Universiteit Utrecht

technicolor

UTRECHT UNIVERSITY
GRADUATE SCHOOL OF NATURAL SCIENCES
DEPARTMENT OF INFORMATION AND COMPUTING SCIENCES
MSC GAME AND MEDIA TECHNOLOGY

Automated Generation of Animated 3D Facial Meshes: A Photogrammetry and Deformation Transfer-Based Model

Ilja Gubins
ilja@gubins.lv
ICA-5663024

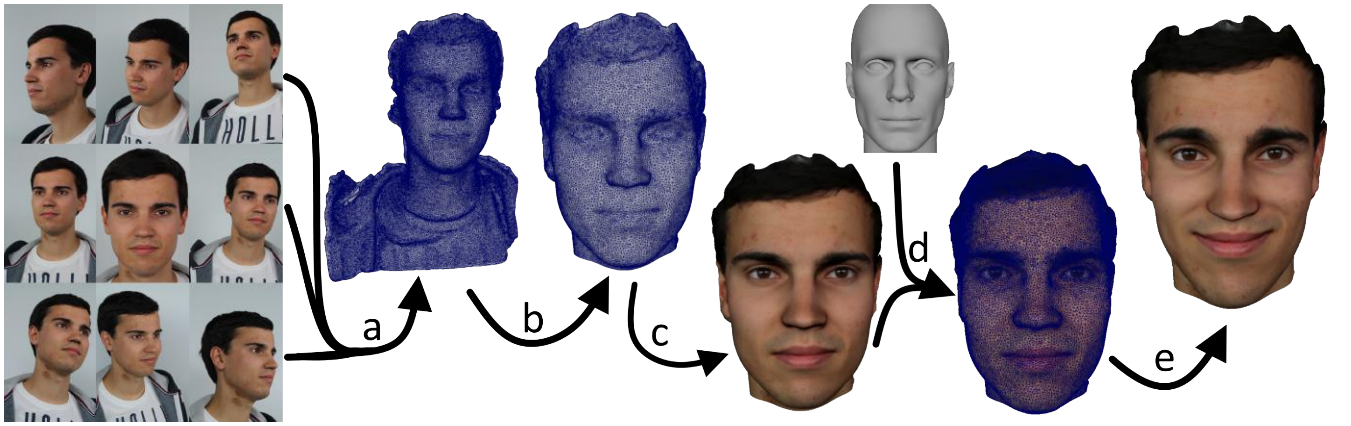
Utrecht, August 16, 2017

Project supervisor:
dr. dr. E. L. van den Broek
Department of Information and Computing Sciences
Utrecht University
e.l.vandenbroek@uu.nl

Daily supervisor:
dr. Quentin Avril
Immersive Lab
Technicolor R&D France
quentin.avril@technicolor.com

Second examiner:
prof. dr. R. C. Veltkamp
Department of Information and Computing Sciences
Utrecht University
r.c.veltkamp@uu.nl

Associate daily supervisor:
dr. Fabien Danieau
Immersive Lab
Technicolor R&D France
fabien.danieau@technicolor.com



Workflow overview: ***(a) Mesh acquisition.*** *We obtain multi-camera view of a person with a camera rig. Using the resulting photos, we first create dense point cloud, and then the triangle mesh.* ***(b) Mesh cleaning.*** *We simplify, clean, and compute UV-mapping.* ***(c) Mesh texturing.*** *We obtain the color of the surface by blending intensities from each photograph.* ***(d) Correspondence construction.*** *We calculate the correspondence between the acquired mesh and the existing fully animated template mesh by using automatic landmarking and ICP registration algorithm.* ***(e) Blendshape transfer.*** *Finally, we transfer blendshape animations from template to the newly created mesh based on the constructed correspondence.*

Abstract

Virtual reality (VR)'s recent, long expected breakthrough to the consumer market has enabled industry to deliver highly immersive experiences to consumers. However, creating photorealistic assets for such experiences is time consuming and requires a high level of expertise. To provide a remedy for this burden, this thesis presents a fully automated and validated model for the generation of animated 3D facial meshes, using photogrammetry and deformation transfer. Using a set of multi-camera photographs of a neutral face, we acquire facial geometry and appearance. Subsequent landmarking and ICP registration provide direct correspondence of the acquired facial mesh and an existing template. Then, using deformation transfer, we transfer existing blendshapes from template source to the obtained facial mesh. To assess the pipeline's effectiveness, we have conducted a set of user experiments. The results show that the automatically produced facial meshes faithfully represent subjects and can convey highly believable emotions. Compared to manual modeling and animation, the pipeline is more than 100x faster and produces results compatible with industry standard game engines, without any manual post-processing. As such, this thesis introduces a unique, fully automated, photogrammetry and deformation transfer-induced model for the generation of animated 3D facial meshes. On the one hand, this model can be expected to have both significant commercial and scientific impact. On the other hand, the model allows sufficient space for future improvement and tailoring.

Keywords: facial acquisition, deformation transfer, facial animation, photogrammetry, pipelines

1 Introduction

The use of 3D photogrammetry becomes increasingly more common for media assets production. Nowadays, movie production studios use photogrammetry for almost every production step, starting from capture of film set for previsualization and reference for artists [Zwerman and Okun 2012], to creation of digital doubles for starring actors and post-production re-lighting of shot scenes [Debevec 2012]. Film production industry already makes heavy use of photogrammetry, but game and media production studios photogrammetry adoption rate, in general, is slower. This is mostly because captured assets are abundantly high-poly and require tedious manual optimizations and post-processing to be run at real-time.

Recent technology advances make photogrammetry more accessible to create and more acceptable to use in real-time projects [Poznanski 2014] [Brown and Hamilton 2016]. At the same time, cutting edge real-time areas like virtual reality (VR) strives to provide immersive hyper-realistic experience. This makes photogrammetry a very favourable choice for static photorealistic VR assets. For dynamic assets, such as non-playable characters (NPC), requirements for achieving photorealism are significantly higher, manifesting as Uncanny Valley problem [Reichardt 1978] in case of human NPCs. One of the biggest contributors to the feeling of fakeness is the lack of emotional facial expressivity [Tinwell et al. 2014].

This paper makes a systemic contribution by establishing and describing a complete pipeline for the creation of a digital face model suitable for VR by using photogrammetry (Section 3) and automatic animation of this model by deformation transfer (Section 4). Photogrammetry allows us to achieve photorealistic quality, while transferred animations bring full spectre of emotional facial expressions. The paper also makes scientific contribution by conducting user experiments evaluating the proposed method (Section 5).

2 Related work

The focus research areas of this paper are 3D facial geometry capture and facial appearance acquisition (Section 2.1), and deformation transfer (Section 2.2). Both topics are active research areas in academia, and are actively used in industry, however no mention of combining them together have been found in the scientific literature.

2.1 3D facial geometry and appearance acquisition

The ultimate goal of facial acquisition is to be able to render the virtual human face, including non-trivial fine details, under arbitrary lighting and from any viewing position. The problem can be split into two parts: 3D facial geometry acquisition, and facial appearance acquisition.

The methods for 3D facial geometry capture developed in the last two decades can be split into active and passive systems:

- Active capture systems require special-purpose hardware, and extra constrains in setup. Such systems usually are based on laser, structured light, gradient-based illumination [Ma et al. 2007], or even requiring spatial multiplexing [Weyrich et al. 2006]. While the results they provide are often very robust, passive systems often are much more versatile and adaptive, allowing different arrangements of setup, numbers of camera and virtually no constrains on camera position, but sacrificing on reliability and accuracy [Beeler et al. 2010].
- Passive techniques have the advantage of non-intrusiveness and capture what is observed. Typically these methods require only a single frame to estimate the structure but usually provide less accurate results. Beeler et al. [Beeler et al. 2010] presented a passive stereo vision system that computes the 3D geometry of the face with reliability and accuracy on par with a laser scanner or a structured light system. In practical terms, that means equal to active systems performance while attaining the advantages of passive system. However, it makes assumption of constant omni-directional illumination, thus limiting it to studio environments. Later that method was extended to arbitrary lighting by estimating the environment map [Wu et al. 2011].

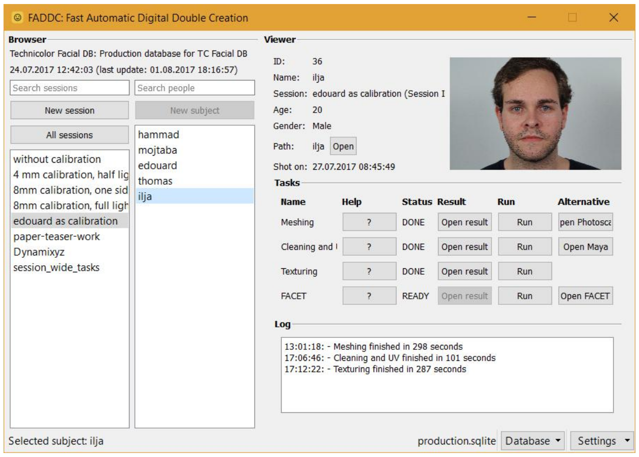
The second part of the problem is facial appearance acquisition: a way to record the way light interacts with skin. Appearance acquisition methods must cope with complexity of light interactions with skin. Two general categories of such methods are distinguished: image-based methods and parametric methods. Image-based methods exhaustively capture the exact face appearance under various lighting and viewing conditions, and then solve the rendering problem through weighted image combinations [Debevec et al. 2000] [Tunwattanapong et al. 2011] [Klehm et al. 2015]. Whereas the parametric methods instead aim at modeling the structure of the skin with suitable approximations. Such representation of the skin is more flexible but at the cost of a potentially inexact reproduction [Fuchs et al. 2005] [Ngan et al. 2005] [Ghosh et al. 2008] [Ghosh et al. 2010].

2.2 Facial animation by deformation transfer

Facial animation can be achieved by a large variety of different methods: skeletons and joints [Magnenat-Thalmann et al. 1988], physically-based muscle models [Waters 1987], linear blendshapes [Bergeron and Lachapelle 1985] and combinations thereof. While every method has a fitting application, using linear blendshape models is the most widely spread approach for high fidelity facial animation. Combining a set of blend shapes produces an arbitrary facial expression. However, the main challenge of using blendshape based model is the creation process of such blendshapes. Creating high quality blendshapes is highly time-consuming and tedious, requiring either high quality motion capture of real actor (and subsequent cleanup and post production) or manual modelling.

Blendshape models are especially convenient, as animations of such blendshapes can be transferred with deformation transfer [Sumner and Popović 2004]. Original research on deformation transfer had number of limitations. It was not contact-aware (addressed in [Sumner 2006]), did not preserve semantic characteristics of the motion (addressed in [Baran et al. 2009]), and did not preserve artist-given character-specific features like wrinkles. The latter was addressed in [Li et al. 2010] by providing a small number of example poses and blendshape coefficients, allowing the system to propagate character-specific features to transferred blendshapes.

All previously mentioned methods require triangle correspondences between source and target meshes, which is problematic if meshes have different topology. Pawaskar et al. [Pawaskar et al. 2013] proposed a technique to transfer blendshapes to a target mesh by first registering source mesh into target mesh using a non-rigid ICP (iterative closest point) algorithm, and then transferring deformation to a new target mesh that has direct triangle-wise correspondence.



FADDC: Fast Automatic Digital Double Creation

Technicolor Facial DB: Production database for TC Facial DB
24.07.2017 12:42:03 (last update: 01.08.2017 18:16:57)

without calibration
4 mm calibration, half lig
8mm calibration, one sid
8mm calibration, full ligh
edouard as calibration
paper-teaser-work
Dynamixyz
session_wide_tasks

hammad
mojtaba
edouard
thomas
ilja

ID: 36
Name: ilja
Session: edouard as calibration (Session I
Age: 20
Gender: Male
Path: ilja
Shot on: 27.07.2017 08:45:49

Name	Help	Status	Result	Run	Alternative
Meshing	?	DONE	Open result	Run	pen Photosca
Cleaning and I	?	DONE	Open result	Run	Open Maya
Texturing	?	DONE	Open result	Run	
FACET	?	READY	Open result	Run	Open FACET

```
13:01:18: - Meshing finished in 298 seconds
17:06:46: - Cleaning and UV finished in 101 seconds
17:12:22: - Texturing finished in 287 seconds
```

Selected subject: ilja
production.sqlite

Figure 2: *Screenshot of the developed pipeline automation application.*

3 3D facial acquisition

In this section we describe capture part of the pipeline, i.e. steps required to capture one's face geometry and appearance. We also describe our implementation details of the pipeline.

3.1 Guiding application

One of the main requirements for the pipeline is the ability to automate the entire process as much as possible and run the whole process without any human supervision. To prove that proposed pipeline fulfills the requirement, we have created a pipeline automation application (Figure 2). It was develop on Python with addition of Python bindings of Qt GUI framework[1] and a number of different frameworks. The application guides the workflow and handles data produced for each subject by running various various scripts for *Autodesk Maya* and Agisoft *Photoscan*.

We developed this application to assist in data gathering. The data is stored in session databases. Every database contains capture sessions. Every session has separate calibration, and every subject can be processed with tasks (individual pipeline steps). Every task gets some kind of input and produces some kind of result, allowing the system to be very flexible, and be easily modified and extended. Sessions database is represented with an *SQLite* database and a data folder, and can be easily transferred to another computer or stored on network location.

3.2 Camera setup

The starting point for the end-to-end pipeline established by this paper is camera setup. Acquiring 3D surfaces with photogrammetry is a very flexible method, allowing virtually any digital camera to be used, to the point where finding the optimal configuration is the main challenge. Fortunately, extensive research has been done

[1] Qt 4.8, The Qt Company, wiki.qt.io/About_Qt

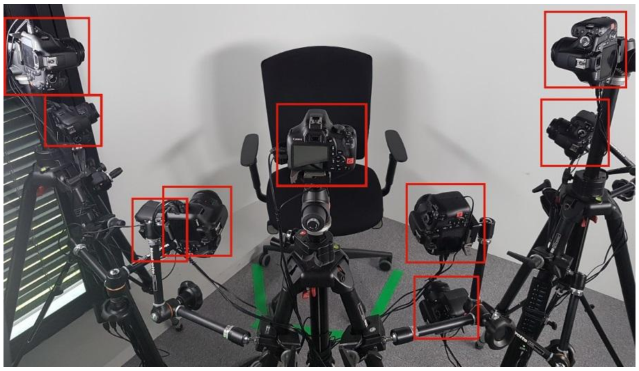


Figure 3: *Photograph of the camera setup. Red rectangles highlight the cameras.*

on defining guidelines for close range photogrammetry [Waldhäusl and Ogleby 1994] [Wenzel et al. 2013].

