Vision, Modeling, and Visualization (2018)
F. Beck, C. Dachsbacher, and F. Sadlo (Eds.)

WithTeeth: Denture Preview in Augmented Reality

Aleksandr Amirkhanov[1], Artem Amirkhanov[1], Matthias Bernhard[2], Zsolt Toth[2],
Sabine Stiller[2], Andreas Geier[3], Eduard Gröller[1], and Gabriel Mistelbauer[4]

[1]TU Wien, Austria
[2]cool IT GmbH, Austria
[3]Denttec KG, Italy
[4]Otto-von-Guericke University Magdeburg, Germany

Figure 1: *Virtual full denture previews in augmented reality. Each pair of images consists of the original image (left) and the virtual dentures preview (right). Our proposed system works in real-time, mimicking a dental virtual mirror.*

Abstract

Dentures are prosthetic devices replacing missing or damaged teeth, often used for dental reconstruction. Dental reconstruction improves the functional state and aesthetic appearance of teeth. State-of-the-art methods used by dental technicians typically do not include the aesthetic analysis, which often leads to unsatisfactory results for patients. In this paper, we present a virtual mirror approach for a dental treatment preview in augmented reality. Different denture presets are visually evaluated and compared by switching them on the fly. Our main goals are to provide a virtual dental treatment preview to facilitate early feedback, and hence to build the confidence and trust of patients in the outcome. The workflow of our algorithm is as follows. First, the face is detected and 2D facial landmarks are extracted. Then, 3D pose estimation of upper and lower jaws is performed and high-quality 3D models of the upper and lower dentures are fitted. The fitting uses the occlusal plane angle as determined mnually by dental technicians. To provide a realistic impression of the virtual teeth, the dentures are rendered with motion blur. We demonstrate the robustness and visual quality of our approach by comparing the results of a webcam to a DSLR camera under natural, as well as controlled lighting conditions. .

CCS Concepts

• ***Computing methodologies*** $\rightarrow$ *Mixed / augmented reality;* • ***Applied computing*** $\rightarrow$ *Consumer health;*

1. Introduction

Dental health care is increasingly using digital technologies in order to provide cost-efficient procedures for the design of personalized treatments with high quality and predictable results. A common pipeline, which has been established for computer-aided dental health care, involves three stages: scanning, computer-aided design (CAD), and computer-aided manufacturing (CAM). During scanning, the jaw of a patient is captured by taking 3D surface scans, either directly by inter-oral scanning, or indirectly from a plaster cast. After capturing the dentition of a patient in a 3D model, dental solutions, comprising bridges, crowns, dentures, healing abutments, or braces, are designed with CAD tools. These tools allow dentists to accurately adjust existing, replaced or treated teeth, as well as other inter-oral parts such as the alveolar bone of the patient. Finally, the CAD result is produced by CAM that supports using various materials for dentistry, such as composite resin, ceramic, or metal alloys. While subtractive manufacturing through milling from a blank is the state-of-the-art approach, the development of additive manufacturing with 3D printers for digital dentistry is advancing.

Our work is concerned with the CAD of teeth replacement treatments, such as implants, partial, or full dentures. CAD tools offer a wide range of tools to assist a dentist or dental technician in designing replacement solutions. They replace traditional procedures of dental technicals that have previously been carried out with physical modeling materials such as acryl or plaster. Moreover, there are also digital versions of apparatuses for functional analysis, like the *Virtual Articulator* [BBV02], which simulates traces of teeth movement. CAD tools, like the *Virtual Articulator*, are used to predict where teeth are potentially colliding, since they would wear off faster in such regions. Overall, there is an established set of tools assisting the CAD of functional dentures, implants, and other treatments. CAD software [3Sh, exo] show to the user by means of visualizations how well the designed objects fit to the jaw of the patient. However, these visualizations do not provide visual feedback on how the dental solutions fit from an aesthetic point of view. It is difficult to formalize teeth aesthetics because it concerns characteristics like beauty and taste, which are rather subjective.Thus, the patient opinion is valuable for the aesthetic analysis.

We define the high-level task of our work as the mapping of a virtual teeth model to an image of the patient's face, thus providing a highly realistic simulation not distinguishable from reality. Through these simulations, dental technicians can communicate with the patients to find a solution according to their taste and preferences. We propose *WithTeeth*, a tool that embeds virtual teeth into a video stream of the patient's face captured by a camera in real-time. A virtual teeth preview is shown in real-time and supports various face positions, orientations, and facial expressions. The main contributions of our work are:

- A real-time system that permits free head movements and allows the user to switch interactively between several different teeth replacement options.
- A mixed-reality solution that achieves compelling visualizations without depending on complex reconstruction algorithms, expensive hardware, or tedious calibration procedures.

2. Workflow and Task Analysis

Many factors contribute to the beauty of a smile, for example, tooth and gingiva colors, teeth shape, texture, transparency, and position. Simulating the physical properties of all teeth materials is a challenging task and involves technical as well as artistic work. In this work, we define the aesthetic model as a function depending on two factors: tooth color and teeth shape. In a Euclidean space, any rigid object can be defined by a position vector and an orientation matrix. In order to find a mapping for the virtual teeth into image space, we determine the position and orientation of the upper and lower dentures. Since the human teeth are rigidly attached to the maxilla and mandible, we specify our low-level tasks as follows:

- **(T1)** Finding position and orientation of the maxilla and the mandible to augment the input image with virtual teeth.
- **(T2)** Provide a rendering of virtual teeth, which allows patients to choose the teeth base color.

WithTeeth detects and tracks the face of the patient, automatically extracting facial landmark points. These points provide an estimate of the 3D pose of the patient's upper (maxilla) and lower (mandible) jaws, as well as the lips' contour lines. The data are then processed to seamlessly place the properly oriented virtual teeth at the correct position on the face of the patient (T1). The visibility of the virtual dentures inside the mouth is limited to the area inside the lips' inner contour when blending virtual and real images during rendering. *WithTeeth* allows the user to switch dentures of different color and shape on the fly to find the one best fitting to the face of the patient from an aesthetic point of view (T2).

Our virtual mirror can be implemented on consumer hardware consisting of a webcam and a personal computer with a screen (e.g., tablet computer). The implementation of the virtual dentures preview works in real-time. It is robust with respect to face orientations, shapes, and expressions. The patient can freely move inside the capturing area and switch between different alternative options of dental replacements.

