Real-time Markerless Facial Motion Capture of Personalized 3D Real Human Research

Zheng-Dong Hou¹, Ki-Hong Kim²*, David-Junesok Lee³, Gao-He Zhang⁴

¹Ph. D student, Department of Visual Contents, Dongseo University, Korea
²*Professor, Department of Visual Animation, Dongseo University, Korea
³Assistant professor, Department of Visual Contents, Dongseo University, Korea
⁴Master D. student, Department of Visual Contents, Dongseo University, Korea
E-mail: 1207711532@qq.com, *khkim@g.dongseo.ac.kr, jsdavid88@g.dongseo.ac.kr, zhanggaohe1005@gmail.com

Abstract

Real human digital models appear more and more frequently in VR/AR application scenarios, in which real-time markerless face capture animation of personalized virtual human faces is an important research topic. The traditional way to achieve personalized real human facial animation requires multiple mature animation staff, and in practice, the complex process and difficult technology may bring obstacles to inexperienced users. This paper proposes a new process to solve this kind of work, which has the advantages of low cost and less time than the traditional production method. For the personalized real human face model obtained by 3D reconstruction technology, first, use R3ds Wrap to topology the model, then use Avatary to make 52 Blend-Shape model files suitable for AR-Kit, and finally realize real-time markerless face capture 3D real human on the UE4 platform facial motion capture, this study makes rational use of the advantages of software and proposes a more efficient workflow for real-time markerless facial motion capture of personalized 3D real human models, The process ideas proposed in this paper can be helpful for other scholars who study this kind of work.

Keywords: Real-time facial motion capture, Real human, Markerless, Animation

1. Introduction

Facial motion capture is a part of motion capture, and synthesizing them on virtual faces becomes a very challenging task. Realistic facial animation has a variety of applications such as movies, games, or emotional interfaces. Motion capture recordings of real people are often used to provide realistic motion to virtual faces. Animation controls need to be established to transfer the captured motion to the virtual face. To mimic the full nature of the captured motion, these animation controls need to be carefully planned beforehand, which costs
a lot of time and money [1]. In recent years, the development of real-time markerless face capture in the field of computer graphics has been very mature, and many methods for real-time facial animation capture have been developed for 3D real humans. E Berson, C Soladie, N Stoiber developed a facial animation modification and refinement system. Taking blend shape animations as input, such as raw motion capture animations, this system successfully filters and enhances animations in real-time regardless of the input frame rate, but depends heavily on the data used during training, the amount and type of data (inevitably with the style of the animator who made it) is likely to strongly influence the animation[2]. Intyanto, G. W., D. A. F. Yuniarti, and A. S. Pawening Using a webcam combined with Blender to achieve markerless facial motion transformation. Although simple and easy to implement, this approach suffers from biased animation effects due to the distance between the object and the camera, capturing the light around the object, and training the system on the method [3]. Another way to capture faces in real-time is to use the Make Blend-Shape method. 3D modelers often need hundreds of Blend-Shape targets for a single character to achieve the widest range of facial expressions. Automating the process of 3D facial expression animation is no easy task, and a lot of research and development has been done to help reduce animators' workload. One of the efforts to reduce the labor and time of developing Blend-Shape is to use motion capture data.

Latoschik M E, Kern F, Stauffert J P Retopology and polygon count reduction and UV mapping using R3dS Wrap in their work to generate realistic and personalized 3D models[4]. Murphy C, Mudur S, Holden D (2021) Wrapping high-resolution scan data into a low-poly base mesh using the R3DS Wrap commercial solution. Wrapping to the base mesh ensures that both the restored mesh and texture have the same topology[5]. Murphy A P, Leopold D A Importing raw 3D model data for each animal into R3DS Wrap in a study testing the rich neural specialization of primate faces, Choose a neutral base model, align the data roughly with the standard base model and topology[6]. At present, R3ds Wrap is widely used in video games and medical neurology, however, it is less used in the field of real-time markerless animation. Avatary is a 3D industrial software developed by FACEGOOD that includes expression modeling, muscle binding, and facial motion capture. It is used for facial expression animation in animation, movies, games, virtual idols, and virtual live broadcasts[7]. This study used Avatary's Auto Face feature, since Avatary is a recently developed software, there are relatively few academic articles.