Following the mentioned guidelines, we have constructed the capture setup (Figure 3). The setup is the result of many iterations. The iterative approach for finding more optimal capture setup was based on comparing the quality of resulting meshes. In terms of hardware, our setup consists of 9 *Canon EOS 100D* DLSR cameras equipped with *Canon EF 50mm* prime (fixed focal length) lenses, and 2 *Kino Flo Tegra 455 DMX* lighting systems equipped with diffusion covers. The most fundamental principle of photogrammetry is the matching of features between photos to form a single contiguous model. To support such matching, a very strong overlap (70+%) is required [Waldhäusl and Ogleby 1994]. Also, it is important to note that even the slightest movements of the subjects can drastically reduce mesh quality. A simultaneous multiple camera trigger system must be used to prevent this.

3.3 Scene calibration

In the context of photogrammetry, calibration usually implies only camera calibration, the process of calculating intrinsic and extrinsic parameters of cameras. However, in our case the calibration also includes alignment of cameras to a coordinate system. The theoretical foundation for camera calibration is well established, and in our system we focused on the practical matter of calibration. Instead of re-implementing existing algorithms, our approach was to use existing tools, preferably ones that would allow the most flexibility (have SDK or API) to be integrated into the pipeline. Comparing different tools, it was decided to use the commercial software Agisoft *Photoscan*[2] for camera calibration. To automate scene calibration, we have reverse-engineered *Photoscan* project file format and created an external tool that can generate a new *Photoscan* project.

Instead of calculating intrinsic and extrinsic parameters of cameras with multiple calibration pattern views, calibration in *Photoscan* works differently. First, feature matching across the input photos is conducted. *Photoscan* detects points in the source photos which are stable under viewpoint and lighting variations and generates a descriptor for each point based on its local neighborhood [Lowe 2004]. These descriptors are used to detect correspondences across photos. After that, using a greedy algorithm *Photoscan* finds approximate camera locations and refines them with a bundle adjustment algorithm [Triggs et al. 2000].

Since *Photoscan* camera location estimation is based on feature points, we can exploit this, by taking photos of an object with a

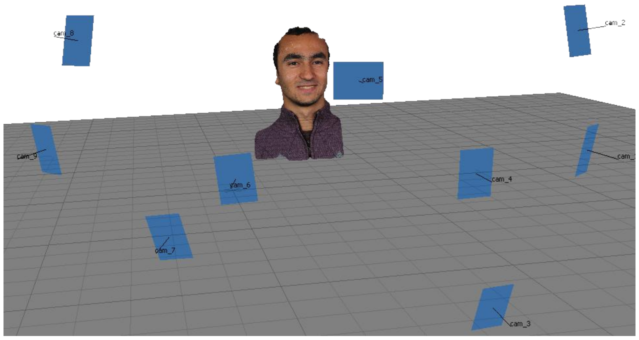


Figure 4: *Estimated position and rotation of cameras in the world space after scene calibration.*

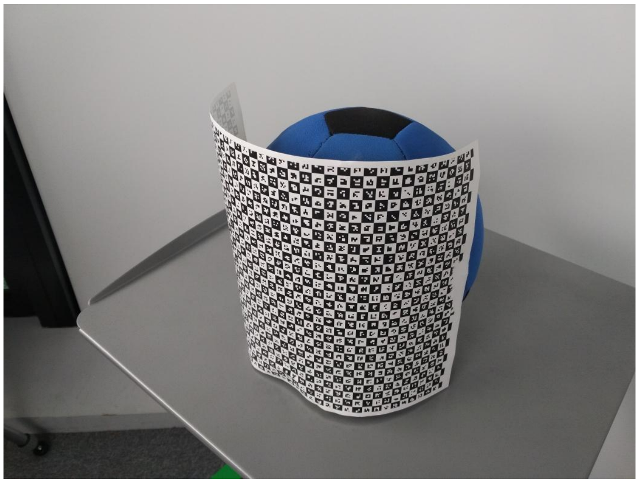


Figure 5: *Calibration pattern on the sphere used for the scene calibration.*

[2] Agisoft Photoscan Standard 1.3.1, Agisoft LLC, www.agisoft.com

strictly unique patterns. Using a tool presented by Atcheson et al. [Atcheson et al. 2010], we generate unique pattern of fiducial points. Inspired by Beeler et al. [Beeler et al. 2010], we put the pattern on a calibration sphere of size comparable to human head (Figure 5). Photos of such calibration sphere provide millimeter accuracy calibration with *Photoscan*, allowing to create high-definition facial meshes later down the pipeline. Figure 4 shows how the setup looks like in the world space after scene calibration.

3.4 Capture and meshing

When the cameras and lights are setup, and the scene is calibrated, it is possible to capture the subjects and create 3D model from the captured images. After acquiring photos, we used *Photoscan* to create a model. Since the whole scene is already calibrated, and camera parameters are defined, *Photoscan* generates a sparse feature point cloud. After that, using pair-wise depth map computation, *Photoscan* constructs a dense point cloud. Based on it, a triangle mesh is obtained by Screened Poisson Surface Reconstruction [Kazhdan and Hoppe 2013].

The obtained mesh contains extremely high number of vertices (up to 10 million) due to oversampling nature of photogrammetry. To make it suitable for real-time VR applications, we have to decimate (simplify) the acquired mesh. The performance of real-time applications depends on many things, not just polygon count of virtual object. For example, texture sizes, code complexity, hardware limitations (draw calls, fill rate, bandwidth limit, etc.) are all influencing the final performance, so defining a single target polygon count for the facial mesh is an ill-posed problem. Moreover, modern game engines rely on dynamically determining which level of details is required for the asset at the run time, and loading mesh version with appropriate polygon count. In our implementation of the pipeline, we used *Photoscan* built-in decimation feature.

3.5 Mesh cleaning

Even with proper calibration and correct capture process, obtained mesh is guaranteed to be noisy at the very least. Depending on (but not limited to) the parameters such as facial structure, skin properties, and facial hair, triangle meshes might have a number of defects like holes in the topology, unconnected components, non-manifold vertices, unreferenced vertices or many others. To fix such defects in the automated pipeline, we used the commercial software *Autodesk Maya*[3]. *Maya* was run in the headless mode executing a MEL script that does the following actions to clean up the acquired mesh:

- Hole filling
- Removal of unconnected isolated pieces
- Removal of unreferenced vertices
- Removal of edges with zero length
- Removal of faces with zero geometry area
- Removal of non-manifold geometry

3.6 Texturing

After actions described in Section 3.5, we obtain a cleaned mesh ready to be textured. One of the most important parts of texturing is UV mapping, also sometimes referred to as mesh parameterization: creating a map that describes how 2D texture map should be projected on to the 3D model. *Photoscan* can generate UV map automatically. However, the space usage efficiency of such automatically created UV map is very low (Figure 6, left column). Moreover, textures for such UV maps are virtually impossible to be modified manually by artist, if such requirement arises.

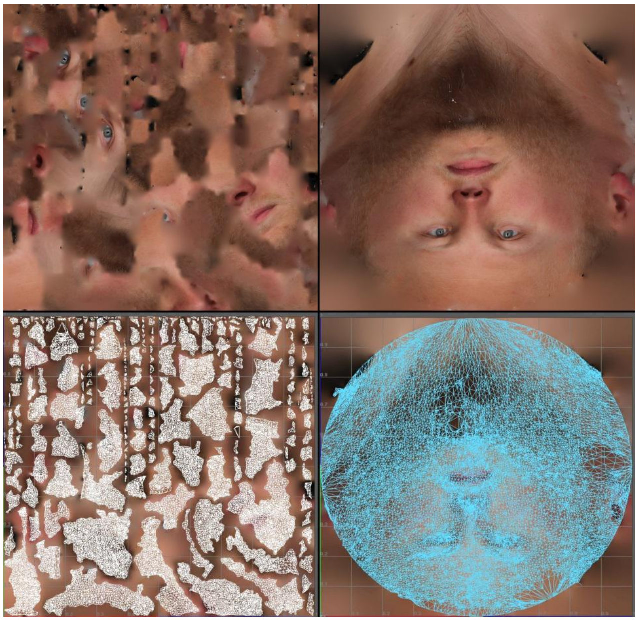


Figure 6: *Comparison of initially generated UV mapping (left column) and unwrapped cylindrical automatic UV mapping (right column).*

All facial meshes obtained by our setup have very similar structure of a half of cylinder. This observation can help us create a significantly more artist-friendly mapping without any stitches by using cylinder as a projection basis. However, obtained mapping almost always contains multiple overlapping regions and does not optimize space usage, so an additional step is required. Mesh cleaning (Section 3.5) ensured that mash is manifold, and that allows us to unwrap the mapping and flatten it into 2D mapping with no overlapping regions. Using such one-layered and continuous resulting mapping, we generate texture with *Photoscan* (Figure 6, right column).

3.7 Facial mesh normalization

The camera locations found with scene calibration (Section 3.3) are in an arbitrary coordinate system, meaning that the resulting 3D mesh is also in the same arbitrary coordinate system. There is no guarantee that during the next session coordinate system will be preserved. This is an unwanted variation, and moreover, landmark localization, which is required for mesh registration and subsequent animation transfer, is highly dependent on the stable mesh pose. There are different ways to handle this.

Some applications, where it is impossible to guide a photogrammetry subject, for example infants, require a very heavy pose correction [Wu et al. 2014]. Such approaches usually rely on localization of three most robust and pose invariant landmarks: nose tip and inner eye corners. Most widely used methods for such 3D landmark localization are least squares fitting [Guo et al. 2013] and HK-curvature analysis [Chang et al. 2006]. Both of these methods support face meshes that are rotated across any pose variations on X, Y, Z axes. The only restriction is that such face meshes must be manifold and contain no self-occlusions.

Both least squares fitting and HK-curvature analysis were tried out

[3] Autodesk Maya 2017, Autodesk Inc., autodesk.com/products/maya/

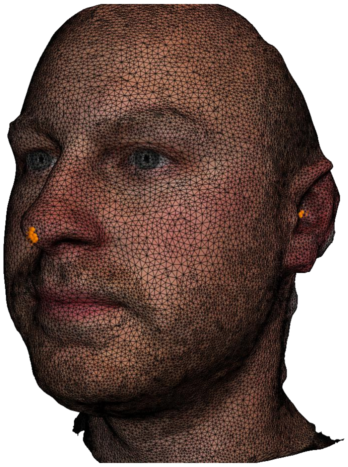


Figure 7: *Example of unwanted successful least square fitting. Orange dots represent top 10 candidate vertices. Out of 10 candidate vertices, 3 are located in cavum concha, the inferior portion of the cavity of the auricle of the ear.*

for normalization in our pipeline. First turned out to be not a reasonable option due to the fact that our meshes can not be assumed to be homogeneously scaled and definitely do not have similar topology. For some meshes, nose can consist of a thousands vertices, but on some others only a hundred. Introducing extra dimensions to account for an arbitrary mesh scaling will make it much more complicated and computationally intense, making the method less robust and even more prone to errors. Another issue with least squares is that some meshes might have unwanted areas that can be fit with the same parameters as wanted landmarks. In our case that was often the case for nose and areas like ears (Figure 7) and chin.

Second approach, HK-curvature analysis, produced very promising results, and that was one of the methods that we seriously considered for the pipeline. Implementing the idea of Colombo et al. [Colombo et al. 2006], we calculated mean (*H*) and Gaussian (*K*) curvature maps (Figure 8), applied thresholding to discard low curvature values:

$$|H(u,v)| >= T_h, |K(u,v)| >= T_k \qquad (1)$$

(where T_h and T_k are predefined experimentally tested thresholds) and then analyzed curvature. Inner eye landmarks (*endocanthion L/R*) are described as elliptical concave regions ($H < 0, K > 0$), while nose top (*pronasale*) is described as elliptical convex region ($H > 0, K > 0$). Even though the preliminary results are quite robust, curvature analysis still can not guarantee landmark detection in most cases. Also it is fairly dependant on captured face region. In cases when subject's wear was included in the capture (shirt collars, sweatshirts' hood etc.) some unwanted parts that fit curvature description of eyes or nose can be found.

Least squares fitting and HK-curvature analysis, while showing good potential, can not provide foolproof results and would require at least minimal human supervision. In our case, all photographed subjects are adults, are capable of holding the face relatively vertically straight and can keep the gaze straight for the moment of capture. During capture, we asked subject to look at the frontal camera. Assuming that the frontal camera is exactly in front of the subject's face, it is possible to use that to align the coordinate system with the new coordinate system, where central camera is located at $(0, 0)$ coordinate point. Using an active appearance model algorithm [Alabort-i-Medina et al. 2014], we obtain 68 facial landmarks (following the Multi-PIE [Gross et al. 2010] 68 points mark-up) with very high precision in 2D space of the frontal camera photo. By projecting them into the 3D space, we obtain landmarks on the mesh which allow us to determine roll and yaw rotation of the face, and compensate for them. The result of such alignment can be seen on Figure 9.