WithTeeth enables dental technicians to communicate with patients through augmented reality (AR). Using the tool, domain experts show to the patients a simulated result of the dental treatment at early stages. For the patients, the tool allows an exploration of virtual teeth and to make a decision on the preferable teeth color and shape. Closing a gap in the CAD software currently available on the dental market, *WithTeeth* introduces a new workflow of the dental aesthetic analysis. The tool brings benefits to both dental technicians and patients. By presenting a denture design to the patients on an early stage, the dental technician can avoid cases where they are not satisfied with the results. Additionally, patients using the tool can increase their confidence in the treatment procedure by actively participating in the design process. The workflow opens new possibilities for remote medicine, minimizing travelling expenses. The dental technicians can send a new denture design to patients while they are at home to get their feedback, saving energy and time.

3. Related Work

Precise teeth detection and denture fitting has been rarely investigated, and the literature related to dental procedures is limited. Our work is related to medical applications that involve capturing facial landmarks and augmented reality (AR). We hereby review these two topics, providing examples of dental applications, if available.

Facial capturing. Capturing 3D facial models is an important research topic in many applications, e.g., game or movie industries, as well as the medical domain. Nowadays, it is possible to create photorealistic models of the human head using special facial scanners or photogrammetry. Facial paintings or markers can be used for capturing high-resolution static or animated facial models [Wil90, HCTW11, GGW*98]. These methods are often cumbersome to use and time-consuming as they require manual marker placement on the face. In our approach, it is important to have a very fast capturing algorithm, since we provide immediate visual dental feedback to the patients. Other methods use structured light and project patterns directly on the face. They do not require prior face markings [ZSCS04, WLVGP09, WHL*04]. How-

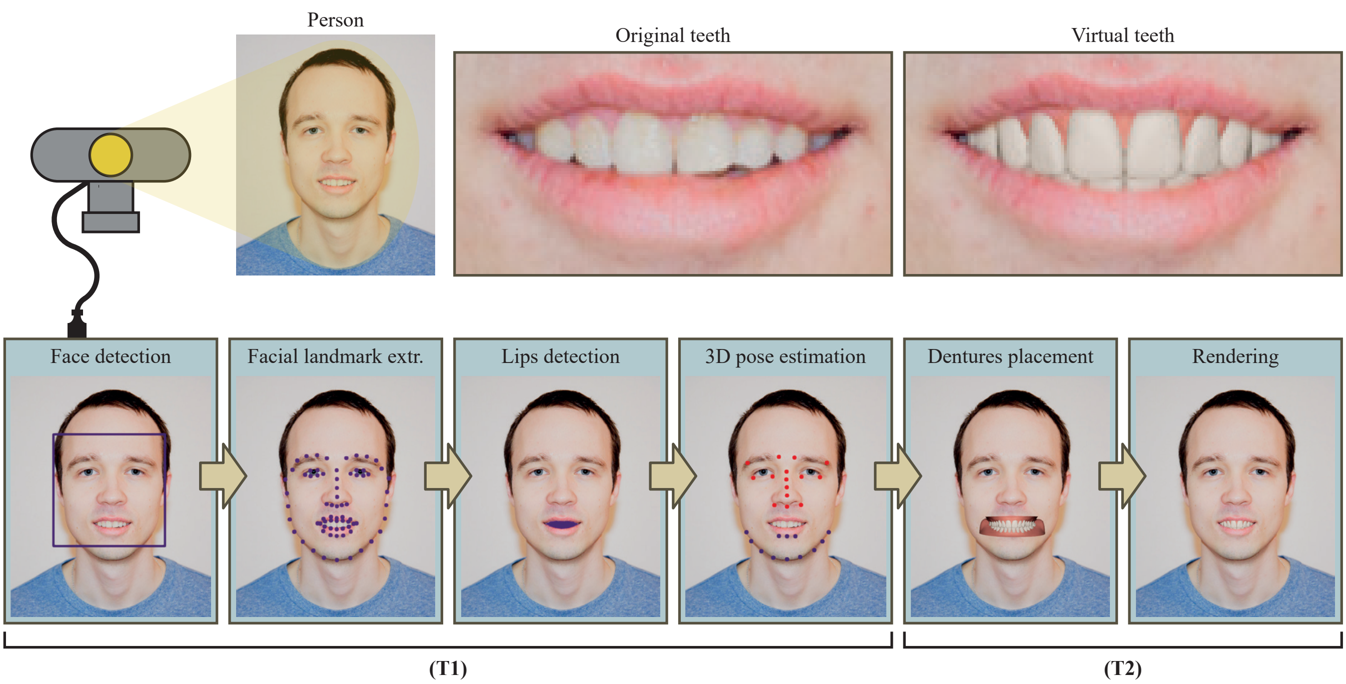


Figure 2: *WithTeeth pipeline, showing all the employed steps.*

ever, the skin texture is modified, which entails a number of limitations for photorealistic rendering applications. Furthermore, the methods reduce patient comfort as the light might be directly projected onto the face. During the last decade, depth cameras, such as Kinect or Intel RealSense, have become popular for facial capture [MGF*15, DMS12]. These depth cameras can provide an accurate head tracking and capturing. Most depth sensing solutions bundle infrared projectors and sensors with RGB video cameras. Such a strong coupling significantly limits flexibility. Built-in cameras often have limited capabilities and upgrading is not possible without replacing the entire assembly. Sophisticated parametric models and approaches have been designed, with a particular focus on the human body and the head. These 3D facial parametric models are often used to reconstruct the head from a photography or a video [CBZB15, CHZ14, PB16]. Richardson et al. presented a method to reconstruct the facial geometry for a singular image [RSK16]. The method uses a convolutional neural network trained on the synthetic dataset consisting of photo-realistic facial images with the known geometric form. Thies et al. [TZS*16] presented an approach to reenact a target video based on the facial expression of an actor, who is recorded live with a webcam. The methods mentioned above require a training set of photos and are sensitive to changes in light conditions. Nevertheless, the techniques are promising for real-time face capturing and tracking. Despite this, the topic of precise teeth detection and denture fitting has not been sufficiently investigated. Wu et al. [WBG*16] presented a teeth reconstruction technique that is based on a set of photographs or a video for detecting the shape and position of the teeth. Their approach requires a database of high-quality teeth for training and is sensitive to light conditions. In addition, their approach does not work for patients with only a few or no teeth, who usually are the target group of dental prostheses. Detecting the lips is another important aspect for providing a plausible virtual denture. Eveno et al. [ECC04] have presented an active contour algorithm for segmenting lips in 2D images. Their approach detects the boundary of the mouth and then fits a spline to the contour of the lips using a parametric model. Garrido et al. [GZW*16] presented an approach for 3D reconstructing the shape of the lips, based on monocular RGB videos. A comprehensive summarization and classification of the state of the art has recently been published by Zollhöfer et al. [ZTG*18].

