPAAM described in [15], which is a variant of SPAAM that forces the user to take calibration measurements at different distances to the calibration point. In contrast to [15], these movements are not strictly enforced but only recommended. Also, the magnetic tracker is replaced with a purely vision-based one. Furthermore, our system improves the user-friendliness of the calibration process by allowing adjustments of the display area used for calibration. The use of a vision-based tracker requires the calibration of the corresponding camera. This paper uses the camera calibration solution introduced in [20]. Other recent approaches dealing with head-mounted display calibration include [10] that use unconstrained eye-cameras and [19, 18] that use a camera based approach to capture as many calibration parameters offline as possible.

3 SYSTEM AND ALGORITHMS

Our proposed system comprises three different hardware components. At the core of every intra-oral scanning system is the *scanner* itself. In addition to the scanner, a dental impression acquisition system requires a *processing unit* responsible for building the complete 3D teeth model by stitching together sequentially acquired single scan images. In current systems, the screen of this processing unit is used to display the progress of 3D teeth modeling. In our proposed system, the modeling progress is superimposed directly onto the patient's teeth. For this purpose, our system also includes an *OST-HMD*. For a more detailed description of the hardware components used for the research prototype of our proposed system see Section 3.2. The partially hardware specific interactions between the different components of our proposed system are then described in Section 3.3.

3.1 System Requirements

Based on the requests and feedback the experienced marketing people at the company got from their large user base, and an additional, more detailed discussion with dentists from the University of Zurich, it was clear that the proposed solution is seen as useful to make intra-oral scanning more user-friendly, but that the following system requirements should be met to maximize the positive impact of the proposed solution on the dentist's scanning performance:

- The displayed image overlay should not occupy large parts of the central field of view of the dentist.

- The system should show the visible as well as the occluded parts of the scanned model to facilitate the assessment of the scan quality.

The first requirement implies that simply using an OST-HMD as an alternative screen for what is normally displayed on the monitor of a processing unit, is not sufficient. Our system meets this requirement by only showing the model at the scale of, and aligned to, the teeth, thus avoiding introducing unnecessary clutter to the dentist's field of view. The second requirement is inherently met by the proposed system. It implies that detecting occlusions of the scanned model and including them in the rendering process is not necessary, and even not desirable.

3.2 Prototype Hardware Specifications

Intra-Oral Scanner   All experiments presented in this paper are performed with an MHT *3D Progress* intra-oral scanner. This scanner captures depth maps of teeth by projecting a laser pattern onto them and varying the focal length of its camera, similar to confocal microscopy. The sequentially acquired single depth maps are stitched into a 3D teeth model.

Processing Unit   In the proposed system, the processing unit is a standard desktop personal computer, running Windows 10, with the following specifications: Intel Core i7-6700 CPU @ 3.40GHz having 4 cores (8 threads), 16GB RAM, and NVIDIA GeForce GTX 960. The GPU is used by the scanner processing library for scan stitching and model rendering. It is currently not used for the algorithms presented in this paper.

OST-HMD   The research prototype of our proposed system uses the Moverio Pro BT-2000 augmented reality glasses, which run Android 4.0.4. The corresponding specifications can be found in [13]. In addition to its display, the Moverio Pro BT-2000 features a number of environmental sensors. For the sake of portability to other AR glasses, we restrict our system's use of sensors to the Moverio Pro BT-2000's monocular color camera and refrain from using its GPS, geomagnetic sensor, IMU, illumination sensor, microphone, stereo camera, and depth image capturing capabilities.
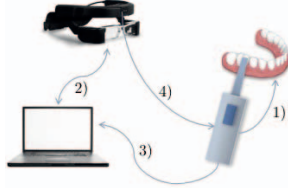
Figure 2: System components and connections.

Furthermore, we modify the Moverio Pro BT-2000 by rigidly binding the optical see-through display to the device's camera bar in order to allow for camera-display relative pose calibration.

3.3 System Connections

The three components of our system, i.e. the scanner, the processing unit, and the OST-HMD are interconnected with each other in various ways. Figure 2 shows the three components as well as the connections, which can be read as follows:

1. The scanner sequentially acquires overlapping depth frames of the teeth and stitches them together. As a byproduct of the stitching process, the scanner calculates the pose of every frame with respect to the first one, thus providing a measure for the current scanner pose relative to the origin of the teeth point cloud model. The current pose, as well as the point cloud model, is accessible through the API of the 3D Progress software.