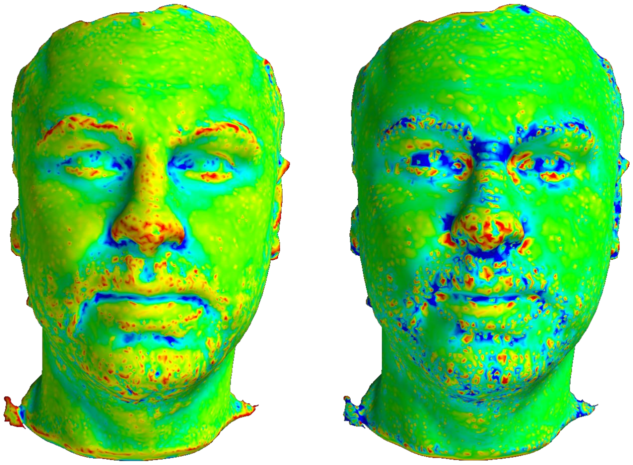


Figure 8: *H-Mean (left) and K-Gaussian (right) curvatures of unsmoothed acquired mesh. Red color indicates higher positive values, blue color indicates lower negative values, and green represents zero values.*

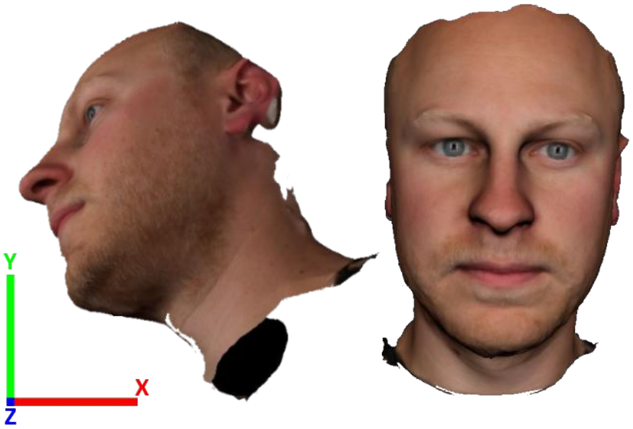


Figure 9: *Before (left) and after (right) automatic mesh normalization, based on camera position assumptions. Mesh is displayed in world coordinates.*

4 Animation transfer

4.1 Automatic face landmarking

Taking into consideration that the mesh is normalized (Section 3.7) and that the acquired mesh has photorealistic texture, we selected approach of using well established 2D facial landmarking by flattening acquired 3D meshes into 2D images. Active appearance model algorithm [Alabort-i-Medina et al. 2014] previously provided a set of 68 2D facial landmarks. However, the retrieved landmarks pose number of problems.

First, 2D landmarks projected onto 3D world space do not guaran-

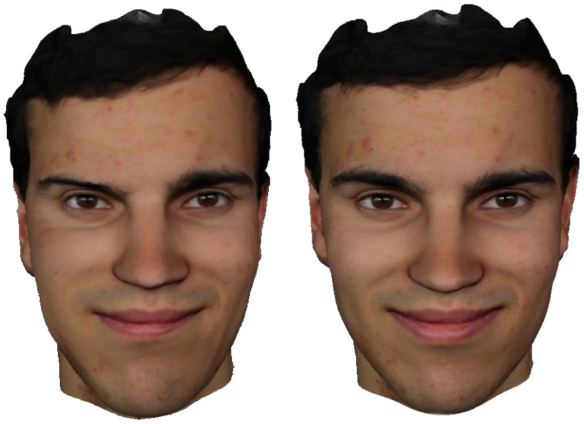


Figure 10: *Side by side comparison of the same animation on the facial mesh corresponded with 68 automatic landmarks (on the left) and the facial mesh corresponded with 24 automatic landmarks (on the right).*

tee that they will be projected onto facial mesh. To do correct blendshape transfer, it is necessary to have a correct correspondence, and it can be built only with the same number of anchor points (landmarks). To get over this issue, we use landmark labels and make sure that if the N-th landmark has not been projected onto the mesh, we remove the Nth landmark from the source mesh landmark set.

Another problem is that an increase in quantity of landmarks does not guarantee increase in quality of correspondence. In fact, the more landmarks there are, the more is the chance that this landmark will introduce an incorrect anchor point, leading to less accurate correspondence. To find the subset of the most stable and accurate landmarks, we have conducted comparison of resulting landmarks with manually annotated ground truth.

The results of such comparison showed that the most stable and accurate landmarks are the frontal landmarks:

- 9 - chin
- 37, 40; 43, 46 - left and right eye
- 18, 20, 22; 23, 25, 27 - left and right eyebrow
- 28, 31, 32, 36 - nose radix, nose tip and outer nostrils
- 49, 55, 52, 63, 67, 58 - mouth contour

Using the 21 landmark subset instead of full 68 set significantly improved the quality of correspondence and, subsequently, of transferred animations (Figure 10).

4.2 Registration and blendshape transfer

Deformation transfer method proposed by Sumner and Popovic [Sumner and Popović 2004] requires meshes to be corresponded. In our case, obtained mesh is a heavily modified triangulated point cloud, meaning that meshes will never have identical topology. We register the source mesh to the acquired facial mesh, following approach by Pawaskar et al. [Pawaskar et al. 2013]. The source mesh will deform so that it records the geometrical features of the acquired facial mesh, but intrinsically it contains the same topology.

Using non-rigid deformation [Amberg et al. 2007], we morph source mesh close to target mesh by solving per-vertex affine transformation. Introduction of Laplacian regularization term minimizes the difference between transformation matrices of adjacent vertices while simultaneously getting rid of noise extremities, small area spikes. Next, a fitting refinement process is applied by iteratively changing the ”stiffness” of regularization term. After deformation convergence, the stiffness is reduced, and the process repeats until a certain threshold is reached, eventually providing us with a morphed source mesh, a mesh that has direct correspondence to target mesh.

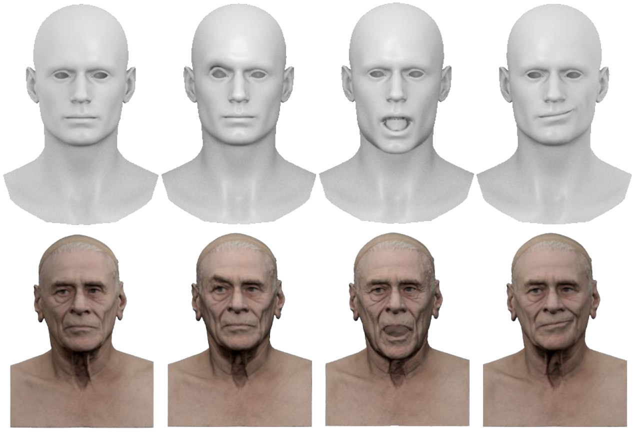


Figure 11: *Source mesh with three manually created blendshapes (top row) and acquired 3D mesh with transferred blendshapes (bottom row). The leftmost mesh of bottom row is the original target mesh, while other meshes of second row are generated blend shapes. Note: the bottom mesh was not acquired with the setup described in this paper and is property of Ten 24 Media Ltd.*

After obtaining the correspondence, we transfer the blendshapes from source to morphed source as described by [Sumner and Popović 2004] with addition of implementing [Sumner 2006] to preserve connectivity between triangles. Figure 11 shows example results of the transfer.

5 Evaluation

To evaluate the proposed pipeline, we assess the believability of produced rigged 3D meshes. To determine it, we have conducted 4 user experiments with 8 participants aged 23 to 53. 4 participants reported professional experience in computer graphics and 4 reported no experience. 6 participants are male and 2 participants are female. Each participant was presented with experiments described in the next subsection.

5.1 Purpose, materials and methods

We have conducted a total of 4 different experiments evaluating results of separate pipeline steps: meshing, cleaning, texturing, and, finally, deformation transfer. Each experiment features 5 faces (Figure 12).

1. **Can participants recognize the person by the model without the texture?**
   The goal of this experiment is to determine if quality of reconstructed face surface is precise enough to preserve subject's face surface uniqueness. We do this by providing participants with an image of 5 subjects faces and access to a 3D model viewer with 5 models shown individually, and asking person to match the photos with the models. The order of appearance in the image is different from the order of appearance of the models. The participant can freely move camera in the viewer. We also measure time needed to complete matching,

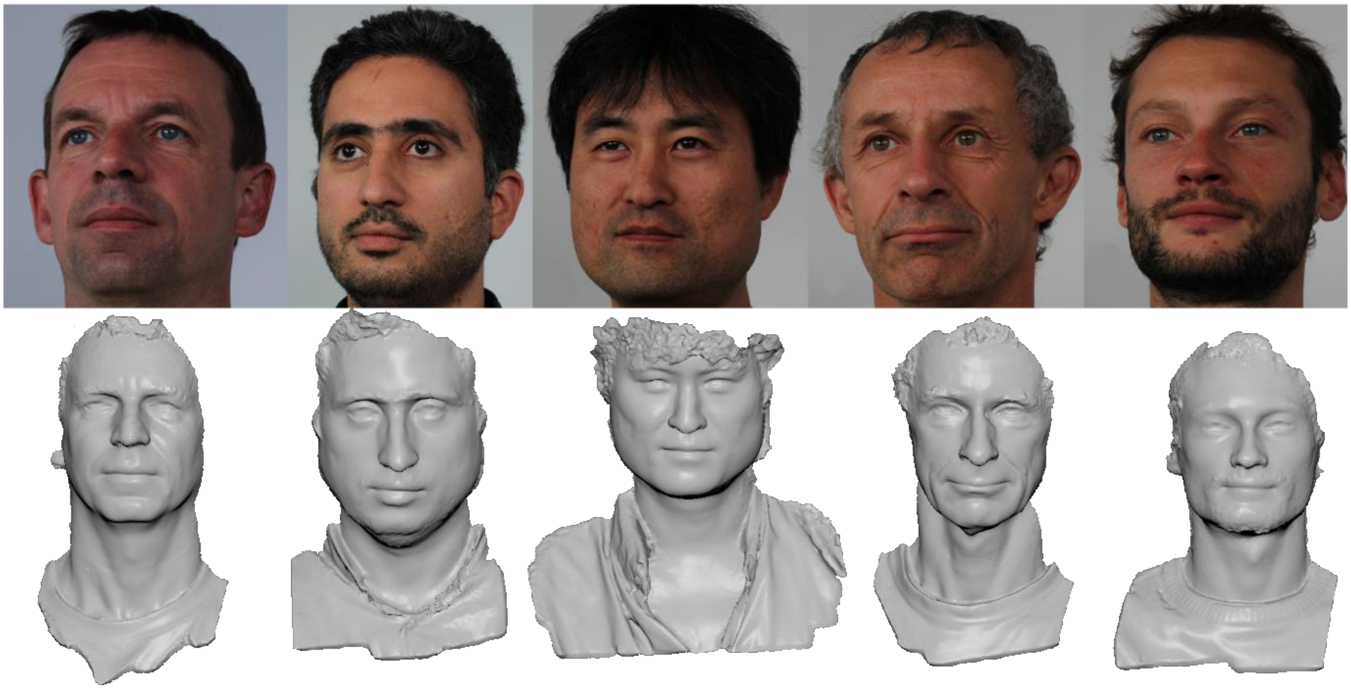


Figure 12: *Photos (top row) and models (bottom row) used in the study participants for the first experiment.*

count how many tries were unsuccessful, and ask the participant their match strategy and on what part of the model they focus the most while matching.

2. **At what camera distance participants notice artifacts on the model?**
   The goal of this experiment is to determine the relative size of the face to the screen, at which artifacts introduced by the facial reconstruction become distinctly visible. We do this by showing participants frontal 2D perspective projections (angle of 30) of the face models at an increasingly bigger size. The distance from the participant to the screen remained constant, at the participant's consideration. The order in which we have shown the models was randomized. The projections are shown without time constrains. The sizes of steps are defined as following: extremely long shot (face takes 4% of image height), long shot (face takes 9% of image height), full figure shot (25%), medium shot (60%), close up (100%) and italian shot (100%+, focused on eyes and nose). Once the participant notices a problem, we show the projection again and ask the participant to denote the area of a noticed artifact.

3. **How faithful is the produced face compared to the face of of the subject?**
   The goal of this experiment is to determine how faithful is the model to the photos. For examples, mesh might be oversmoothed, losing fine details of the skin, or the other way, too bumpy while the skin is actually smooth. The experiment is done by presenting participants a model and a set of photographs, from which the model was created, without any viewing time constrains. The order of model appearance was randomized. We then ask the participant to rate the resulting accuracy on the scale of 0 to 7, where 0 is ”nothing like” and 7 is ”completely accurate”. The participant can freely move camera around the model in the viewer. We also ask participants to denote the area that differs the most.

4. **How genuine are the transferred animations?**
   The goal of this experiment is to determine how believable are the transferred animations. First, we manually select blendshape weights to show three emotions: sadness (0.5 * (1+4+15 FACS AU): Inner Brow Raiser, Brow Lowerer, Lip Corner Depressor)), happiness (0.5 * (6+12 FACS AU)): Cheek Raiser, Lip Corner Puller) and anger (1.0 * (4+5+7+23 FACS AU)): Brow Lowerer, Upper Lid Raiser, Lid Tightener, Lip Tightener). Then, we apply the weights to 5 automatically obtained models, resulting in 15 different views. 9 of them can be seen on the Figure 13. The experiment is done by showing participants a neutral expression frontal photo and showing the models showcasing simulated emotions. The order of appearance was randomized. Participant's objective is to define what emotion the face is showing and rate how believable the emotion is on the Likert scale from 0 to 7, where 0 is ”very unrealistic” and 7 is ”absolutely realistic”. If the rating is not 7, we also ask participants why and what areas are the ones that seem the most unrealistic.

5.2 Results

1. **Facial surface details**
   The goal of the first experiment was to determine if surface details are preserved well enough to distinguish between models without texture. Participants, on average, spent 32 seconds to correctly match the models with their input photographs. All participants correctly identified subjects on the first try. Such results suggests that the quality of reconstructed face surfaces is preserving subject's facial surface features and is good enough to be able to robustly identify a person's just by surface geometry.