Augmented reality (AR). Many AR applications use virtual mirror approaches, e.g., in commercial products [RTHS10] and anatomical education applications [MFB*13]. Chu et al. [CDWB10] presented an approach where users can try different types of jewelry. Rahman et al. [RTHS10] proposed a method that previews features of make-up, before using a cosmetic tool. In the field of anatomical education, Blum et al. [BKBN12] use a depth camera to align a volumetric dataset to the human body. Stefan et al. [SWO*14] presented a system for bone anatomy education. Users are presented with a puzzle consisting of bones to assemble, employing their own body as a reference frame. Moreover, Mercier-Ganady et al. [MGLLE*14] visualize brain activity in real-time. Their approach teaches users to control different mental states, such as concentration or relaxation. In the dental domain, Rhienmora et al. [RGH*10] discussed a dental training simulator in AR. This simulator includes a virtual mirror for indirect vision and utilizes haptic force feedback. For dental reconstruction, a commercial software is available from Kapanu AG [Kap]. The software enables previewing dental reconstructions in AR, however, there is no publication explaining the technique. Other medical AR applications include image-guided surgery, e.g., the work of Salah et al. [SPS*10] for brain tumor resection. An example of focus-and-context visualization in medical AR is the work of Kutter et al. [KAB*08], which inspects bone structures to support trauma surgery. Virtual mirrors have been employed in numerous previous works [BSHN07, BHFN09, NFB07, Bic10]. Illustrative visualization has been employed in the past in combination with AR, in order to show vascular structures around tumors [HWR*10].

4. Methodology

In this section we describe our method in detail. It consists of the following steps: face detection, facial landmark extraction, lips' inner contour detection, 3D pose estimation of maxilla and mandible, dentures placement, and finally rendering. An overview of our approach is depicted in Figure 2. In the face detection step, the face region is identified in the image. Then, the facial landmarks are extracted in the face region. Using these facial landmarks the inner contour of the lips is detected as the virtual dentures have to be cropped there. Then, the 3D poses of the maxilla and mandible are estimated from the facial landmarks, in order to determine the positions of the lower and upper dentures. In the last step, the original teeth are replaced with the virtual dentures, resulting in a realistic dental preview in real-time.

4.1. Input

Our approach requires certain equipment and environmental settings. The goal of our work is to build a robust system for previewing dentures that require only consumer hardware, such as a computer and a webcam. We are targeting low-cost webcams, however, our system can work with more expensive and better quality cameras as well. For the facial capture, the recording environment is very important to obtain a good performance. We recommend using environments with bright and indirect light. Bright light allows the recording device to use a low ISO sensitivity, which results in a better image quality. In cases where indirect light is not available, the light source should illuminate the face from the camera direction. The face of the patient should be homogeneously illuminated without any hard shadows. A good contrast between the face of the user and the background is recommended, while the face should not be occluded by any other objects.

Our approach works for videos as well as for single images. For videos, an additional step is required to filter the noise in the input video stream produced by the camera sensor. To reduce noise, we apply a Kalman filter [Kal60] that predicts the actual value based on a set of observations. For simplicity, we subsequently explain the steps of our pipeline for a single image, except at places where we are talking about the Kalman filter.

4.2. Face Detection

Face detection is the first step of the *WithTeeth* pipeline. In the input image, all human faces, which are present in the field of view of the camera, are detected using a histogram of oriented gradients [DT05, FGMR10]. The face detector module searches for faces

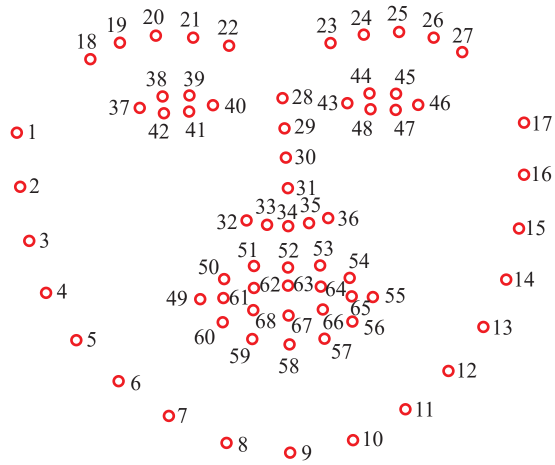


Figure 3: *The facial landmark extraction provides 68 facial points. The outer contour of the lips has more points than the inner contour. We use a heuristic to predict points on the inner lip contour using information from the outer lip contour.*

that are oriented towards the camera. Empirically, the angle between the face and the camera direction should be less than 45°. Otherwise, the face will not be captured appropriately, as shown in Figure 9. Due to performance reasons, the input frame is downsampled by the factor of 5 for the face detection. The frame itself remains untouched for the next steps. The downsampling factor depends on the resolution of the camera and how far the user is located from the device. We empirically found that for a webcam with a resolution of 1920×1080 pixels and a horizontal field of view of 70.5°, a downsampling factor of 5 is sufficient to detect a face at a distance up to one meter from the capturing device.

4.3. Facial Landmark Extraction

In the second step, landmarks are extracted from the facial region. We use a 68-point facial model for the landmark detection (see Figure 3), which has been trained on the ibug 300-W dataset [SAT*16]. Due to noise in the input image and limitations of the facial landmark detection technique, the points include some error. To reduce noise in the points between the video frames, a Kalman filter is applied [Kal60]. It reduces the noise in the measurements of the landmarks and predicts their actual values. For each landmark point in a frame, the Kalman filter is applied separately. The filter has two main parameters: the process-noise covariance and the measurement-noise covariance. These parameters are adjusted for each specific camera and the recording environment. For the webcam that is used for all our results, we set the process-noise covariance to two and the measurement-noise covariance to four. After the facial landmarks have been extracted, we detect the lips to determine the mouth region.