2. The AR glasses are connected with the processing unit via TCP-IP. In this light, a TCP server application for the AR glasses and a corresponding client application for the processing unit were implemented. These applications are also used during OST-HMD calibration.

3. The connection between the scanner and the processing unit is part of the intra-oral scanning system of 3D Progress. On the side of the processing unit, the 3D Progress software provides an API over which the scanner can be accessed. This API allows for the following: Initialize, start, pause, resume, and stop scanning process, read and set scanner configuration values (e.g. scanning plaster teeth as opposed to real ones requires a different configuration), read information on teeth model point cloud and its normals, display a rendered version of the current model, as well as read and set current rotation and translation of the scanner relative to the point cloud model. Since the projection matrix of the scanner software's rendering framework is fixed and cannot be set via API, it is necessary to read out the point cloud information and render it separately.

4. The direct connection between the AR glasses and the scanner is only unidirectional. The camera of the glasses tracks the scanner.

3.4 Calibration

The calibration method used in our system is a variant of SPAAM [16] with several modifications, such as a purely vision-based tracker, an adjustable calibration display area for higher user comfort, and a variable depth of the calibration points.

Calculation of G The aim of our OST-HMD calibration approach is to find two 3×4 projection matrices G_C^L and G_C^R which directly relate homogeneous 3D points given in the AR glasses' camera coordinates C, to their corresponding homogeneous 2D left and right display projection coordinates. Given at least six 3D-to-2D correspondences, G_C^L and G_C^R can be found as described in [16].
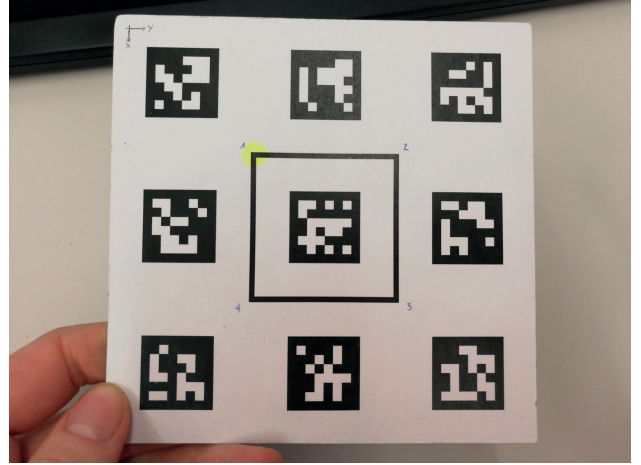

Figure 3: Calibration board for OST-HMD calibration. The calibration point is marked in yellow.

Calibration Procedure During the calibration procedure, the necessary 3D-to-2D correspondences for calculating G_C^L and G_C^R are captured. Based on the feedback from several users, the proposed procedure calibrates the left and the right displays sequentially. At the beginning of the calibration process, the user is shown a rectangle on each screen, indicating the display area in which 2D calibration points will be shown. The initial size of the rectangles is set to be suitable for most users but can also be adjusted manually. This allows the user to exclude areas of the glasses' displays which are entirely invisible to them or which they can only see by wearing the glasses in an uncomfortable manner.

After the calibration display area has been confirmed, 32 3D-to-2D correspondences are captured (16 per eye). For this purpose, the user is asked to align a cross-hair, displayed on the current side of the OST-HMD, to a calibration point, indicated in yellow, on a cardboard calibration board as shown in Figure 3. The cross-hairs are displayed in predefined 2D display positions on a rectangular grid which is designed to cover as much of the user defined calibration area as possible. The 3D location of the calibration point with respect to the AR glasses' camera is calculated using the fiducial markers on the calibration board. The user is instructed to use variable distances between AR glasses and calibration board for capturing the 3D-to-2D correspondences.

Camera Calibration In order to capture the pose of the OST-HMD calibration board, it is necessary that the AR glasses' camera is calibrated. Camera calibration is performed using a chessboard pattern.