   We also asked people to describe their match strategy. The most popular way was the jaw shape matching with 7 participants claiming to pay the most attention to it. Two other participants used nose shape, and one used eyes. Interesting to note that 9 of 10 participants matched models not in the

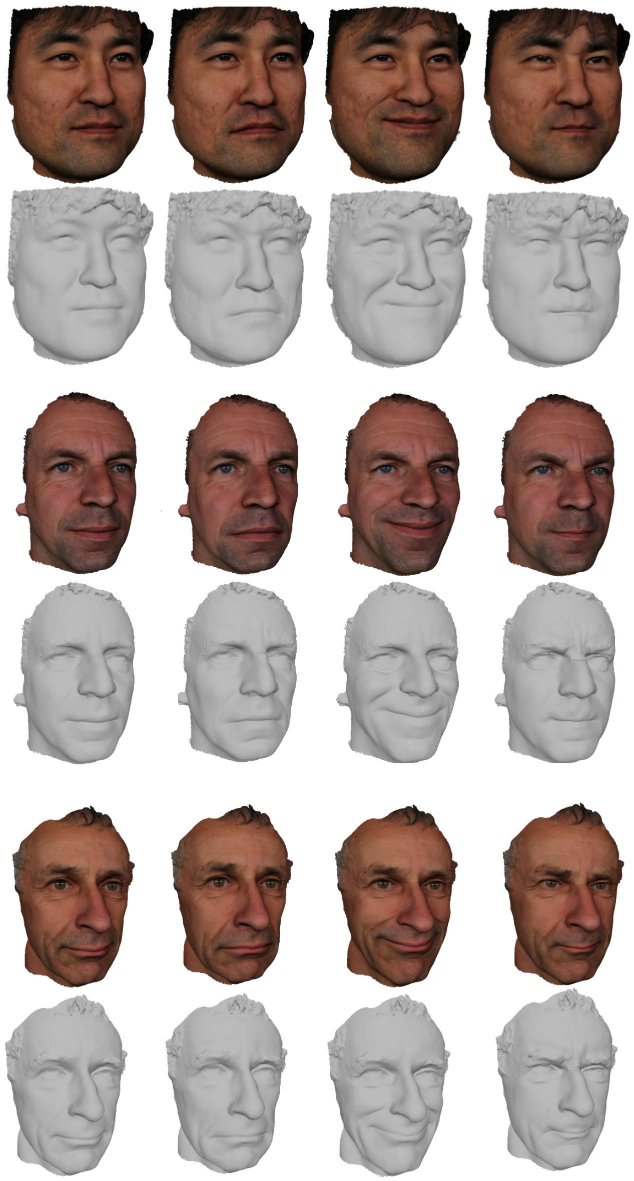


Figure 13: *Synthesized views and untextured models (every second row), for viewpoints different from the original camera images, across subjects of varying appearance. First column shows original photogrammetrically obtained meshes (neutral face), while three other columns show synthesized poses. The weights of blendshapes for poses of the same emotions are the same across subjects.*

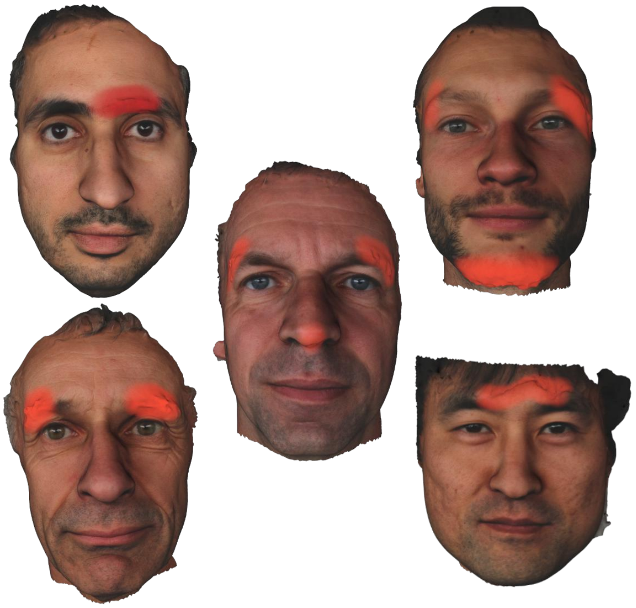
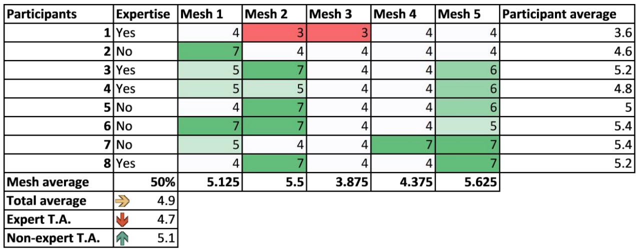


Figure 14: *Textured models that were created with our pipeline and used in the user experiments. Red smudges represent areas where participants noticed artifacts the most.*

Table 1: *Numerical results of Experiment 2.*



Participants	Expertise	Mesh 1	Mesh 2	Mesh 3	Mesh 4	Mesh 5	Participant average
1	Yes	4	3	3	4	4	3.6
2	No	7	4	4	4	4	4.6
3	Yes	5	7	4	4	6	5.2
4	Yes	5	5	4	4	6	4.8
5	No	4	7	4	4	6	5
6	No	7	7	4	4	5	5.4
7	No	5	4	4	7	7	5.4
8	Yes	4	7	4	4	7	5.2
Mesh average	**50%**	**5.125**	**5.5**	**3.875**	**4.375**	**5.625**	
Total average	➡ 4.9						
Expert T.A.	⬇ 4.7						
Non-expert T.A.	⬆ 5.1						

order of appearance, but what seemed random.

2. **Mesh artifacts**
   The goal of the second experiment was to subjectively rate quality of the resulting 3D models at different distance from the camera. To represent the results numerically, we have matched the various shot sizes to the Likert scale from 0 to 7, where 0 is seeing a problem already on the first zoom level and 7 not seeing any problems in an extreme close-up zoom level. The average zoom level at which participants noticed problems is 4.9, translating to somewhere between the medium shot (face taking 60% of image height) and close up shot (face taking 100% of image height). The minimum zoom level when artifacts were noticed was 3 (25% of image height), 2 times. The zoom level of 7 (no problems even in the extreme close-up) was noted 9 times. Table 1 shows results of the experiment in one table.

   Apart from collecting the zoom level at which artifacts become visible, we asked participants to describe the artifacts and their location. Figure 14 highlights the artifacts that participants noticed. Results of the this experiment suggest that the reconstructed face can be confidently used for the background characters, whose faces will not be shown up close.

3. **Accuracy of person's representation**

Table 2: *Numerical results of Experiment 3.*

Participants	Expertise	Mesh 1	Mesh 2	Mesh 3	Mesh 4	Mesh 5	Participant average
1	Yes	6	5	6	5	6	5.6
2	No	5	6	5	6	6	5.6
3	Yes	6	7	6	6	6	6.2
4	Yes	6	6	5	6	7	6
5	No	5	7	6	5	7	6
6	No	6	7	6	6	6	6.2
7	No	6	6	7	7	7	6.6
8	Yes	5	5	7	5	6	5.6
Mesh average	**50%**	**5.625**	**6.125**	**6**	**5.75**	**6.375**	
Total average	➡ 5.975						
Expert T.A.	⬇ 5.85						
Non-expert T.A.	⬆ 6.1						

Table 3: *Numerical results of Experiment 4.*

Participants	Expertise	Mesh 1	Mesh 2	Mesh 3	Mesh 4	Mesh 5	Participant average
1	Yes	7	6	7	6	6	6.4
2	No	6	7	7	5	6	6.2
3	Yes	6	7	6	5	7	6.2
4	Yes	7	7	7	7	5	6.6
5	No	7	7	7	7	5	6.6
6	No	7	7	7	7	7	7
7	No	7	7	6	6	7	6.6
8	Yes	4	7	5	6	7	5.8
Mesh average	**50%**	**6.375**	**6.875**	**6.5**	**6.125**	**6.25**	
Total average	➡ 6.425						
Expert T.A.	⬇ 6.25						
Non-expert T.A.	⬆ 6.6						

The goal of the third experiment was to determine how faithful is the model to the photos. On the scale from 0 to 7, where 0 is ”nothing like” and 7 is ”completely accurate”, on average our meshes received user rating of 5.975, with lowest rating of 5 occurring 9 times, and highest rating of 7 occurring also 9 times. Table 2 shows results of the experiment in one table. This rating suggests that the models are reasonably accurate and viable for faithful representation of human faces in it's current implementation. However, participants also noted the areas that are not faithful. According to them, the most problematic areas are the eyebrows, followed by the facial hair. The root of problems in those two areas is due to the nature of photogrammetry, not allowing to capture nearly-transparent and very thin objects. We are aware of the issue and have proposed a potential solution in Future work (Section 6.1).

4. **Believability of the transferred emotions**
The goal of the fourth experiment was to assess believability of the transferred emotions. All participants correctly distinguished the base shown emotions: sadness, happiness and anger. The average rating received rating is 6.425, with lowest rating of 4 occurring once and highest rating of 7 occurring 24 times. Table 3 shows results of the experiment in one table. Such high average rating suggests that the synthesized emotions can be correctly conveyed to the end user.

In 11 cases participants noted feeling of fakeness of the conveyed emotions, but without doubting the believebility that the person can show such emotion. For example, some participants explicitly distinguished happiness as ”tired but content”, ”shy smile” and in one case even as ”polite smile to avoid being rude”. When asked to elaborate what exactly gave that feeling, participants noted discrepancies in intensities of smile parts. For example, the skin around the eyes was tightened significantly less intense than expected by participants to convey genuine smile.

As expected, on average, participants with expertise in computer graphics rated meshes as slightly less accurate, realistic or believable. This can be explained by their exposure to the research area.

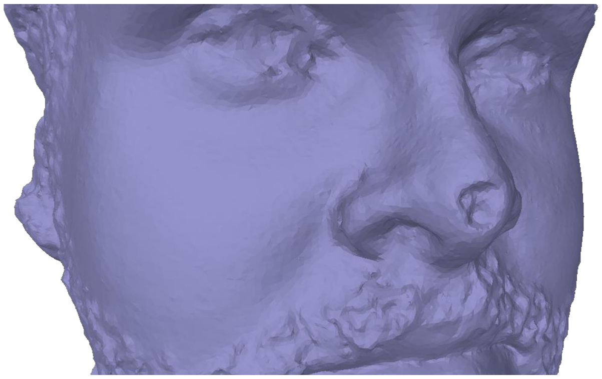


Figure 15: *Example of artifacts produced by specularity of the nose tip.*

6 Discussion

The proposed pipeline is established and validated. A number of limitations has been identified, as well as areas that can benefit from further improvement.

Specular areas of the face distort the mesh, for example the specularity on the tip of the nose (Figure 15). Such areas reflect the direct light source, even if it is heavily diffused. Indirect lighting and cross-polarization can prevent this before meshing process. Alternatively, a plausible simulated reconstruction can fix affected area post-factum.

Another limitation is that the proposed pipeline heavily depends on assumption of how the camera rig is constructed. If there are no cameras that can capture subject's face *en face*, we are not able to obtain landmarks, and subsequently not able to normalize the mesh, construct the correspondence or animate it.

6.1 Future work

Photogrammetry result quality directly corresponds to the quality of capture photos [Wenzel et al. 2013]. However, it is not clear how exactly, and without any strictly defined rules, other than just a number of ”guidelines”. Every photogrammetry subject type has different optimal capture setup, and finding the most optimal way to capture human face is an open problem.

Instead of using 2.5D landmarking, one obvious improvement is to use 3D landmarking, for example with methods mentioned in section 3.7. Potentially, such methods can be more precise as they use not just intensity (color), but also shape information. 3D landmarks would also allow universal facial mesh normalization based on landmarks, instead of assumptions of camera positions or subject's pose.

The current result of the pipeline is a single mesh, which includes eyes and hair (including facial) as a part of surface. Extracting eyes and hair, and being able to control them separately would significantly improve quality of the end result. An extensive research has been done on capture of the eyes by Berard et al. [Bérard et al. 2014] and the hair by Beeler et al. [Beeler et al. 2012]. Both of methods require manual supervision, and it would be interesting future work to automate such separation.

Face is one of the (if not the most) crucial parts of person's identity. However, personalized face by itself is not enough to believable represent real person in virtual space. Capturing and faithfully representing person's body is the next logical advancement for bringing

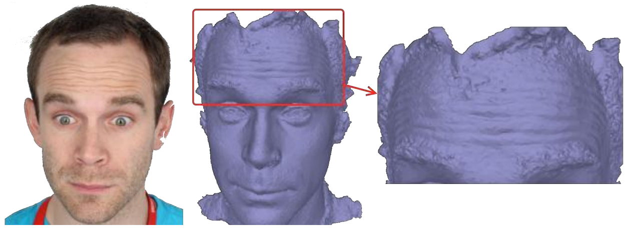


Figure 16: *Example of capture fidelity in our setup, left to right: cropped area of one of nine input photos, reconstructured raw model, close up of forehead with heavy skin geometric variation.*

humans into the virtual space.

6.2 Conclusions

We have presented a new technique for automatic creation of rigged facial meshes based on a set of multi-cameras photos of a neutral face. First, using photogrammetry, we obtain a mesh we acquire facial geometry and appearance. Using automatical landmarking and ICP registration algorithm, we obtain direct correspondence of the acquired facial mesh and the existing template. Then, using deformation transfer, we transfer existing blendshapes from the template to our new mesh. Figure 13 shows a possible variety of facial expressions of meshes created with the proposed approach. Figure 16 demonstrates the fidelity of photogrammetric reconstruction. Figure 11 shows results of a deformation transfer.

Creating a rigged facial mesh ready for game engines is a highly manual process that requires work of a team with a dedicated skillset on the scale of man-weeks. Areas that require even more fidelity, such as visual effects production, require amount on the scale of man-months[4]. The processing time of our workflow implementation, from image input to output of a rigged 3D model, takes around 10-20 minutes (depends on the number of transferred animations) on a high-end computer[5]. Such efficiency brings improvement in speed compared to manual work in the order of two magnitudes. It is important to note that the current implementation has had no additional optimizations or parallelization, and we believe that it is possible to reduce compute time to even less.

The conducted user experiments suggest that the obtained rigged facial meshes:

- do not lose unique surface details of the subjects,
- are out of the box suitable for secondary characters without close up shots,
- fairly accurate represent the subject faces,
- can be used to confidently convey emotions.

To our knowledge, this is the first method for completely automatic generation of rigged facial meshes (without deep learning) currently described in the literature. A system of this type would be invaluable for generation of rigged facial datasets. One such database, collected with the workflow described in this study, is expected to be released with a separate publication later this year. We believe this database will be helpful for the scientific community, especially in machine learning research area.

Acknowledgements

I would like to thank my first supervisor dr. dr. Egon L. van den Broek for his invaluable feedback. His guidance and open-mindedness made this work possible and allowed me to complete this international project on time.

I want to express my immense gratitude to Technicolor supervisors, dr. Quentin Avril and dr. Fabien Danieau, for letting me be a part of Immersive Lab in Technicolor for the past 6 months. Their productive supervision and deep knowledge provided me with a ton of indispensable experience.