4.4. Lips Detection

The facial landmark extraction provides 68 points distributed over the face including the region of the lips (see Figure 3). We found

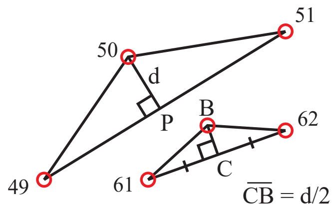


Figure 4: *Example of our heuristic to increase the number of points on the lips' inner contour. A point between points 61 and 62 is added, based on the positions of points 49, 50, and 51.*

that the number of points on the inner lip contour is insufficient to smoothly replace the original teeth with the virtually rendered dentures. Especially points at the corners of the lips are missing, where the curvature is often the highest. To address this, we developed a heuristic to increase the number of points on the inner contour of the lips. We also observed that the inner contour has a similar shape as the outer contour of the lips. Using this characteristic, we add four additional points that are located between points 61 and 62, 64 and 65, 65 and 66, 68 and 61 (see Figure 4). The subsequent steps determine a point between points 61 and 62:

$$\mathbf{a} = 51 - 49, \qquad \mathbf{b} = 50 - 49, \tag{1}$$

$$\mathbf{c} = 62 - 61, \qquad \mathbf{n} = \begin{pmatrix} -\mathbf{c}.y & \mathbf{c}.x \end{pmatrix}^{\mathsf{T}}, \tag{2}$$

$$P = 49 + \frac{\langle \mathbf{a}, \mathbf{b} \rangle}{|\mathbf{a}|} \cdot \hat{\mathbf{a}}, \tag{3}$$

$$C = 61 + 0.5 \cdot \mathbf{c}, \tag{4}$$

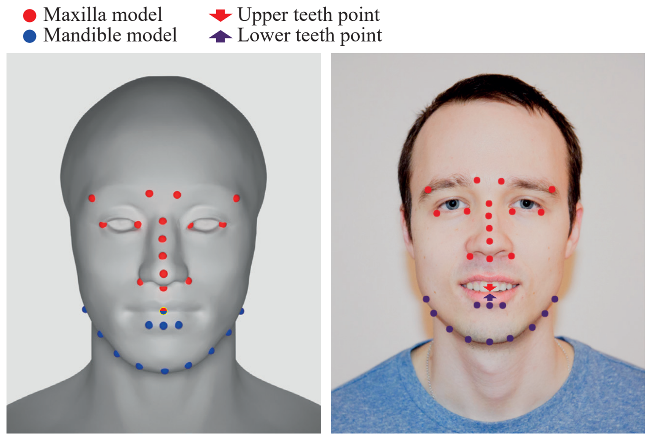
$$B = C + 0.5 \cdot \hat{\mathbf{n}} \cdot \|50 - P\|, \tag{5}$$

with $\hat{\mathbf{v}}$ being the unit vector of $\mathbf{v}$. We proceed analogously for the three other pairs of points.

After the reference points of the lips' inner contour are detected, we calculate the mask of all the pixels inside the polygon formed by the points. Then, we apply a morphological erosion and a normalized box filter, with kernel size of 5×5 pixels, to the mask. The erosion prevents the virtual dentures from leaking into the area of the lips. The Gaussian blur softens the transition from the real to the virtual region, thus making it less distractive.

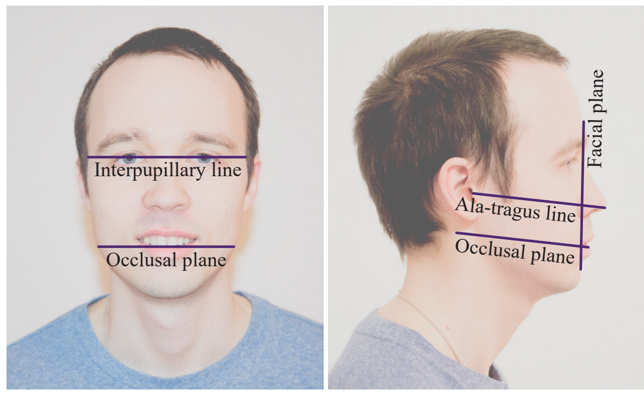
4.5. Pose Estimation

The pose estimation of a 3D object in a 2D image is done by finding the 3D-2D point correspondence. To create 3D models of the maxilla and mandible, we use the head mesh from the BodyParts3D dataset [MFT*09] (see Figure 5a). The maxilla model consists of 15 points that are located on the nose, eye, and eyebrows. The mandible model consists of 12 points that are placed on the chin and the lower lips. To estimate the pose of the maxilla, we use an iterative method based on Levenberg-Marquardt optimization [Lev44] by fitting the 3D model to the facial landmarks detected in the previous steps. The method minimizes the reprojection error between the points of the 3D model and the facial landmarks. It provides a transformation matrix and a translation vector of points from the model coordinate system to the world coordinate system. To estimate the pose of the mandible we perform a similar procedure as



(a) (b)

Figure 5: *We have developed two 3D models of the human head. The red points indicate the model of the maxilla and the blue points indicate the model of the mandible. For the maxilla we find more stable points that are almost motionless during facial emotion or smile. For the mandible, points are placed on more dynamic regions. Therefore, the quality of the pose estimation for the mandible is worse than for the maxilla.*



(a) (b)

Figure 6: *The occlusal plane is parallel to the interpupillary line (a), and to the Ala-Tragus line (b).*

with the maxilla. As only difference we use the extrinsic guess and provide to the method the rotation matrix and translation vector as approximate initial position of the mandible. We assume that the mandible has a similar position and orientation as the maxilla, which applies to most of the cases.

4.6. Dentures Placement

We determine the occlusal plane to properly fit the virtual teeth to the original ones (see Figure 6). The functional occlusal plane goes through the molar and premolars and indicates the relative positions of the teeth with respect to the head. The functional occlusal plane is always parallel to the interpupillary and Ala-Tragus lines. The interpupillary line connects the centers of the eyes' pupils (see Figure 6a). The Ala-Tragus line connects the wing of the nose with the

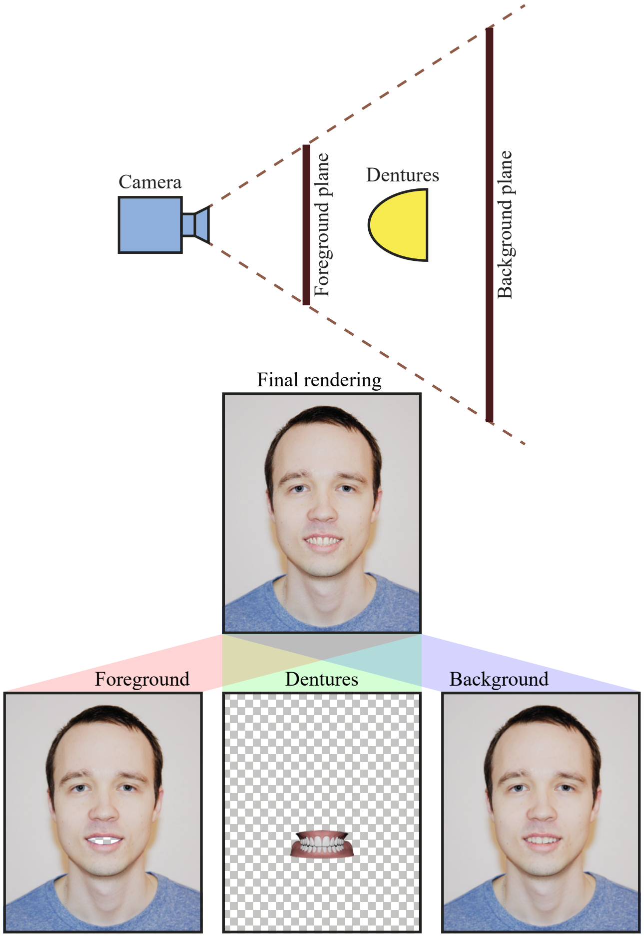


Figure 7: *Rendering the 3D scene involves the camera, the foreground plane, virtual dentures and the background plane. Intrinsic camera parameters should match the parameters of the real-word capturing camera (field of view, aspect ratio). The virtual dentures are positioned, at the screen position of the real teeth. The distances of the foreground and background planes to the camera are defined so that the virtual teeth are always in between these two planes.*