3.5 Indirect 3D Tracking

To properly overlay the scanned model onto the corresponding teeth being scanned, it is necessary to track the 3D pose of the teeth in all six degrees of freedom. This type of problem is also known as pose estimation. However, directly tracking the teeth leads to several issues:

- Teeth are mostly untextured, specular and slightly translucent. This has a negative impact on any vision-based approach of tracking them.
- Teeth are naturally occluded by lips, tongue, cheeks, and other teeth. During intra-oral scanning, additional occlusions are caused by the dentist's hand, the tip of the scanner, and other tools, such as dental mirrors. An example of this can be seen in Figure 4.
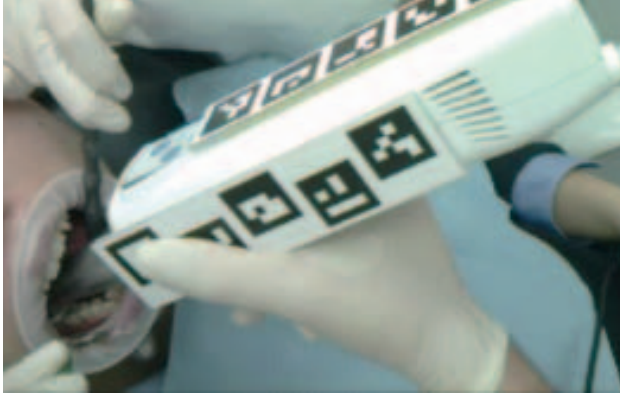
Figure 4: Intra-oral scanning process as seen from the head-mounted camera of a Moverio Pro BT-2000.
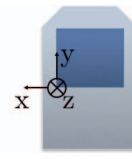

Figure 5: Scanner model coordinates.

Figure 6: Scanner with an additional marker at the tip (the coordinate system origin) during image acquisition for SfM.

- The 3D teeth model is built incrementally during the scanning process. Consequentially, at the beginning of the process, the model to be tracked is very limited in size and might strongly resemble several areas of the patient's teeth. Even if the teeth were to be tracked with a generic model, the question on where the initial scanned model fragment is located within the generic model remains.

It is for these reasons that the proposed system uses an indirect pose estimation approach. This approach refrains from explicitly tracking the teeth. Instead, it integrates the model pose information provided by the scanner processing library. The scanner's software calculates the pose of every newly scanned depth map in relation to the model point as a part of its stitching process. Leveraging this information, we can track the teeth indirectly by determining the pose of the scanner instead. In the following paragraphs, the pose estimation of the scanner and the integration of the scanner and teeth poses are explained in more detail. Furthermore, a brief discussion on the presented approach is given.

Scanner Pose Estimation   The 3D scanner pose with respect to the AR glasses' camera is estimated using a model based approach. Due to the lack of reliably and repeatedly localizable features on the glossy scanner surface, five planar fiducial markers, as described in [3], are placed on each of the three planar sides of the 3D Progress scanner. The large number of markers adds some redundancy in case of occlusions and increases robustness against inaccuracies in marker corner detection. The corners of the markers are reconstructed into a 3D virtual scanner model by employing structure from motion (SfM) techniques. For the research prototype scanner model, marker corner locations from a sequence of 475 undistorted images are used. The 3D reconstruction is done using Bundler [14], an open-source SfM software.

The obtained scanner model is defined up to a 3D similarity transformation. It is beneficial to register the SfM scanner model origin with the lower left corner of the tip of the scanner as shown in Figure 5. This simplifies the concatenation of relative teeth and scanner poses. To facilitate the calculation of the corresponding similarity transform, an additional marker is added to the tip of the scanner during the image acquisition for SfM. The marker's sides are aligned with the scanner window and its first corner coincides with the coordinate system origin as shown in Figure 6.

Given the 3D scanner model the relative pose of the AR glasses' camera with respect to the scanner is calculated for every frame as follows:

1. The received frame is undistorted.

2. The 2D corner locations of all markers visible in the frame are detected.

3. The 3D scanner coordinates corresponding to the detected marker corners are retrieved from the scanner model.

4. Using the pose of the previous frame as a prior, the rotation and translation of the camera relative to the scanner tip is calculated based on the established 3D-to-2D point correspondences and the intrinsic camera parameters obtainable by camera calibration. This type of problem is referred to as a perspective-n-point (PnP) problem [6].