I also want to thank prof. dr. Remco C. Veltkamp for finding time to read and evaluate this work, prof. dr. Marc van Kreveld for helping me with administrative aspects of the project before dr. van den Broek became my supervisor, and of course UU International Office, for making this international experience possible.

Finally, I want to thank my family, and especially my parents. They are the ultimate role models, life guides and my biggest fans. Everything I have achieved in my life is due to their love and support.

References

ALABORT-I-MEDINA, J., ANTONAKOS, E., BOOTH, J., SNAPE, P., AND ZAFEIRIOU, S. 2014. Menpo: A comprehensive platform for parametric image alignment and visual deformable models. In *Proceedings of the ACM International Conference on Multimedia*, ACM, New York, NY, USA, MM '14, 679–682.

AMBERG, B., ROMDHANI, S., AND VETTER, T. 2007. Optimal step nonrigid icp algorithms for surface registration. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, IEEE, 1–8.

ATCHESON, B., HEIDE, F., AND HEIDRICH, W. 2010. CALTag: High Precision Fiducial Markers for Camera Calibration. In *Vision, Modeling, and Visualization (2010)*, The Eurographics Association, R. Koch, A. Kolb, and C. Rezk-Salama, Eds.

BARAN, I., VLASIC, D., GRINSPUN, E., AND POPOVIĆ, J. 2009. Semantic deformation transfer. *ACM Trans. Graph. 28*, 3 (July), 36:1–36:6.

BEELER, T., BICKEL, B., BEARDSLEY, P., SUMNER, B., AND GROSS, M. 2010. High-quality single-shot capture of facial geometry. *ACM Trans. Graph. 29*, 4 (July), 40:1–40:9.

BEELER, T., BICKEL, B., NORIS, G., BEARDSLEY, P., MARSCHNER, S., SUMNER, R. W., AND GROSS, M. 2012. Coupled 3d reconstruction of sparse facial hair and skin. *ACM Trans. Graph. 31*, 4 (July), 117:1–117:10.

BÉRARD, P., BRADLEY, D., NITTI, M., BEELER, T., AND GROSS, M. 2014. High-quality capture of eyes. *ACM Trans. Graph. 33*, 6 (Nov.), 223:1–223:12.

BERGERON, P., AND LACHAPELLE, P. 1985. Controlling facial expressions and body movements in the computer generated animated short 'tony de peltrie'.

BROWN, K., AND HAMILTON, A., 2016. Photogrammetry and 'star wars battlefront', Mar.

CHANG, K. I., BOWYER, K. W., AND FLYNN, P. J. 2006. Multiple nose region matching for 3d face recognition under varying facial expression. *IEEE Trans. Pattern Anal. Mach. Intell. 28*, 10 (Oct.), 1695–1700.

[4] For example, Technicolor reports approximately 400 man-days (20 artists working for 20 days) for human characters that will be shown in close-up shots

[5] Intel i7 3.1Ghz, 64Gb RAM, GeForce GTX 1080 Ti

COLOMBO, A., CUSANO, C., AND SCHETTINI, R. 2006. 3d face detection using curvature analysis. *Pattern Recognition 39*, 3, 444 – 455.

DEBEVEC, P., HAWKINS, T., TCHOU, C., DUIKER, H.-P., SAROKIN, W., AND SAGAR, M. 2000. Acquiring the reflectance field of a human face. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, SIGGRAPH '00, 145–156.

DEBEVEC, P. 2012. The light stages and their applications to photoreal digital actors. *SIGGRAPH Asia 2012 Technical Briefs*.

FUCHS, M., BLANZ, V., LENSCH, H., AND SEIDEL, H. P. 2005. Reflectance from images: a model-based approach for human faces. *IEEE Transactions on Visualization and Computer Graphics 11*, 3 (May), 296–305.

GHOSH, A., HAWKINS, T., PEERS, P., FREDERIKSEN, S., AND DEBEVEC, P. 2008. Practical modeling and acquisition of layered facial reflectance. *ACM Trans. Graph. 27*, 5 (Dec.), 139:1–139:10.

GHOSH, A., CHEN, T., PEERS, P., WILSON, C. A., AND DEBEVEC, P. 2010. Circularly polarized spherical illumination reflectometry. *ACM Trans. Graph. 29*, 6 (Dec.), 162:1–162:12.

GROSS, R., MATTHEWS, I., COHN, J., KANADE, T., AND BAKER, S. 2010. Multi-pie. *Image and Vision Computing 28*, 5, 807–813.

GUO, J., MEI, X., AND TANG, K. 2013. Automatic landmark annotation and dense correspondence registration for 3d human facial images. *BMC Bioinformatics 14*, 1, 232.

KAZHDAN, M., AND HOPPE, H. 2013. Screened poisson surface reconstruction. *ACM Trans. Graph. 32*, 3 (July), 29:1–29:13.

KLEHM, O., ROUSSELLE, F., PAPAS, M., BRADLEY, D., HERY, C., BICKEL, B., JAROSZ, W., AND BEELER, T. 2015. Recent advances in facial appearance capture. *Comput. Graph. Forum 34*, 2 (May), 709–733.

LI, H., WEISE, T., AND PAULY, M. 2010. Example-based facial rigging. In *ACM SIGGRAPH 2010 Papers*, ACM, New York, NY, USA, SIGGRAPH '10, 32:1–32:6.

LOWE, D. G. 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision 60*, 2 (Nov), 91–110.

MA, W.-C., HAWKINS, T., PEERS, P., CHABERT, C.-F., WEISS, M., AND DEBEVEC, P. 2007. Rapid acquisition of specular and diffuse normal maps from polarized spherical gradient illumination. In *Proceedings of the 18th Eurographics Conference on Rendering Techniques*, Eurographics Association, Aire-la-Ville, Switzerland, Switzerland, EGSR'07, 183–194.

MAGNENAT-THALMANN, N., LAPERRIÈRE, R., AND THALMANN, D. 1988. Joint-dependent local deformations for hand animation and object grasping. In *Proceedings on Graphics Interface '88*, Canadian Information Processing Society, Toronto, Ont., Canada, Canada, 26–33.

NGAN, A., DURAND, F., AND MATUSIK, W. 2005. Experimental analysis of brdf models. In *Proceedings of the Sixteenth Eurographics Conference on Rendering Techniques*, Eurographics Association, Aire-la-Ville, Switzerland, Switzerland, EGSR '05, 117–126.

PAWASKAR, C., MA, W. C., CARNEGIE, K., LEWIS, J. P., AND RHEE, T. 2013. Expression transfer: A system to build 3d blend shapes for facial animation. In *2013 28th International Conference on Image and Vision Computing New Zealand (IVCNZ 2013)*, 154–159.

POZNANSKI, A., 2014. Visual revolution of the vanishing of ethan carter, Mar.

REICHARDT, J. 1978. *Robots: Facts, Fiction and Prediction*. Viking Penguin.

SUMNER, R. W., AND POPOVIĆ, J. 2004. Deformation transfer for triangle meshes. *ACM Trans. Graph. 23*, 3 (Aug.), 399–405.

SUMNER, R. W. 2006. *Mesh Modification Using Deformation Gradients*. PhD thesis, Cambridge, MA, USA. AAI0809158.

TINWELL, A., GRIMSHAW, M., AND ABDEL-NABI, D. 2014. Nonverbal communication in virtual worlds. ETC Press, Pittsburgh, PA, USA, ch. The Uncanny Valley and Nonverbal Communication in Virtual Characters, 325–341.

TRIGGS, B., MCLAUCHLAN, P. F., HARTLEY, R. I., AND FITZGIBBON, A. W. 2000. *Bundle Adjustment — A Modern Synthesis*. Springer Berlin Heidelberg, Berlin, Heidelberg, 298–372.

TUNWATTANAPONG, B., GHOSH, A., AND DEBEVEC, P. 2011. Practical image-based relighting and editing with spherical-harmonics and local lights. In *2011 Conference for Visual Media Production*, 138–147.

WALDHÄUSL, P., AND OGLEBY, C. 1994. 3 x 3 rules for simple photogrammetric documentation of architecture. *International Archives of Photogrammetry and Remote Sensing 30*, 426–429.

WATERS, K. 1987. A muscle model for animation three-dimensional facial expression. *SIGGRAPH Comput. Graph. 21*, 4 (Aug.), 17–24.

WENZEL, K., ROTHERMEL, M., FRITSCH, D., AND HAALA, N. 2013. Image acquisition and model selection for multi-view stereo. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci 40*, 251–258.

WEYRICH, T., MATUSIK, W., PFISTER, H., BICKEL, B., DONNER, C., TU, C., MCANDLESS, J., LEE, J., NGAN, A., JENSEN, H. W., AND GROSS, M. 2006. Analysis of human faces using a measurement-based skin reflectance model. *ACM Trans. Graph. 25*, 3 (July), 1013–1024.

WU, C., WILBURN, B., MATSUSHITA, Y., AND THEOBALT, C. 2011. High-quality shape from multi-view stereo and shading under general illumination. In *CVPR 2011*, 969–976.

WU, J., TSE, R., AND SHAPIRO, L. G. 2014. Automated face extraction and normalization of 3d mesh data. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 750–753.

ZWERMAN, S., AND OKUN, J. 2012. *Visual Effects Society Handbook: Workflow and Techniques*. Taylor & Francis.

Appendix A. 3D Facial Geometry Acquisition and Facial Appearance Capture: Literature Review

Abstract—**This paper examines the literature on topics of active and passive methods of 3D facial geometry acquisition, and image-based and parametric facial appearance methods. Various approaches are described and compared by critical dimensions, describing both seminal works on the topic and current state-of-the-art. Paper also summarizes methodological limitations and recommendations for further work in the areas.**

I. Introduction

Modeling and rendering realistic human characters is not a trivial task. Faces of believable realistic characters that cross the Uncanny Valley require high-quality geometry, texture maps, reflectance properties, subsurface scattering shading, surface detail at the microlevel of skin pores and skin wrinkles, and realistic facial expressions under arbitrary lighting [16]. This problem has inspired research on some great number of solutions in facial modeling, rendering, and facial animation. However, reproducing human faces is still a challenge in computer graphics because humans are sensitive to facial appearance and quickly, on a subconscious level, recognize any anomalies in face 3D geometry or dynamics. Moreover, game designers are even advised to create less realistic humanoid characters, because highly-realistic characters even with small abnormalities seem to be perceived as uncanny and produce negative reactions [10] [36].

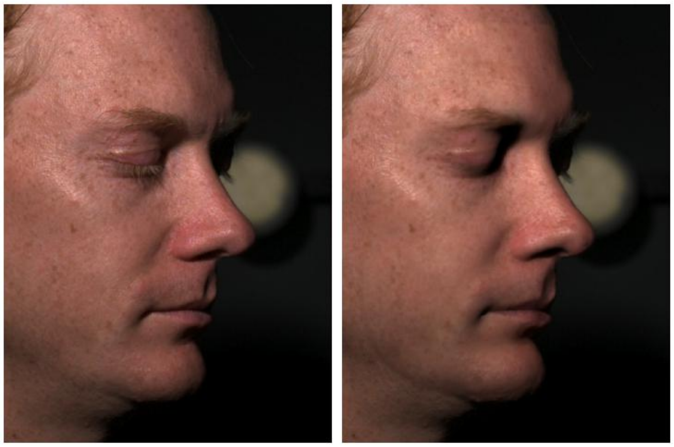


Fig. 1. Photograph and a rendering done with skin reflectance model presented by Weyrich et al. [42]

With decreasing cost and increasing technological progress, 3D photogrammetric imaging systems are more and more common for all kinds of applications. Instead of modeling human face, it is possible to do a photogrammetric acquisition of human face geometry. This is of interest in multiple fields of research: in medicine, in the movie and games industries, for archival purposes, for crime investigations and many other domains.

Facial acquisition is essentially a process consisting of there steps:

- **Capture** - The equipment should be properly setup and tuned for capturing parts of the face surface in high detail under bright ambient illumination. Usually, using a computer-controlled remote control, all capturing equipment is simultaneously taking pictures, which later are used for the next steps.
- **Face geometry reconstruction** - Using various methods, for example such as iterative binocular stereo method, it is possible to reconstruct a single high-resolution mesh of facial geometry from 2D images acquired in the previous step.
- **Appearance recreation** - The images allow not only reconstruct mesh, but, with enough capturing quality, to also acquire the surface data with micro-surface details such as skin pores and hair follicles, enabling to also create a high-quality texture for the model. This steps involves application of various captured data, such as reflectance, diffuse and normal maps.

This study is concerned with a systematic review of existing literature on the topics of 3D facial geometry acquisition and facial appearance capture. Both topics are reviewed with applying of scientific strategies that limit bias by the systematic assembly, critical appraisal and synthesis of all relevant studies on the specified topics. We used explicit and reproducible methods which should enforce the systematicness of the literature review.

The second part of the thesis already comes with a defined hardware inventory available, more specifically:

- 9 DLSR cameras Canon 100D + 9 lenses FF 50mm
- 3 tripods Manfrotto + 6 magic arms Manfrotto
- 2*4 neon light Kino Flo Tegra

and we are not going to review methods that require equipment other than listed above.

A. Facial geometry acquisition

The methods for 3D facial geometry capture developed in the last two decades can be split into active and passive systems. Active capture systems require special-purpose hardware, extra constrains in setup and often employ time-division multiplexing for methods like polarization. Such systems usually are based on laser, structured light or gradient-based illumination. While the results they provide are often very robust, passive systems often are much more versatile and adaptive, allowing different arrangements of setup, numbers of camera and virtually no constrains on camera position, but sacrificing on reliability and accuracy [2].

This paper reviews both types of capture systems, with the exclusion of methods that are not possible to recreate with the previously mentioned hardware inventory, for example methods that require projections or lasers. However, for the sake of completeness of the study, we are also going to include some facial performance capture literature into the review. While studying performance capture is not the goal of this paper, very often facial geometry acquisition plays a centerpiece part in performance capture.

B. Facial appearance capture

The ultimate goal of facial acquisition is to be able to render a the under arbitrary lighting and from any viewing position, including non-trivial fine details of a human face. Facial geometry acquisition is just one part of the solution, and the other part is acquisition of facial appearance. In the nutshell, the main challenge of facial appearance capture is to record the way light interacts with skin, so that later on it the recorded data can be used to complete the previously mentioned ultimate goal.