small pointed eminence of the outer ear (see Figure 6b). Capturing the interpupillary line is not necessary for our method. We assume that the previous 3D pose estimation of the maxilla provides us with a transformation matrix that is already aligned to the interpupillary line. To find the Ala-Tragus line, the user has to place four points on an image of the patient's head, which is shown in a profile view. These points are at the following locations: chin, forehead, wing of the nose, and pointed eminence of the ear. The line defined by the first two points is referred to as *facial line*. It depicts the orientation of the head in space. The line defined by the third and fourth points indicates the Ala-Tragus line. Afterwards, the angle between the occlusal plane and the facial plane is measured and used to place the dentures, as subsequently described.

After determining the transformation matrices and rotation vectors for the maxilla and mandible, we verify the mandible position.

Figure 8: *Example of dentures rendering.*

Typically, the mandible transformation is noisier and has a larger error than the one of the maxilla. This is because the mandible is the moving part and the maxilla is attached to the skull. The point where to find the virtual teeth is defined within the facial 3D models (see Figure 5). In some cases, this point has to be adjusted to match the anatomy of the person. We provide a method to shift the point from its original position in order to achieve a good fitting. In addition, we check that the mandible is at about the same distance to the camera as the maxilla. We apply the initial rotation to each denture in order to match the occlusal plane. Subsequently, we apply transformations of dentures to the world coordinate system. In the next step of the *WithTeeth* pipeline, we realistically render the dentures to provide a natural and aesthetic impression.

4.7. Rendering

The composition of the denture rendering is shown in Figure 7. It consists of textured foreground and background planes, a camera, and the virtual teeth placed in between these two planes. The camera looks through the foreground plane towards the background plane. The foreground plane presents the texture of the patient's face with the mouth region transparently rendered in order to see through the mouth. The teeth are placed so that the camera can see them in the mouth region. The virtual teeth are placed in the same way on the screen as the original teeth. The background plane presents the texture of the patient's face as well as the foreground plane, but it does not remove the mouth region. As a consequence, the tongue remains visible in the result image. The light source is located close to the camera and casts light on the teeth. It is important to match the light color with the real light. In our tool, it is possible to adjust the intensity of the light as well as the color temperature. In this way, we can adjust the virtual teeth to match the real ones in appearance.

5. Implementation

The face detection is implemented using the *dlib* library [Kin09]. *OpenCV* [Bra00] is used for pose estimation and the *Unreal Engine* is taken for rendering the virtual dentures [Unr]. *dlib* and *OpenCV* offer state-of-the-art algorithms for face detection and facial landmark detection. The *Unreal Engine* enables fast prototyping for rendering and works as the binding framework for all the involved libraries. Moreover, *dlib*, *OpenCV*, and *Unreal Engine* allow us to compile the application on many platforms such as Windows, macOS, Linux, iOS, and Android. Being able to execute the application on multiple platforms is a crucial requirement. It allows the dental technicians to run the application on their workstations and the patients to preview the simulation results on portable and affordable devices like mobile phones or tablet computers. There are other libraries for face detection and facial landmark extraction such as Apple Vision or ARCore or ARToolKit. However, some of them are platform-dependent, for example, Apple Vision supports only macOS and iOS; and ARCore supports only Android.

We have created two different materials in *Unreal Engine* for the teeth and the gingiva. Both materials use the subsurface shading model [Sub] implemented in Unreal Engine. The material of the gingiva has the following parameters: Base color (R, G, B) = (0.694, 0.251, 0.243); Subsurface color (R, G, B) = (0.694, 0.251, 0.243); Roughness = 0.23; Specular = 1; Subsurface amount = 0.2. The material of the teeth has the following parameters: Base color (R, G, B) = (0.847, 0.426, 0.066); Subsurface color (R, G, B) = (1, 1, 1); Roughness = 0.13; Specular = 1; Subsurface amount = 0.19. An exemplar rendering of virtual dentures, using our proposed approach, is presented in Figure 8. To provide a realistic impression of the teeth during movement, we use motion blur. Ambient occlusion is also used to give the teeth a more volumetric and natural appearance.

6. Results

In this section, we summarize the most significant results of our approach, illustrating scenarios of application and demonstrating its functionality, and performance.

Figure 1 shows three participants with their original teeth (left) and with virtual dentures (right). Pose estimation works for a large range of head positions (arbitrary yaw, pitch, and roll). Our approach works for participants with glasses (see second and third columns) and with different facial shapes and sizes.

Figure 9 illustrates the performance of our approach on different facial orientations (rows). Each step of the *WithTeeth* pipeline is displayed in a separate image (columns). In the **first** row the face is oriented towards the camera. The face detection method finds the facial region in the correct place and the facial landmarks are precisely extracted. The mouth region is detected with sufficient quality. The borders of the lips are not exact. However, no real teeth are visible in the final result. Markers of the pose estimation approach are placed at the correct positions and close to facial landmarks. The rendered dentures look natural like real teeth. The **second** row demonstrates the results if the face is slightly turned away from the camera. Each step is performed without visible deviations in quality from the first row.

In the **third and fourth** rows no quality degradation is visible in the images as well. In the **fifth** row, the accuracy of the inner contour detection of the lips is starting to degrade. Since some key points of the 3D maxilla and mandible models are occluded, the pose estimation step is not performing so well, resulting in a significant decrease in quality of the denture placement step. This is visible in the result image where the dentures are displaced from the original teeth position. The results are becoming even worse in the **sixth** row, where the head is turned further away from the camera. The lips' inner region is becoming less accurate. It includes part of the right cheek and more teeth are becoming visible from the left side. The lips detection is failing due to the imprecise facial landmark extraction. In the image with the facial landmarks (see column 3, row 6 in Figure 9), the feature points around the right corner of the mouth are imprecise. Although the region leaks into the right cheek, this region looks fine in the final rendered image. The reason is that the inner region of the lips is used only to occlude the virtual dentures. Also, locations like the right corner of the eyebrow or the right corner of the right eye are not placed exactly, since their correct positions are occluded. As a consequence, they are placed on the edge of the face. Imprecise landmark detection leads to an error in the pose estimation and, therefore, to an erroneous denture placement.