The function used for the final step employs an iterative, Levenberg-Marquardt optimization based approach to solve the PnP problem. This method minimizes the reprojection error E, which for a set of n image points $(x_i, y_i)^T$ and the corresponding projected model points $(\tilde{x}_i, \tilde{y}_i)^T$ is calculated as

$$E = \sum_{i=1}^{n} \left[(x_i - \tilde{x}_i)^2 + (y_i - \tilde{y}_i)^2 \right]. \tag{1}$$

Pose Integration   In the following, all coordinate systems relevant for pose integration are described:

Scanner model S The coordinates in which the model of the scanner is defined. By design choice, the origin of this model lies at the tip of the scanner. This coordinate system is right handed and measured in centimeters.

Teeth model T The coordinates in which the teeth model point cloud is defined. For the initial scan its origin coincides with the one of the scanner model. After that, the relation of these systems is given as the current transformation provided by the scanner's API. This coordinate system is left handed and measured in millimeters.

Head-mounted Camera C What is needed to properly render the teeth model, given the $\mathtt{G}_C^L$ and $\mathtt{G}_C^R$ matrices for the left and right display transformations, found during OST-HMD calibration, is the relative pose of the teeth model and the AR glasses' camera. This coordinate system is right handed and measured in centimeters.

For a given point $(x_T, y_T, z_T)^T$ of the scanner model point cloud, its projection $(x_L, y_L)^T$ onto the left display can be calculated using the following equations:

$$\begin{pmatrix} x_S \\ y_S \\ z_S \\ 1 \end{pmatrix} = \begin{pmatrix} 0.1 & 0 & 0 & 0 \\ 0 & -0.1 & 0 & 0 \\ 0 & 0 & 0.1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \mathtt{T}_T^S \begin{pmatrix} x_T \\ y_T \\ z_T \\ 1 \end{pmatrix} \tag{2}$$

$$\alpha \begin{pmatrix} x_L \\ y_L \\ 1 \end{pmatrix} = \mathtt{G}_C^L \cdot \mathtt{T}_S^C \begin{pmatrix} x_S \\ y_S \\ z_S \\ 1 \end{pmatrix} \tag{3}$$

	Video see-through	OST-HMD
Transmission [ms]	5.16 ± 0.91	124.66 ± 24.96
Pose estimation [ms]	15.77 ± 1.38	14.86 ± 1.92
Rendering [ms]	73.81 ± 38.80	83.35 ± 32.12
Total round trip time [ms]	92.75 ± 28.37	217.97 ± 30.61

Table 1: Time consumption of the current system prototype. Times are given per fame as mean duration plus or minus standard deviation.

where $\mathtt{T}_T^S$ is transformation from the teeth model coordinates to scanner model coordinates, i.e. the current transformation provided by the scanner API, $\mathtt{T}_S^C$ is the transformation from the scanner model coordinates to camera coordinates as calculated during scanner pose estimation, and $\mathtt{G}_C^L$ is the projection from 3D camera coordinates to the 2D coordinates of the left display as calculated during the OST-HMD calibration. The projection of the same 3D point onto the right display, $(x_R, y_R)^T$, can be computed in a similar manner, just using projection matrix $\mathtt{G}_C^R$ instead.

Indirect 3D Tracking Discussion   The tracking approach described in this section avoids many issues related to directly tracking teeth. However, it assumes that the current transformations provided by the scanner software are accurate. This assumption holds true in general, especially with regard to translations, except for cases when the scanner loses track of the teeth being scanned and is unable to stitch the current scan to the existing model. Such occurrences are caused by excessively fast movements of the scanner and can be avoided when scanning slowly and steadily.

While the use of large and easy to recognize markers allows the approach to track the scanner reliably, such large markers may not be desirable in a consumer product. For a commercial version of the proposed system, the markers can be replaced by a texture spread over the surface of the scanner case. Then, the only constraint on the texture is that it should be rich in unique features, such that correspondence can still be established uniquely.

4   EVALUATION

4.1   Computational efficiency

Table 1 summarizes the time consumption of the current system prototype. The numbers have been calculated for the full system and a simpler video see-through version. This video see-through system uses a Logitech C170 webcam and the same processing unit as the full system. The resulting live video with added model overlay is displayed on the processing unit's monitor. This modality allows for the verification of scanner tracking and model alignment accuracy independent of OST-HMD calibration performance.

The numbers in Table 1 are based on three scans per system, taking into consideration point clouds with 30'000 to 100'000 points. This results in a total of 150 frames for OST-HMD, and 300 frames for video see-through.