The human skin is a complex multi-layered structure, and light interacts with it in a convoluted way. A part of light is reflected off the oily skin surface, resulting view-dependent highlights which also uncover details like pores and wrinkles. Another part of light does not reflect off the surface, but travels through the layers of the skin. Each layer has different characteristics, and therefore interacts with the light differently too. In addition to the complex effect of skin layers on color, light can also leave the skin from a different position than it entered. This produces characteristic soft appearance and smooths out the reflectance [18].

Facial acquisition methods must cope with complexity of light interactions with skin, and two broad categories can be distinguished based on how method handles it: image-based methods and parametric methods. Image-based methods exhaustively capture the exact face appearance under various lighting and viewing conditions, and then solve the rendering problem through weighted image combinations. In turn, parametric methods instead aim at modeling the structure of skin with suitable approximations, which then allows a more flexibly representation of the skin at the cost of a potentially inexact reproduction.

This paper reviews literature on both of the approaches, however as was mentioned previously, the hardware is already defined. This allows us to exclude methods that require movement of lights in space to more degrees of freedom than simple switching on and off of the positioned lights.

II. Literature search methods

As was mentioned in Section I, to successfully conduct a systematic review of the existing literature, a proper and well defined guidelines must be followed. The search strategy for this study is heavily inspired by PRISMA guidelines [29], often used in medicine research. Unfortunately, in it's original way, PRISMA guidelines can't be used for a systematic literature review on computer science domain studies. For example, quality assessment step (assessment of likelihood of random errors, and likelihood of systematic errors and bias), as well as sample size, funding bias and assessment of clinical trials are usually not applicable to studies on computer graphics about facial acquisition.

The search of existing literature (on February 17, 2017) was was performed via Google Scholar, a web search engine that indexes the scholarly literature across an array of publishing formats and electronic databases, such as Springer Publishing, IEEE Xplore Digital Library and ACM Digital Library. A number of keywords were used:

1) facial capture,
2) facial geometry,
3) facial geometry acquisition,
4) facial photogrammetry,
5) facial appearance capture,
6) facial performance capture.

The keywords were combined between each other, but various combinations of keywords in separate groups were used.

Considering the way Google Scholar (as well as separate computer science studies libraries such as ACM and IEEE) works, the amount of results to go through is not realistic. For example, search for keywords ”facial capture” alone yield more than 400000 result hits. It is not rational to go through all of them, so some limits must be applied. In all searches, we review only first 100 hits (10 pages with 10 results each), partly because they are already sorted by relevance, and partly due to empirical experience showing that only first 3-4 pages yield useful results.

All titles and abstracts were read to identify relevant papers. No limitations were made based on publishing country or year of publication, except for the fact that we have reviewed only English-language studies. All papers were viewed equally. After reviewing abstract, it was decided whether the paper fits inclusion criterion mentioned in section I-A and I-B. In case of uncertainty regarding the eligibility of the paper based on the content of the title and abstract, the full text version of the paper was retrieved and evaluated. The full text version of all papers that met the inclusion criteria were retrieved for further assessment and data extraction.

A. Selection process

The selection process is represented in Figure 2. The initial search resulted in 600 hits. The titles and abstracts were screened, and only 128 papers remained. After removal of duplicates, 71 remained. The remaining papers were checked against inclusion criterion and 44 were removed. That leaves us with 27 papers to review. Figure 2 shows the selection process.

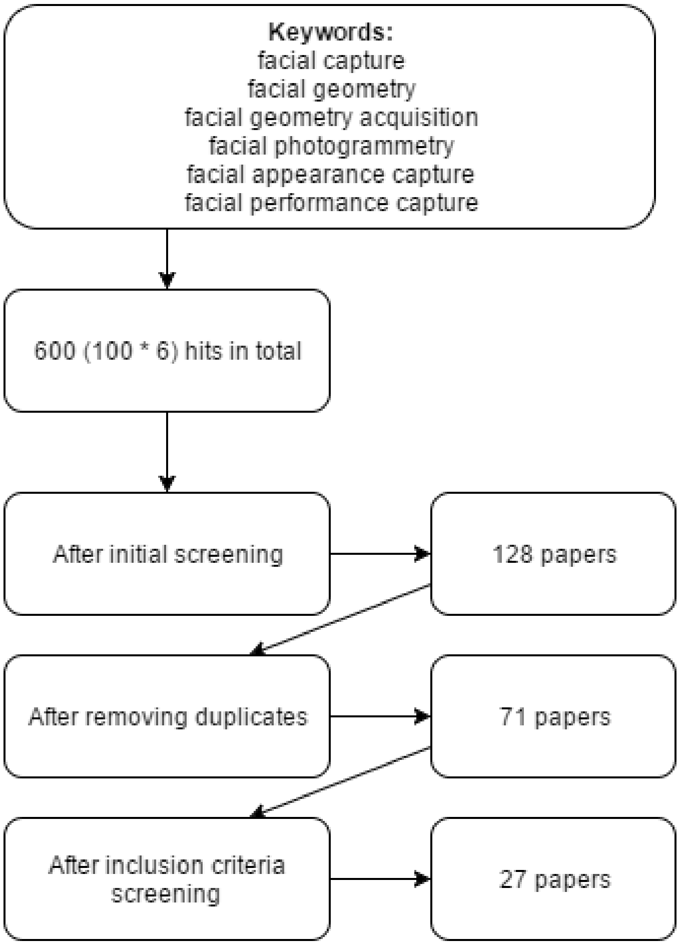

Keywords:
facial capture
facial geometry
facial geometry acquisition
facial photogrammetry
facial appearance capture
facial performance capture
600 (100 * 6) hits in total
After initial screening
128 papers
After removing duplicates
71 papers
After inclusion criteria screening
27 papers

Fig. 2. Selection process of literature

III. Facial geometry acquisition

As was mentioned previously, 3D facial geometry acquisition methods can be split into active and passive approaches.

A. Active approaches

The best current techniques for capturing geometry of a human face are active. The first and study that opened the domain was work on stereo capture of a face augmented with skin markings [34]. With time, the research moved to marker-less geometry capturing, into image-based face modeling. They have provided highly realistic representations because they implicitly capture complex, very hard to reproduce, effects like self-shadowing, inter-reflections and subsurface scattering.

While most methods gather geometry as part of something bigger like getting the whole facial appearance, some methods describe only generation of surface normals, which is equally important for a realistic face model. Even though surface normals can always be gathered from the reconstructed geometry, resolution of such normals is not high enough to show fine details of the skin. Active methods that are focused on normals are more or less all based on the seminal work introducing the technique of photometric stereo [44]. Using three lighting conditions, it is sufficient enough to estimate normals on Lambertian surfaces. However, in the context of facial surface structure, this method does not provide good results.

To get over the fact that face is non-flat, non-convex, causes self-shadowing and that the reflectance can not be accurately described as Lambertian, various ideas were proposed. Weyrich et al. [42] rely on a light stage (figure 3) and record 150 various illumination conditions with 16 cameras, temporally multiplexed over 25 seconds. Even though the results are of very high quality, the exposure time requirement is problematic and since it is impossible for a person to not have at least micro-movements during such long capture process, the captured data is often inaccurate, requiring over a minute to complete full capture consisting of thousands images. Other methods tried to overcome this issue by reducing the illumination condition count (and effectively reducing the needed time for capture) [27] and using optical flow [20] to align captured data [43]. Reducing illumination count does not completely remove the problem of micro-movements, and optical flow does not work too well for anything more than micro-movements. Some authors suggest using spectral multiplexing of illumination conditions instead [39] (figure 4 or apply a white makeup [24] to cancel the spectrally dependent attenuation of the skin.

Further work on the subject yielded study by Ma et al. [27], which presented a system for acquiring high-quality 3D model and normals using polarized gradient-based illumination to generate high-resolution 3D reconstructions. The same approach was advanced by Ghosh et al. [15], in which an improvement was proposed for a more expressive facial reflectance model from just 20 photographs captured from a single viewpoint. It also provided a way to model layered facial reflectance, by acquiring multiple layers, such as specular reflectance, single scattering, and shallow and deep subsurface scattering. Combining layers together yield ultra-realistic face models. At this moment, it is still considered state-of-the-art approach for capturing face geometry. More recent studies focus on using image-based modeling for facial performance capture, improving speed, robustness in lighting conditions and ease-of-use, but the geometry and texture acquisition algorithm is the same.

B. Passive approaches

In contrast to active methods, passive techniques have the advantage of non-intrusiveness, capturing what's observed. Typically these methods require only a single frame to estimate the structure but usually provide less accurate results. Beeler et al. 2010 [2], presented a passive stereo vision system that computes the 3D geometry of the face with reliability and accuracy on a par with a laser scanner or a structured light system. In practical terms, that means equal to active systems performance while attaining the advantages of passive system. However, it makes assumption of constant omni-directional illumination, thus limiting it to studio environments. Later that method was extended to arbitrary lighting by estimating the environment map [45].

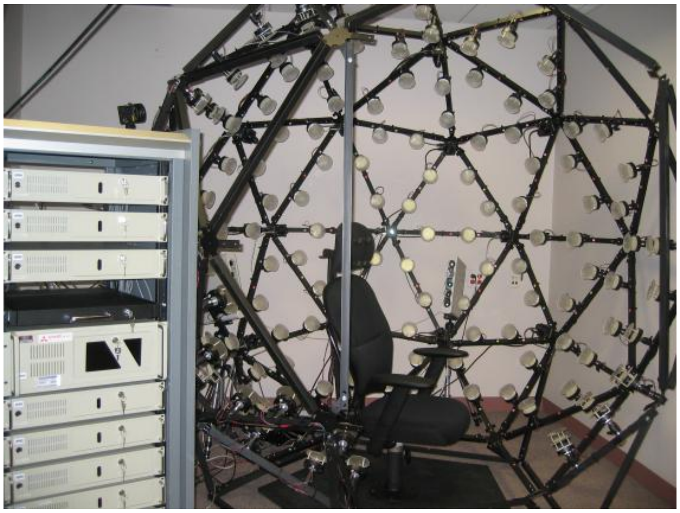


Fig. 3. The face-scanning light dome, consisting of 16 cameras, 150 LED light sources, and a commercial 3D face-scanning system, all used by Weyrich et al. [42]

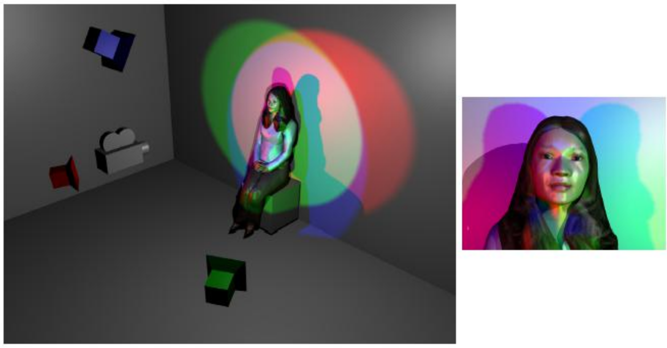


Fig. 4. Acquisition setup used by Vogiatzis et al. [39]. The subject stands in front of three lights of different frequencies and a video camera. The frame is shown on the right.

C. Conclusion

To summarize, active approaches can generally acquire geometry and normals at extremely high resolution, but usually require special setups, time multiplexing and sometimes even makeup. In comparison to active methods, passive approaches do not require any additional hardware and just ”observe” the scene. However, all passive methods have a core problem of separating shading and texture. Such problem is ill-posed for a single shot techniques and passively, can be overcame only with temporal information [3].

A number of critical dimensions was identified for facial geometry acquisition that would be of most importance for the second part of the master thesis. Overview comparison of the listed papers can be found in table I

IV. Facial appearance capture

A. Image-based approaches

The main technique behind image-based methods is the weighted interpolation of captured images in different illumination settings. Contrary to parametric approaches, these captured images are used directly for rendering and are not dependent on the facial geometry. Since photographs essentially capture all real-world effects of human skin, the quality of the rendering, with enough image data available, is incredibly realistic. Another plus of this approach is that rendering images usually does not require extensive computation. However, this approach also brings cons, such as challenging artistic control and editing, and dependence on the high number of captured data to provide realistic result, making image-based methods very data sensitive.

For image-based techniques, only the observed end-result of light-face interaction is relevant, essentially allowing to consider face a ”black box” which can be fully described by input and output. Reflectance field is a way to mathematically describe the incoming and outgoing light rays at the face surface. Concept of a reflectance field in the context of facial appearance capture was proposed by Debevec et al. [7] and since then sparked a large body of research and had an incredible impact on the movie industry. Reflectance field allowed a post-capture relighting of an actor, marking an important step towards the fully virtual actor.

The original problem of reflectance field is described with nine degrees of freedom: time, four degrees of incoming light (light entering at a point x_i from a direction $\vec{w_i}$) and four degrees of outgoing light (light entering at a point x_o from a direction $\vec{w_o}$). Sampling all nine dimensions is challenging, and depending on the use case authors reduce the dimensionality by fixing certain dimensions, for example, removing the temporal dimension and considering only static scenes or assuming distant illumination, which means that light hits face at all points equally, thus removing the spatial variability and reducing the dimensionality by two [25].

The latter limitation, combining with capturing the resulting light only from two cameras are the limitations Debevec et al. imposed for their reflectance field method, essentially bringing this problem to only four dimensions. The dimensions were sampled by moving a light to fixed positions around the subject and capturing images with the two cameras, at different viewpoints (figure 5. While thoroughly sampling incoming illumination enables high-quality relighting, the method is still tied to the viewpoints of capturing of outgoing illumination. To synthesize from novel viewpoints, a dense sample of viewpoints required. This requires excessive amount of data storage and effort.