As the first four rows in Figure 9 present good results we conclude that our approach works well if the head of the user has not an extreme angle away from the camera. Mostly, this is due to the facial landmark detection step that yields worse results if the head is turned further away from the camera. We present our tool as a virtual mirror for previewing dentures. Since it is not possible to see oneself in the mirror from extreme viewing directions, our virtual mirror provides high-quality previewing.

7. Limitations and Future Work

The major contribution of this work is the development of a pipeline for pre-treatment dental previewing that can produce realistic results in real-time. The pipeline consists of multiple steps, and each of them can be individually improved. In this section, we discuss limitations of the pipeline steps and outline possible improvements for future work.

Facial landmark detection. The current module of facial landmark detection was trained on frontal faces and, therefore, does not work well if the head is turned far away from the camera. A solution to this limitation might be training another module for finding facial landmarks for such head positions or extending the current training database with more annotated images of the head in such positions.

Also, facial landmarks are prone to noise in the input image and they are detected at slightly different positions for each frame in the video. This noise is propagated to subsequent steps and causes jittering of the augmented dentures. To reduce the noise in the input signal we use a Kalman filter that predicts the correct position of landmarks based on previous observations. This filter provides good results if the face undergoes just small movements in the frame or the motion is predictable, e.g., the face is moving on a line. However, if the face movements are too fast and random, the error of fitting dentures is increased. For example, if a fast motion just starts, the dentures will be moving slightly slower than the head and they will be close to their original position. The same happens when the fast movement stops. The Kalman filter continues predicting the movement and dentures will sway a bit further along the movement trajectory. This limitation does not allow patients

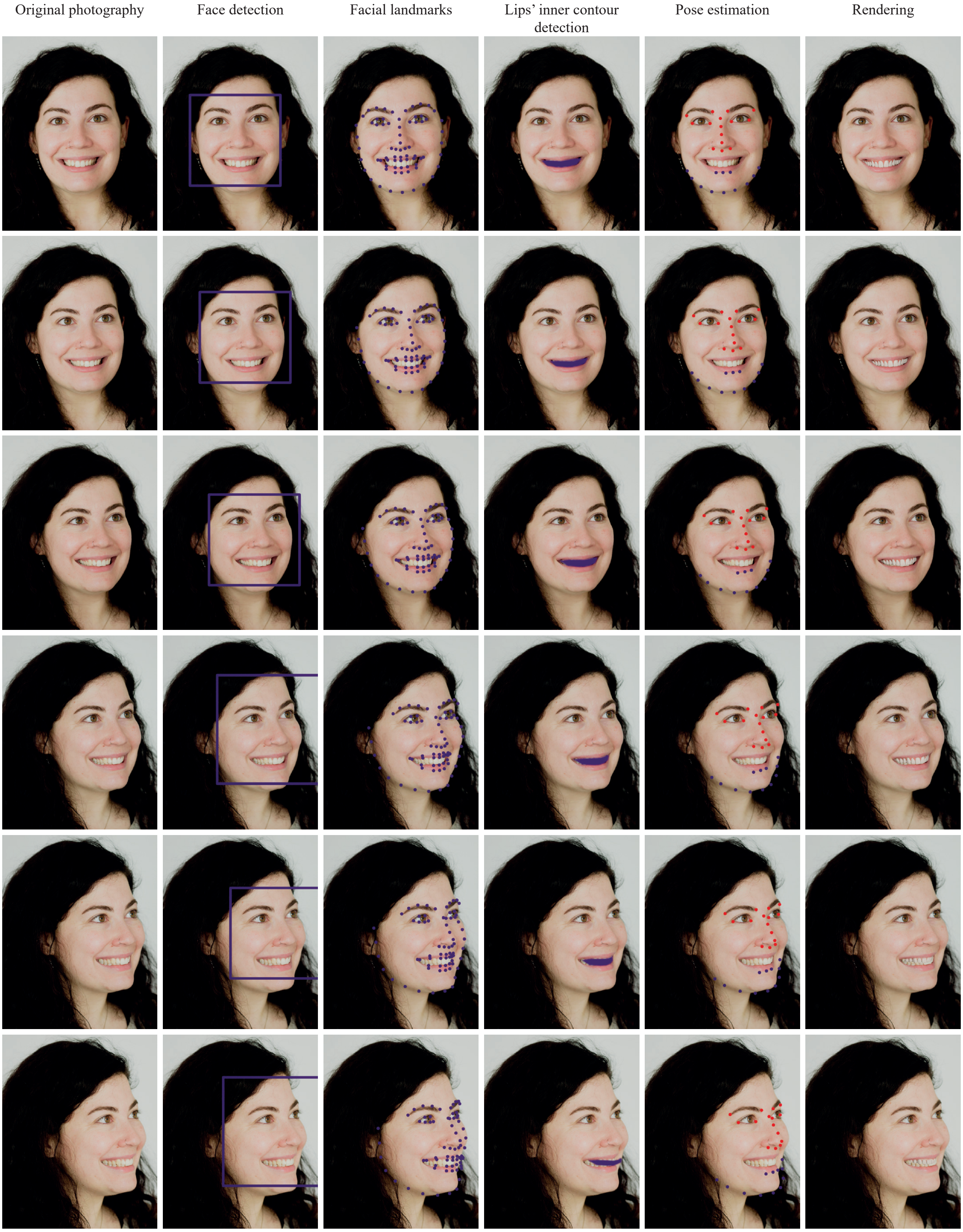


Figure 9: *WithTeeth pipeline for different face poses for the same person.*

to fully immerse into the mixed reality scene. In most of the cases, though, patients do not move fast and often explore the virtual teeth using rather static positions.

Lip detection. The current lip detection method is based on the facial landmarks. In some extreme facial expressions, the landmarks are not detected correctly, e.g., if the mouth is wide open. The facial landmark detector is based on a machine learning approach and strongly depends on the training set. To avoid this drawback, more sophisticated algorithms can be used such as, for example, boosted edge learning.