The current frame rate is sufficient to demonstrate the effectiveness of the proposed system and algorithms. If a higher frame rate is desired, e.g. to create a consumer product based on the system described in this paper, the model rendering speed should be addressed. First experiments with an adapted version of the scanner software, which allows for the setting of projection matrices, show that the rendering time can be reduced to an average of $20.80.75 \pm 7.68$ milliseconds for video see-through and 21.90 ± 7.40 for OST-HMD. The resulting video see-through frame rate meets real-time requirements. For the OST-HMD case, the transmission time could be reduced by replacing the frame-based connection from processing unit to HMD with a video streaming protocol.
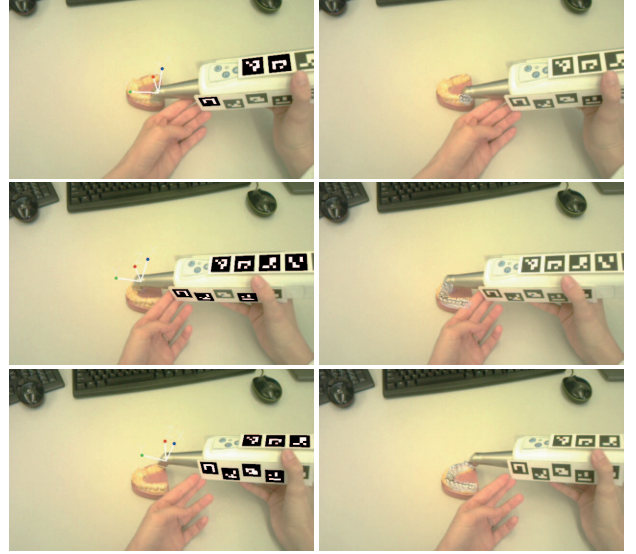

Figure 7: Three sample frames of a teeth scanning process on a plaster model. *Left column:* Images captured by the head-mounted camera with the estimated scanner poses. *Right column:* Images with video see-through augmentations of the scanned teeth model.

Figure 8: Optical overlay augmentations of the scanned teeth model for the left and right display, corresponding to the model in the bottom row of Figure 7.

4.2   Visualization of Calibration and Tracking Results

Figure 7 illustrates the performance of our system with regard to teeth pose estimation, which is crucial for the proposed application. The figure shows the processing results for three sample frames achieved in a laboratory setting on an artificial (plaster) teeth model with the prototype of our proposed system. A plaster model was chosen, because with the indirect tracking approach, all direct interaction with the scanned object is handled by the scanner. Hence, scanning on real teeth does not pose an additional challenge for the system proposed in this paper. However, the absence of occlusions such as lips makes the visual assessment of augmentation alignment easier on plaster models.

The first column of images in Figure 7 shows the scanner and the artificial teeth as seen from the monocular camera of the Moverio Pro BT-2000. The detected ArUco markers are highlighted and the estimated pose of the tip of the scanner is projected into the image. The second column shows the model as a video see-through augmentation, projected using the six degrees of freedom pose information obtained with our implicit pose estimation algorithm and the pre-calibrated intrinsic and distortion parameters of the monocular Moverio Pro BT-2000 camera. Figure 8 shows the model projections for the optical overlay augmentations for the left and right displays of the head-mounted display corresponding to the bottom row model of Figure 7. The results presented in Figure 7 indicate, that the alignment achieved with our approach is sufficiently accurate for the intended application.

5 CONCLUSION

In this paper, we have introduced a novel application of augmented reality in the field of intra-oral scanning. We have developed the necessary application-tailored 3D tracking and calibration algorithms and have created a prototype of the proposed system, employing these algorithms. The current prototype provides the user with promisingly accurate teeth model augmentations that we plan to test extensively in clinical settings. We can therefore conclude that our system does have the potential to noticeably facilitate intra-oral scanning in the future.

REFERENCES

[1] Kapanu. http://www.kapanu.com/, March 2017.

[2] A. L. Fuhrmann, R. Splechtna, and J. Přikryl. Comprehensive calibration and registration procedures for augmented reality. In *Immersive Projection Technology and Virtual Environments 2001*, pages 219–227. Springer, 2001.

[3] 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(6):2280–2292, 2014.