A number of works appeared as the result of seminal Debevec et al. work, working on reduction of the number of required photographs [37]. However, absolute majority worked on removing the temporal dimension limitation, to allow composition of actor’s performance in a virtual environment and not just static face. Unfortunately, reviewing dynamic

Paper	Capture approach	Possible with current setup	Capture time	Acceptable quality	Surface normals scale
Woodham 1980 [44]	Active	✓	few seconds	X	mesoscale
Weyrich et al. 2006 [42]	Active	X	25 seconds	✓	mesoscale
Ma et al. 2007 [27]	Active	X	few seconds	✓	mesoscale
Wilson et al. 2010 [43]	Active	X	few seconds	✓	microscale
Vogiatzis et al. 2012 [39]	Active	X	near-instant	✓	microscale
Beeler et al. 2010 [2]	Passive	✓	near-instant	✓	microscale
Wu et al. 2011 [45]	Passive	X	near-instant	✓	microscale
Beeler et al. 2012 [3]	Passive	X	near-instant	✓	microscale

TABLE I
COMPARISON OF FACIAL GEOMETRY ACQUISITION METHODS.

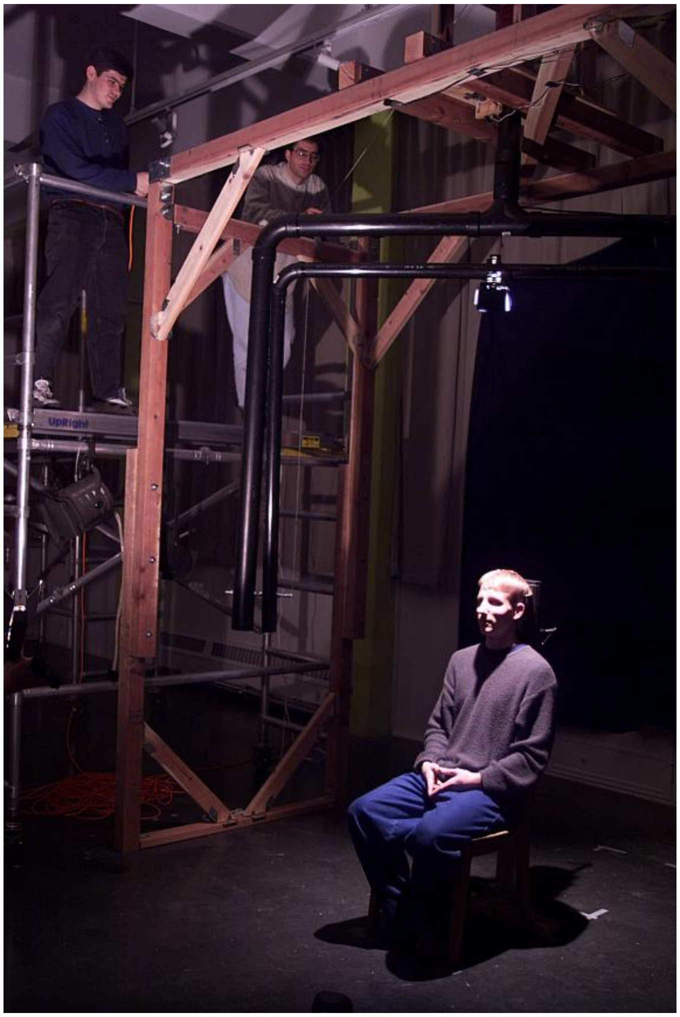


Fig. 5. The face-scanning light stage used by Debevec et al. [7] consists of two-axis rotation system and a directional light source. The outer black bar is rotated about the central vertical axis and the inner bar is lowered one step for each outer black bar's rotation. Video cameras outside the stage record the facial appearance. The inset shows a long-exposure photograph of the light stage in operation.

facial acquisition literature goes beyond the scope of this report.

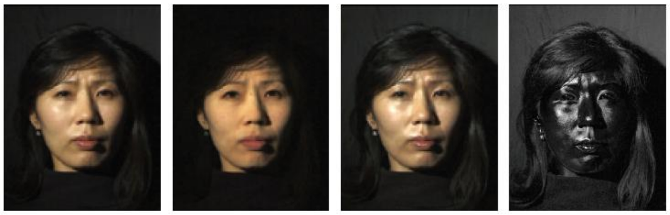


Fig. 6. Diffuse-specular separation using polarization done by Debevec et al. [7]. From left to right: reference image without polarizers, cross-polarization yielding diffuse reflectance only, parallel-polarization yielding diffuse and specular reflectance, substraction of cross-polarization image from the parallel-polarized image yielding the specular reflectance only.

B. Parametric approaches

Parametric approaches have a number of compelling advantages over image-based approaches. First of all, after using the original measurements and captured data to fit into a some kind of parametric appearance model they can be safely discarded. This massively reduces the amount of data that needs to be stored. Secondly, such parametric models usually implicitly support interpolation between and extrapolation beyond original observations. Finally, such models typically allow easier editing, artistic styling and re-use of them.

However, they are also more complex. While image-based approaches can be described as ”black-box”, parametric approaches need to accurately reproduce all of the visible effects seen in the real world. To simplify the problem, most approaches deconstruct the full facial appearance model into several, smaller scale models, each responsible for a separate component. For example, splitting it into surface scattering model and subsurface scattering model. The main challenge of that approach is in separating effects both in the measurement data and resulting model.

In the previously mentioned paper, Debevec et al. proposed idea of splitting facial reflectance into diffuse reflection which is view-independent and specular reflection which is view-dependent. This is a shared challenge of most parametric approaches. The most convenient way to do this is by purely optical means, relying on light polarization and exploiting the fact that specular reflections preserve the polarization state of light, while light that has scattered multiple times rapidly loses polarization (figure 6). Even though polarization requires two captures to get cross- and parallel- polarization for diffuse

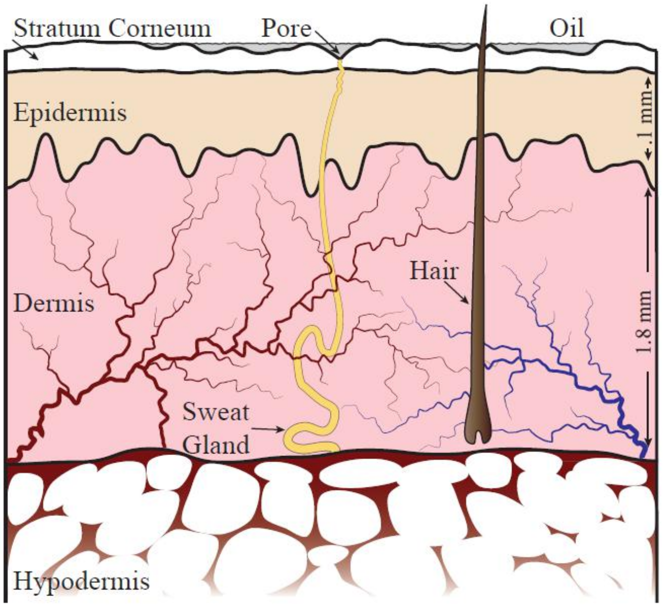


Fig. 7. Diagram showing a decomposition of skin into layers. From top to bottom we see the oil layer mixed with the Stratum Corneum. Below that we have the epidermis layer consisting mainly of melanin pigments responsible for the brownish-yellowish color of skin. Then we have dermis mostly consisting of blood cells with hemoglobin which give the strong red color. Finally there is hypodermis which consists mostly of fat cells that reflect most of the light. [25]

and specular reflections, its simplicity makes this method the preferred choice for diffuse-specular separation. However, since our hardware inventory does not include polarization filters, we will focus only on computational ways to conduct the separation.

Some computational approaches for diffuse-specular separation use the idea that at least on some view/light direction specular reflection vanishes, which allows extracting of images that have only diffuse reflection and eventually the isolation of specular reflectance [7] [42]. Other approaches rely on having a special kind of illumination, for example using spherical harmonic illumination which allows to isolate diffuse reflectance with the first two orders of spherical harmonics [38], or using light of different frequency pattern, for instance binary square wave which with various techniques allows extraction of diffuse reflectance [26]. Finally, some papers focus on optical light transport analysis, using structured light to perform surface and subsurface reflectance separation [30] [32].

1) Surface reflectance: As was mentioned previously, skin is the largest and quite complex human organ consisting of different layers (figure 7) and the most common way to model it's reflectance model is to decompose it into a top layer with a surface reflectance model (also known as BRDF, *bidirectional reflectance distribution function*) and lower layers with a volumetric approximation.

While some papers focus on capturing surface reflectance and using empirical data to create a model [9], majority of recent techniques rely on micro-facet models, which have a strong theoretical basis and have been thoroughly validated [31]. Micro-facet models assume that surface is composed of randomly oriented micro-facets which partly specularly reflect and partly refract light (figure 8). The distribution of micro-facet normals is the parameter that characterizes ”roughness” of these parametric BRDF models. Such distributions are often defined by analytic functions like Blinn [5], Beckmann [1] or the GGx [40].

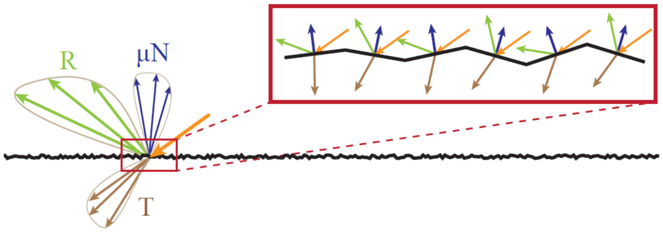


Fig. 8. A diagram showing the various types of interactions between light and a rough dielectric surface. The events depicted in this figure are Surface Reflection (R) and Surface Transmission (T). A zoom-in of the surface which is approximated as a collection of microfacets along with their normals (N) is highlighted on the top. [25]

A device called gonioreflectometer can be used to capture the specular reflectance at every surface point, however in case of skin, and even more of a facial skin, it is not practical, and usually is just approximated, for example by Weymrich et al. [42] with the previously mentioned (section III-A) light dome. After making the diffuse-specular separation, they subtract the diffuse component from the observations and fit the specular reflectance parameters into a number of BRDF models. While Weyrich et al.'s approach produces good results, it takes significant amount of time. Other approach is to use the fact that a single photograph captures the reflectance for many orientations on curved surfaces at once. If we assume that surface properties are unique for a region, it is possible to combine data from all pixels in the same region to obtain measurements with fewer captures. Such segmentation into region wast used by many researches, including Ghosh et al. [15], Weyrich et al. [42] and Fuchs et al. [9]. To achieve full per-pixel parameterization using very few photographs some other methods were introduced, based on capturing statistical properties of reflectance instead of discretely sampling the reflectance function by gradient illumination [12] or by analyzing polarization state of light [14] with Mueller calculus.

2) Sub-surface reflectance: The bottom layers of skin can be described with surface reflectance models, however, they are not expressive enough to account for the soft, translucent appearance of skin due to subsurface scattering. Sub-surface relfectance models (also known as BSSRDF, *bidirection surface scattering reflectance distribution function*) overcome this by modeling skin surface as a boundary of a volumetric participating medium.

The straightforward approach is to actually simulate the light propagation with volumetric path tracing [18]. However,

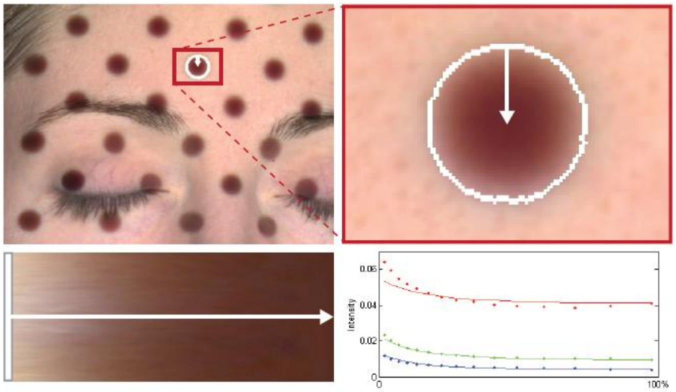


Fig. 9. Ghosh et al. [15] project black dots to sparsely acquire subsurface scattering profiles to fit the parameters. The lower left images shows the reparametrization, which encodes lines from the dot boundary to the center as rows in the image. Hence, each row represents a scattering profile (lower right).

Paper	Diffuse-specular separation	Roughness	Sampling strategy
Marschner et al. 1999 [28]	polarization	object	CS
Georghiades 2003 [11]	none	face	G
Fuchs et al. 2005 [9]	none	region	CS
Ghosh et al. 2009 [13]	polarization	pixel	GI
Ghosh et al. 2010 [14]	polarization	pixel	P
Graham et al. 2012 [17]	polarization	region	CS
Weyrich et al. 2006 [42]	computational	pixel	G
Ghosh et al. 2008 [15]	polarization	region	CS

TABLE II

COMPARISON OF SURFACE REFLECTANCE ESTIMATION METHODS. THE ACRONYMS IN THE SAMPLING STRATEGY COLUMN ARE AS FOLLOWS: G - GONIOMETER, CS - CURVED SURFACE, GI - GRADIENT ILLUMINATION, P - POLARIMETRY.

Paper	Profile from	Fitting	Granularity
Jensen et al. 2001 [22]	beam	optimization	object
Tariq et al. 2006 [35]	stripes	lookup table	pixel
Weyrich et al. 2006 [42]	beam	optimization	face
Ghosh et al. 2008 [15]	black dots	lookup table	region
Zhu et al. 2013 [46]	curvature	direct	pixel

TABLE III

COMPARISON OF SUBSURFACE REFLECTANCE ESTIMATION METHODS.

while the results are very faithful, the simulation is computationally expensive, and the required potentially volumetrically-varying setup is too complex, so the practical models simplify the problem one way or another.

One of the most common ways to simplify the problem is to assume that the medium in which subsurface scattering occurs is homogeneous. That way, only two parameters estimate the diffusion: the absorption coefficient and the reduced scattering coefficient. Jensen et al. [21] introduce the dipole method to computer graphics and also describe how to approximate these parameters from a photograph of a material sample illuminated by normally incident beam of light. Unfortunately, due to setup and time requirements this method is not viable for facial subsurface scattering.

A more practical approach is to use total diffuse reflectance, while doesn't allow you to estimate previously mentioned parameters, allows to compute the reduced albedo [21], which together with the translucency does provide absorption coefficient and reduced scattering coefficient. The measurement of the total diffuse reflectance is trivial and requires just a simple full-on illumination capture. The mentioned translucency has to be estimated from an observed profile. Weyrich et al. [42] captured diffusion profile with a single beam of incident light with a sensor head that could be placed safely on the face of a subject. However, they had to average translucency and total diffuse reflectance, obtaining homogeneous surbsurface scattering parameters.