Dentures placements. The placement of the dentures depends on the pose estimation of the upper skull and the chin. We developed 3D models to find the positions using a publicly available dataset of the human body. The pose estimation performance can be improved, e.g., by personalizing the models. For example, the 3D face of the user can be captured using a photogrammetry approach or a depth camera. Then, the facial landmark points will be searched and personalized on the head of the user mesh. This will potentially improve the dentures placement step. It might require additional equipment or steps in the pipeline. In the current approach, we use a Levenberg-Marquardt optimization that fits the facial mask consisting of a set of 3D points to 2D facial landmarks. The pose estimation can be further improved using a 3D-to-3D optimization, requiring depth information of the facial landmarks that can be obtained using, for example, a depth camera.

In some cases, only a single tooth of the patient has to be treated or replaced. The current approach does not support to preview the dental treatment outcome for a single tooth. The individual tooth can be occluded by other teeth, thus the tooth visibility or the visibility mask has to be determined for a correct fitting. This can be calculated by using an individual teeth extraction based on the head position and a priori knowledge of a typical tooth layout. Additionally, the teeth visibility mask will help to hide patient's teeth if they are shorter than the dentures.

During testing, we found that in some cases it is useful to zoom in to the teeth to see small details. In the current implementation this is not possible. If the camera is placed close to the mouth of the patient, other important parts of the face, such as eyes or chin, are outside of the camera view. Thus, some facial landmarks cannot be used for detecting the face orientation. This limitation can be overcome, by adding more landmarks to the mouth region.

Rendering. The current rendering approach can be improved by measuring physical parameters, e.g., the bidirectional reflectance distribution function, of real teeth and adjusting rendering settings accordingly. To make the lighting conditions realistic, environmental maps can be used. We consider a precise color previewing of dentures as future work, because it requires all the rendering improvements mentioned above. In addtion, precise color previewing requires a calibrated camera or a recording environment with controlled light. Also, shadows from lips can be implemented and ambient occlusion can be taken into account more accurately.

An informal interview with domain experts revealed the need of dental technicians for simulating skin deformations caused by wearing dentures of different sizes. The dental state, shape, and size of teeth have a big influence on the shape of the upper and lower lips. For example, if teeth are missing, the skin will refract towards the mouth of the patient. Or vice versa, if the teeth are located too far in front of the mouth the lips will be bigger and more visible. The skin deformation has a big influence on the patient appearance and has to be taken into account in demonstrating the outcome of dental treatment. We consider a realistic model of skin deformation as an interesting direction of future work.

8. Conclusions

We presented *WithTeeth*, a virtual mirror approach, targeting the dental industry. Using our method, dental technicians can demonstrate possible treatment outcomes in advance to the patients to get their feedback. Patients are involved in the design procedure that might increase their confidence and satisfaction in the outcome. Virtual dentures can be exchanged on the fly and various characteristics can be controlled such as tooth color and size. Our tool works in real-time and requires only basic equipment comprising a computer with display, and a consumer-level webcam or a DSLR. Also, our approach does not require a priori training. The lighting parameters must be adjusted manually to match the real lighting of the environment. We described our method and provided implementation details where necessary. We discussed limitations and suggested directions for further research. Our approach, following modern trends of the dental industry, shows how digital technologies may reform and improve the dental reconstruction.

Acknowledgments

This work is part of the Visual Analytics for Realistic and Aesthetic Smile Design (SmileAnalytics) project, supported by the Austrian Research Promotion Agency (FFG) project no. 861168.

References

[3Sh] 3shape. https://www.3shape.com/de-de. Accessed: 2018-09-07. 2

[BBV02] Bisler A., Bockholt U., Voss G.: The virtual articulator – applying VR technologies to dentistry. In *Proceedings Sixth International Conference on Information Visualisation* (2002), pp. 600–602. 2

[BHFN09] Bichlmeier C., Heining S. M., Feuerstein M., Navab N.: The virtual mirror: A new interaction paradigm for augmented reality environments. *IEEE Transactions on Medical Imaging 28*, 9 (2009), 1498–1510. 4

[Bic10] Bichlmeier C.: *Immersive, interactive and contextual in-situ visualization for medical applications*. PhD thesis, Technische Universität München, 2010. 4

[BKBN12] Blum T., Kleeberger V., Bichlmeier C., Navab N.: mirracle: An augmented reality magic mirror system for anatomy education. In *Proceedings of IEEE Virtual Reality Workshops* (2012), pp. 115–116. 3

[Bra00] Bradski G.: The OpenCV Library. *Dr. Dobb's Journal of Software Tools* (2000). 6

[BSHN07] Bichlmeier C., Sielhorst T., Heining S. M., Navab N.: Improving depth perception in medical AR. In *Proceedings of Bildverarbeitung für die Medizin* (2007), Horsch A., Deserno T. M., Handels H., Meinzer H.-P., Tolxdorff T., (Eds.), Springer Berlin Heidelberg, pp. 217–221. 4

[CBZB15] Cao C., Bradley D., Zhou K., Beeler T.: Real-time high-fidelity facial performance capture. *ACM Trans. Graph. 34*, 4 (2015), 46:1–46:9. 3

[CDWB10] Chu M., Dalal B., Walendowski A., Begole J.: Countertop responsive mirror: supporting physical retail shopping for sellers, buyers and companions. In *Proceedings of SIGCHI Conference on Human Factors in Computing Systems* (2010), pp. 2533–2542. 3

[CHZ14] Cao C., Hou Q., Zhou K.: Displaced dynamic expression regression for real-time facial tracking and animation. *ACM Trans. Graph. 33*, 4 (2014), 43:1–43:10. 3

[DMS12] Datta S. K., Morrow P., Scotney B.: *Facial Feature Extraction Using a 4D Stereo Camera System*. Springer Berlin Heidelberg, 2012, pp. 209–218. 3

[DT05] Dalal N., Triggs B.: Histograms of oriented gradients for human detection. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* (2005), vol. 1, pp. 886–893. 4

[ECC04] Eveno N., Caplier A., Coulon P. Y.: Accurate and quasi-automatic lip tracking. *IEEE Transactions on Circuits and Systems for Video Technology 14*, 5 (2004), 706–715. 3

[exo] exocad. https://exocad.com/. Accessed: 2018-09-07. 2

[FGMR10] Felzenszwalb P. F., Girshick R. B., McAllester D., Ramanan D.: Object detection with discriminatively trained part-based models. *IEEE Transactions on Pattern Analysis and Machine Intelligence 32*, 9 (2010), 1627–1645. 4

[GGW*98] Guenter B., Grimm C., Wood D., Malvar H., Pighin F.: Making faces. In *Proceedings of the Conference on Computer Graphics and Interactive Techniques* (1998), pp. 55–66. 2