[4] Y. Genc, F. Sauer, F. Wenzel, M. Tuceryan, and N. Navab. Optical see-through hmd calibration: A stereo method validated with a video see-through system. In *Augmented Reality, 2000.(ISAR 2000). Proceedings. IEEE and ACM International Symposium on*, pages 165–174. IEEE, 2000.

[5] Y. Genc, M. Tuceryan, A. Khamene, and N. Navab. Optical see-through calibration with vision-based trackers: Propagation of projection matrices. In *Augmented Reality, 2001. Proceedings. IEEE and ACM International Symposium on*, pages 147–156. IEEE, 2001.

[6] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, second edition, 2003.

[7] V. Lepetit, P. Fua, et al. Monocular model-based 3d tracking of rigid objects: A survey. *Foundations and Trends® in Computer Graphics and Vision*, 1(1):1–89, 2005.

[8] K. R. Moser, S. Anreddy, and J. E. Swan. Calibration and interaction in optical see-through augmented reality using leap motion. In *Virtual Reality (VR), 2016 IEEE*, pages 332–332. IEEE, 2016.

[9] K. R. Moser and J. E. Swan. [poster] improved spaam robustness through stereo calibration. In *Mixed and Augmented Reality (ISMAR), 2015 IEEE International Symposium on*, pages 200–201. IEEE, 2015.

[10] A. Plopski, J. Orlosky, Y. Itoh, C. Nitschke, K. Kiyokawa, and G. Klinker. Automated spatial calibration of hmd systems with unconstrained eye-cameras. In *Mixed and Augmented Reality (ISMAR), 2016 IEEE International Symposium on*, pages 94–99. IEEE, 2016.

[11] L. Qian, A. Winkler, B. Fuerst, P. Kazanzides, and N. Navab. Modeling physical structure as additional constraints for stereoscopic optical see-through head-mounted display calibration. In *Mixed and Augmented Reality (ISMAR-Adjunct), 2016 IEEE International Symposium on*, pages 154–155. IEEE, 2016.

[12] J. R. Rambach, A. Tewari, A. Pagani, and D. Stricker. Learning to fuse: A deep learning approach to visual-inertial camera pose estimation. In *Mixed and Augmented Reality (ISMAR), 2016 IEEE International Symposium on*, pages 71–76. IEEE, 2016.

[13] Seiko Epson Corporation. *MOVERIO Pro BT-2000 User's Guide*, 2015.

[14] N. Snavely, S. M. Seitz, and R. Szeliski. Photo tourism: exploring photo collections in 3d. In *ACM transactions on graphics (TOG)*, volume 25, pages 835–846. ACM, 2006.

[15] A. Tang, J. Zhou, and C. Owen. Evaluation of calibration procedures for optical see-through head-mounted displays. In *Mixed and Augmented Reality, 2003. Proceedings. The Second IEEE and ACM International Symposium on*, pages 161–168. IEEE, 2003.

[16] M. Tuceryan, Y. Genc, and N. Navab. Single-point active alignment method (spaam) for optical see-through hmd calibration for augmented reality. *Presence: Teleoperators and Virtual Environments*, 11(3):259–276, 2002.

[17] T. Weise, T. Wismer, B. Leibe, and L. Van Gool. In-hand scanning with online loop closure. In *Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on*, pages 1630–1637. IEEE, 2009.

[18] F. Wientapper, T. Engelke, J. Keil, H. Wuest, and J. Mensik. [demo] user friedly calibration and tracking for optical stereo see-through augmented reality. In *Mixed and Augmented Reality (ISMAR), 2014 IEEE International Symposium on*, pages 385–386. IEEE, 2014.

[19] F. Wientapper, H. Wuest, P. Rojtberg, and D. Fellner. A camera-based calibration for automotive augmented reality head-up-displays. In *Mixed and Augmented Reality (ISMAR), 2013 IEEE International Symposium on*, pages 189–197. IEEE, 2013.

[20] Z. Zhang. A flexible new technique for camera calibration. *IEEE Transactions on pattern analysis and machine intelligence*, 22(11):1330–1334, 2000.

[21] F. Zhou, H. B.-L. Duh, and M. Billinghurst. Trends in augmented reality tracking, interaction and display: A review of ten years of ismar. In *Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 193–202. IEEE Computer Society, 2008.