Less intrusive ways based on illumination proposed projecting various patterns on the face [15] [35] (figure 9), allowing to estimate parameters with per-region resolution, and introducing back multi-layer model, instead of assuming subsurface is homogeneous.

C. Conclusion

To conclude, facial appearance capture methods are split into two categories, image-based approaches and parametric approaches. Image-based approaches have advantage of capturing the full light-skin interaction limited only by the resolution of sampling. Parametric approaches have advantage of being much less data intensive and providing full range of appearance (rendering from any viewpoint and under any illumination), at cost of being not accurate, only approximating the interaction and quite intensive estimation of parameters.

Number of critical dimensions were identified for facial appearance capture, split into image-based (table IV) and parametric approaches, latter of which was split further into into surface (table II) and subsurface scattering (table III) parameter estimation methods.

V. DISCUSSION

Accuracy of Parametric Models. Current parametric models are still not able to show the fully detailed facial appearance. In fact, most existing models rely on simplifications and approximations, which provide not completely faithful results, usually invalid in practice. While these limitations do not prevent reaching acceptable results, there is definitely room for improvements.

Comprehensive human model. Since the ultimate goal is to capture the full facial model, it is worth noting that none of the mentioned in this review papers focus on various remaining challenges, including: capturing eyes [4], teeth, hair [33], the tongue, the rest of the body, and clothing, which all would have to be integrated to acquire the fullest virtual human appearance.

Artistic control. For various applications it is common to have artists change the appearance of virtual characters.

Paper	Relighting	Novel viewpoints	Dynamic appearance	Sampling dimensionality	Illumination basis
Debevec et al. 2000 [7]	✓	✓	X	6D	canonical
Debevec et al. 2002 [8]	X	X	✓	3D	canonical
Hawkins et al. 2004 [19]	✓	✓	✓	7D	canonical
Wenger et al. 2005 [41]	X	X	✓	5D	canonical & hadamard
Chabert et al. 2006 [6]	✓	✓	✓	7D	canonical
Jones et al. 2006 [23]	✓	X	✓	5D	canonical & structured
Tunwattanapong et al. 2013 [38]	✓	X	X	4D	canonical & spherical

TABLE IV
COMPARISON OF IMAGE-BASED FACIAL APPEARANCE CAPTURE METHODS.

Usually it implies editing of texture maps that encode different attributes of the model. While some attribute editing is straightforward, like editing of diffuse texture map, editing of other attributes like subsurface scattering can be difficult to predict even for advanced users. Bridging the gap between parameters that are intuitive to edit while having a physical meaning is an interesting area for future work.

REFERENCES

[1] P. Beckmann and A. Spizzichino. *The scattering of electromagnetic waves from rough surfaces*. 1987.

[2] Thabo Beeler, Bernd Bickel, Paul Beardsley, Bob Sumner, and Markus Gross. High-quality single-shot capture of facial geometry. *ACM Trans. Graph.*, 29(4):40:1–40:9, July 2010.

[3] Thabo Beeler, Derek Bradley, Henning Zimmer, and Markus Gross. Improved reconstruction of deforming surfaces by cancelling ambient occlusion. In *Proceedings of the 12th European Conference on Computer Vision - Volume Part I*, ECCV'12, pages 30–43, Berlin, Heidelberg, 2012. Springer-Verlag.

[4] Pascal Bérard, Derek Bradley, Maurizio Nitti, Thabo Beeler, and Markus Gross. High-quality capture of eyes. *ACM Trans. Graph.*, 33(6):223:1–223:12, November 2014.

[5] James F. Blinn. Models of light reflection for computer synthesized pictures. *SIGGRAPH Comput. Graph.*, 11(2):192–198, July 1977.

[6] Charles-Félix Chabert, Per Einarsson, Andrew Jones, Bruce Lamond, Wan-Chun Ma, Sebastian Sylwan, Tim Hawkins, and Paul Debevec. Relighting human locomotion with flowed reflectance fields. In *ACM SIGGRAPH 2006 Sketches*, SIGGRAPH '06, New York, NY, USA, 2006. ACM.

[7] Paul Debevec, Tim Hawkins, Chris Tchou, Haarm-Pieter Duiker, Westley Sarokin, and Mark Sagar. Acquiring the reflectance field of a human face. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '00, pages 145–156, New York, NY, USA, 2000. ACM Press/Addison-Wesley Publishing Co.

[8] Paul Debevec, Andreas Wenger, Chris Tchou, Andrew Gardner, Jamie Waese, and Tim Hawkins. A lighting reproduction approach to live-action compositing. *ACM Trans. Graph.*, 21(3):547–556, July 2002.

[9] M. Fuchs, V. Blanz, H. Lensch, and H. P. Seidel. Reflectance from images: a model-based approach for human faces. *IEEE Transactions on Visualization and Computer Graphics*, 11(3):296–305, May 2005.

[10] Tom Geller. Overcoming the uncanny valley. *IEEE Comput. Graph. Appl.*, 28(4):11–17, July 2008.

[11] Athinodoros S. Georghiades. Recovering 3-d shape and reflectance from a small number of photographs. In *Proceedings of the 14th Eurographics Workshop on Rendering*, EGRW '03, pages 230–240, Aire-la-Ville, Switzerland, Switzerland, 2003. Eurographics Association.

[12] Abhijeet Ghosh, Tongbo Chen, Pieter Peers, Cyrus A. Wilson, and Paul Debevec. Estimating specular roughness and anisotropy from second order spherical gradient illumination. *Computer Graphics Forum*, 28(4):1161–1170, 2009.

[13] Abhijeet Ghosh, Tongbo Chen, Pieter Peers, Cyrus A. Wilson, and Paul Debevec. Estimating specular roughness and anisotropy from second order spherical gradient illumination. In *Proceedings of the Twentieth Eurographics Conference on Rendering*, EGSR'09, pages 1161–1170, Aire-la-Ville, Switzerland, Switzerland, 2009. Eurographics Association.

[14] Abhijeet Ghosh, Tongbo Chen, Pieter Peers, Cyrus A. Wilson, and Paul Debevec. Circularly polarized spherical illumination reflectometry. *ACM Trans. Graph.*, 29(6):162:1–162:12, December 2010.

[15] Abhijeet Ghosh, Tim Hawkins, Pieter Peers, Sune Frederiksen, and Paul Debevec. Practical modeling and acquisition of layered facial reflectance. *ACM Trans. Graph.*, 27(5):139:1–139:10, December 2008.

[16] Paul Graham, Graham Fyffe, Borom Tonwattanapong, Abhijeet Ghosh, and Paul Debevec. Near-instant capture of high-resolution facial geometry and reflectance. In *ACM SIGGRAPH 2015 Talks*, SIGGRAPH '15, pages 32:1–32:1, New York, NY, USA, 2015. ACM.

[17] Paul Graham, Borom Tunwattanapong, Jay Busch, Xueming Yu, Andrew Jones, Paul Debevec, and Abhijeet Ghosh. Measurement-based synthesis of facial microgeometry. In *ACM SIGGRAPH 2012 Talks*, SIGGRAPH '12, pages 9:1–9:1, New York, NY, USA, 2012. ACM.

[18] Pat Hanrahan and Wolfgang Krueger. Reflection from layered surfaces due to subsurface scattering. In *Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '93, pages 165–174, New York, NY, USA, 1993. ACM.

[19] Tim Hawkins, Andreas Wenger, Chris Tchou, Andrew Gardner, Fredrik Göransson, and Paul Debevec. Animatable facial reflectance fields. In *Proceedings of the Fifteenth Eurographics Conference on Rendering Techniques*, EGSR'04, pages 309–319, Aire-la-Ville, Switzerland, Switzerland, 2004. Eurographics Association.

[20] Berthold K.P. Horn and Brian G. Schunck. Determining optical flow. *Artificial Intelligence*, 17(13):185 – 203, 1981.

[21] Henrik Wann Jensen and Juan Buhler. A rapid hierarchical rendering technique for translucent materials. *ACM Trans. Graph.*, 21(3):576–581, July 2002.

[22] Henrik Wann Jensen, Stephen R. Marschner, Marc Levoy, and Pat Hanrahan. A practical model for subsurface light transport. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '01, pages 511–518, New York, NY, USA, 2001. ACM.

[23] A. Jones. *IET Conference Proceedings*, pages 127–133(6), January 2006.

[24] M. Klaudiny and A. Hilton. High-detail 3d capture and non-sequential alignment of facial performance. In *2012 Second International Conference on 3D Imaging, Modeling, Processing, Visualization Transmission*, pages 17–24, Oct 2012.

[25] Oliver Klehm, Fabrice Rousselle, Marios Papas, Derek Bradley, Christophe Hery, Bernd Bickel, Wojciech Jarosz, and Thabo Beeler. Recent advances in facial appearance capture. *Comput. Graph. Forum*, 34(2):709–733, May 2015.

[26] B. Lamond, P. Peers, A. Ghosh, and P. Debevec. Image-based separation of diffuse and specular reflections using environmental structured illumination. In *2009 IEEE International Conference on Computational Photography (ICCP)*, pages 1–8, April 2009.

[27] Wan-Chun Ma, Tim Hawkins, Pieter Peers, Charles-Felix Chabert, Malte Weiss, and Paul Debevec. Rapid acquisition of specular and diffuse normal maps from polarized spherical gradient illumination. In *Proceedings of the 18th Eurographics Conference on Rendering Techniques*, EGSR'07, pages 183–194, Aire-la-Ville, Switzerland, Switzerland, 2007. Eurographics Association.

[28] Stephen R. Marschner, Stephen H. Westin, Eric P. F. Lafortune, Kenneth E. Torrance, and Donald P. Greenberg. *Image-Based BRDF Measurement Including Human Skin*, pages 131–144. Springer Vienna, Vienna, 1999.

[29] David Moher, Alessandro Liberati, Jennifer Tetzlaff, Douglas G. Altman, and The PRISMA Group. Preferred reporting items for systematic

reviews and meta-analyses: The prisma statement. *PLOS Medicine*, 6(7):1–6, 07 2009.

[30] Shree K. Nayar, Gurunandan Krishnan, Michael D. Grossberg, and Ramesh Raskar. Fast separation of direct and global components of a scene using high frequency illumination. *ACM Trans. Graph.*, 25(3):935–944, July 2006.

[31] Addy Ngan, Frédo Durand, and Wojciech Matusik. Experimental analysis of brdf models. In *Proceedings of the Sixteenth Eurographics Conference on Rendering Techniques*, EGSR '05, pages 117–126, Aire-la-Ville, Switzerland, Switzerland, 2005. Eurographics Association.

[32] Matthew O'Toole, Ramesh Raskar, and Kiriakos N. Kutulakos. Primal-dual coding to probe light transport. *ACM Trans. Graph.*, 31(4):39:1–39:11, July 2012.

[33] Sylvain Paris, Will Chang, Oleg I. Kozhushnyan, Wojciech Jarosz, Wojciech Matusik, Matthias Zwicker, and Frédo Durand. Hair photobooth: Geometric and photometric acquisition of real hairstyles. *ACM Trans. Graph.*, 27(3):30:1–30:9, August 2008.

[34] Frederic Ira Parke. *A Parametric Model for Human Faces.* PhD thesis, 1974. AAI7508697.

[35] Sarah Tariq, Andrew Gardner, Ignacio Llamas, Andrew Jones, Paul Debevec, and Greg Turk. Efficient estimation of spatially varying subsurface scattering parameters.

[36] Angela Tinwell, Mark Grimshaw, Debbie Abdel Nabi, and Andrew Williams. Facial expression of emotion and perception of the uncanny valley in virtual characters. *Comput. Hum. Behav.*, 27(2):741–749, March 2011.

[37] B. Tunwattanapong, A. Ghosh, and P. Debevec. Practical image-based relighting and editing with spherical-harmonics and local lights. In *2011 Conference for Visual Media Production*, pages 138–147, Nov 2011.

[38] Borom Tunwattanapong, Graham Fyffe, Paul Graham, Jay Busch, Xueming Yu, Abhijeet Ghosh, and Paul Debevec. Acquiring reflectance and shape from continuous spherical harmonic illumination. *ACM Trans. Graph.*, 32(4):109:1–109:12, July 2013.

[39] George Vogiatzis and Carlos Hernández. Self-calibrated, multi-spectral photometric stereo for 3d face capture. *Int. J. Comput. Vision*, 97(1):91–103, March 2012.

[40] Bruce Walter, Stephen R. Marschner, Hongsong Li, and Kenneth E. Torrance. Microfacet models for refraction through rough surfaces. In *Proceedings of the 18th Eurographics Conference on Rendering Techniques*, EGSR'07, pages 195–206, Aire-la-Ville, Switzerland, Switzerland, 2007. Eurographics Association.

[41] Andreas Wenger, Andrew Gardner, Chris Tchou, Jonas Unger, Tim Hawkins, and Paul Debevec. Performance relighting and reflectance transformation with time-multiplexed illumination. *ACM Trans. Graph.*, 24(3):756–764, July 2005.

[42] Tim Weyrich, Wojciech Matusik, Hanspeter Pfister, Bernd Bickel, Craig Donner, Chien Tu, Janet McAndless, Jinho Lee, Addy Ngan, Henrik Wann Jensen, and Markus Gross. Analysis of human faces using a measurement-based skin reflectance model. *ACM Trans. Graph.*, 25(3):1013–1024, July 2006.

[43] Cyrus A. Wilson, Abhijeet Ghosh, Pieter Peers, Jen-Yuan Chiang, Jay Busch, and Paul Debevec. Temporal upsampling of performance geometry using photometric alignment. *ACM Trans. Graph.*, 29(2):17:1–17:11, April 2010.

[44] Robert J. Woodham. Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 19(1):191139–191139–, 1980.

[45] C. Wu, B. Wilburn, Y. Matsushita, and C. Theobalt. High-quality shape from multi-view stereo and shading under general illumination. In *CVPR 2011*, pages 969–976, June 2011.

[46] Y. Zhu, P. Garigipati, P. Peers, P. Debevec, and A. Ghosh. Estimating diffusion parameters from polarized spherical-gradient illumination. *IEEE Computer Graphics and Applications*, 33(3):34–43, May 2013.