[GZW*16] Garrido P., Zollhöfer M., Wu C., Bradley D., Pérez P., Beeler T., Theobalt C.: Corrective 3D reconstruction of lips from monocular video. *ACM Trans. Graph. 35*, 6 (2016), 219:1–219:11. 3

[HCTW11] Huang H., Chai J., Tong X., Wu H.-T.: Leveraging motion capture and 3d scanning for high-fidelity facial performance acquisition. *ACM Trans. Graph. 30*, 4 (2011), 74:1–74:10. 2

[HWR*10] Hansen C., Wieferich J., Ritter F., Rieder C., Peitgen H.-O.: Illustrative visualization of 3D planning models for augmented reality in liver surgery. *International Journal of Computer Assisted Radiology and Surgery 5*, 2 (2010), 133–141. 4

[KAB*08] Kutter O., Aichert A., Bichlmeier C., Traub J., Euler E., Navab N.: Real-time volume rendering for high quality visualization in augmented reality. In *Proceedings of Workshop on Augmented environments for Medical Imaging including Augmented Reality in Computer-aided Surgery* (2008). 4

[Kal60] Kalman R. E.: A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering 82*, Series D (1960), 35–45. 4

[Kap] Kapanu ag. http://www.kapanu.com/. Accessed: 2018-09-07. 3

[Kin09] King D. E.: Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research 10* (2009), 1755–1758. 6

[Lev44] Levenberg K.: A method for the solution of certain non-linear problems in least squares. *Quarterly of Applied Mathematics 2*, 2 (1944), 164–168. 5

[MFB*13] Meng M., Fallavollita P., Blum T., Eck U., Sandor C., Weidert S., Waschke J., Navab N.: Kinect for interactive AR anatomy learning. In *Proceedings of IEEE Symposium on Mixed and Augmented Reality* (2013), pp. 277–278. 3

[MFT*09] Mitsuhashi N., Fujieda K., Tamura T., Kawamoto S., Takagi T., Okubo K.: BodyParts3D: 3d structure database for anatomical concepts. *Nucleic Acids Research 37*, suppl 1 (2009), D782–D785. 5

[MGF*15] Meyer G. P., Gupta S., Frosio I., Reddy D., Kautz J.: Robust model-based 3d head pose estimation. In *Proceedings of IEEE Conference on Computer Vision* (2015), pp. 3649–3657. 3

[MGLLE*14] Mercier-Ganady J., Lotte F., Loup-Escande E., Marchal M., Lécuyer A.: The mind-mirror: See your brain in action in your head using EEG and augmented reality. In *Proceedings of IEEE Virtual Reality* (2014), pp. 33–38. 3

[NFB07] Navab N., Feuerstein M., Bichlmeier C.: Laparoscopic virtual mirror new interaction paradigm for monitor based augmented reality. In *Proceedings of IEEE Virtual Reality Conference* (2007), pp. 43–50. 4

[PB16] Piotraschke M., Blanz V.: Automated 3d face reconstruction from multiple images using quality measures. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* (2016), pp. 3418–3427. 3

[RGH*10] Rhienmora P., Gajananan K., Haddawy P., Dailey M. N., Suebnukarn S.: Augmented reality haptics system for dental surgical skills training. In *Proceedings of ACM Symposium on Virtual Reality Software and Technology* (2010), pp. 97–98. 3

[RSK16] Richardson E., Sela M., Kimmel R.: 3d face reconstruction by learning from synthetic data. *CoRR abs/1609.04387* (2016). 3

[RTHS10] Rahman A. S. M. M., Tran T. T., Hossain S. A., Saddik A. E.: Augmented rendering of makeup features in a smart interactive mirror system for decision support in cosmetic products selection. In *Proceedings of IEEE/ACM Symposium on Distributed Simulation and Real Time Applications* (2010), pp. 203–206. 3

[SAT*16] Sagonas C., Antonakos E., Tzimiropoulos G., Zafeiriou S., Pantic M.: 300 faces in-the-wild challenge: database and results. *Image and Vision Computing 47* (2016), 3–18. 4

[SPS*10] Salah Z., Preim B., Samii A., Fahlbusch R., Rose G.: Enhanced intraoperative visualization for brain surgery: A prototypic simulated scenario. In *Proceedings of CURAC* (2010), pp. 125–130. 4

[Sub] Subsurface profile shading model. https://docs.unrealengine.com/en-us/Engine/Rendering/Materials/LightingModels/SubSurfaceProfile. Accessed: 2018-09-07. 7

[SWO*14] Stefan P., Wucherer P., Oyamada Y., Ma M., Schoch A., Kanegae M., Shimizu N., Kodera T., Cahier S., Weigl M., Sugimoto M., Fallavollita P., Saito H., Navab N.: An AR edutainment system supporting bone anatomy learning. In *Proceedings of IEEE Virtual Reality* (2014), pp. 113–114. 3

[TZS*16] Thies J., Zollhöfer M., Stamminger M., Theobalt C., Niessner M.: Face2face: Real-time face capture and reenactment of rgb videos. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2016), pp. 2387–2395. 3

[Unr] Unreal engine. https://www.unrealengine.com/. Accessed: 2018-09-07. 6

[WBG*16] Wu C., Bradley D., Garrido P., Zollhöfer M., Theobalt C., Gross M., Beeler T.: Model-Based Teeth Reconstruction. *ACM Trans. Graph. 35*, 6 (2016), 220:1–220:13. 3

[WHL*04] Wang Y., Huang X., Lee C.-S., Zhang S., Li Z., Samaras D., Metaxas D., Elgammal A., Huang P.: High resolution acquisition, learning and transfer of dynamic 3-d facial expressions. *Computer Graphics Forum 23*, 3 (2004), 677–686. 2

[Wil90] Williams L.: Performance-driven facial animation. In *Proceedings of the Conference on Computer Graphics and Interactive Techniques* (1990), pp. 235–242. 2

[WLVGP09] Weise T., Li H., Van Gool L., Pauly M.: Face/off: Live facial puppetry. In *Proceedings of ACM SIGGRAPH/Eurographics Symposium on Computer Animation* (2009), pp. 7–16. 2

[ZSCS04] Zhang L., Snavely N., Curless B., Seitz S. M.: Spacetime faces: High resolution capture for modeling and animation. *ACM Trans. Graph. 23*, 3 (2004), 548–558. 2

[ZTG*18] Zollhöfer M., Thies J., Garrido P., Bradley D., Beeler T., Pérez P., Stamminger M., Niessner M., Theobalt C.: State of the Art on Monocular 3D Face Reconstruction, Tracking, and Applications. *Computer Graphics Forum (Eurographics State of the Art Reports 2018) 37*, 2 (2018). 3