The goal of this paper is to use existing technologies and software to try to create a specific workflow that can be integrated into a standalone production of a project to achieve real-time markerless facial motion capture animation of 3D real humans. By dividing the 3D real-time human face to capture animation work into 2 main stages: model creation, data acquisition, and application, a new workflow can be defined that can significantly reduce the required resources, especially in time aspects, while guaranteeing a higher quality of the final product. The workflow proposed in this study is mainly aimed at using the 3D real human model obtained by photogrammetry, using Wrap to topology the model, and then using Avatary to create 52 Blend-Shape model files suitable for AR-Kit and realize the 3D real human model in UE4 platform. Real-time markerless facial motion capture. This paper is divided into two main parts. The first part is based on the model topology. After the 3D personalized real human model is obtained, the Wrap software is used to topology the model. The second part focuses on the production of blend shapes and the realization of real-time markerless facial motion capture. This part requires the use of Avatray to create 52 Blend-Shape model files suitable for AR-Kit technology and then realize real-time facial motion capture in the UE4 platform. In conclusion, we analyze and summarize each link between/with/to this technical process. The purpose of this paper is to realize the real-time face capture of the 3D personalized real human through more convenient and lower cost, which has reference significance of the work with the same needs.
2. 3D real human model topology

In 3D modeling, the concept of topology refers to the point-line-surface layout, structure, and connection of a polygonal mesh model. While keeping the appearance and size of the model the same, optimize the arrangement of internal vertices, edges, and faces. The purpose is to make the appearance of the model wiring more clean and regular, improve the working efficiency of modeling, and modify and operate the whole and details of the model faster and more accurately, so as to better reflect the structural characteristics of the object. For the human face model, if the topology conforms to the skeleton and muscle structure of the real human head, the deformation during animation can also be smoother, more natural, and more realistic. Film and game production, as well as research applications, require credible, detailed digital faces. Eric Patterson, Jessica Baron, Devin Simpson (2018) propose a process for automatically localizing dense 2D facial landmarks to topology meshes using images accompanying 3D scans, with the specific goal of creating a model for facial animation[7]. 3D artists go to great lengths to build meshes polygon by polygon, the problem with this approach is that it is very technical and difficult. 3D facial reconstruction using photogrammetry technology is a more concise process, which can quickly construct a real human facial structure and obtain a 3D model. The disadvantage is that the surface of the obtained model is uneven and has many messy curves as shown in Figure 1. This paper chooses to use Wrap to topology the face model, this method is faster than the traditional topology.

![Figure 1. 3D reconstructed face model detail display](image1)

First, create a model node LoadGeom and Basemesh node in Wrap to import the scan model and the preset base model. Then, create a SelectPointPairs node, and create marker points on the two models respectively. In order to obtain better results, add marker points as detailed as possible to the key parts of the face, as shown in Figure 2.

![Figure 2. Add marker result in Wrap](image2)

R3DS Wrap 3 is a commercial tool for transferring shape topology via non-rigid registration. The tool provides initial correspondences in the form of hard constraints. Therefore, the operator only needs to mark the control points accurately. Compared with other topology fabrication methods, the operator does not need to be proficient in the facial muscle structure, nor does it have difficult technical problems. After adding
markers, use the SelectPolygons node to differentiate the faces, highlighting each part of the face model with a unique color. At the same time, the SelectPoints node is added to mark the key parts of the face in the basic model. This link is to improve the accuracy of the topology results, as shown in Figure 3.

![Figure 3. Distinguish and add key points in key parts](image)

After the preparations are completed, add the Blen-Wrapping node for topology, click Compute to start the automatic solution. Observing the model, it can be found that some parts have inaccurate structures, such as the turning of the surface part being too stiff, so add a Brush node and adjust the inaccurate results to obtain the final model file and obtain the model as shown in Figure 4.

3. **Blend shape making and the realization of real-time markerless facial motion capture in the UE4 platform**

Blend-Shape directly controls the displacement of vertices to obtain animation. The traditional production method requires the artist to independently produce each model required, which requires the artist's own technical level and generates a high time cost. This paper proposes to use Avatary to automate the production of the models required by Blend-shape. Avatary's Auto Face function connects two face models by adding control points. The basic principle of the Auto Face function is deformation transfer, which applies the deformation presented by the source triangle mesh to a different target triangle mesh. Deformation transfer computes a set of transformations caused by the deformation of the source mesh, maps the transformations through the correspondence from source to target, and solves the optimization problem to apply the transformations consistently to the target shape. It should be noted that enough control points are added to the key parts of the face, and then clicking Create can quickly generate 52 OBJ model files that can be used in AR-Kit, as shown in Figure 5.

![Figure 5. Add control points (left) and generate Blendshape (right)](image)

After the 52 OBJ model files are made, use the Rigger function of Avatary to import the files into Maya
and create a Blendshape. Using the Rigger function to import Maya is to omit the time to modify the file name. Figure 6 shows the final result of Blendshape, finally exporting the required FBX file.

![Figure 6. Rigger tool (left) and Blend-Shape final result (right)](image)

Finally, the model file needs to be imported to the UE4 platform. It should be noted that the import deformation target is checked, and then the animation blueprint is created, and the real-time link pose node and event blueprint are established to update the animation, and the final animation effect can be observed, as shown in Figure 7.

![Figure 7. Real-time facial motion capture effect display in UE4](image)

Because of the development of mobile technology, the threshold of video presentation has been lowered to the point where video producers can use facial capture technology to create complex video content[8]. These methods are limited by cost and technical personnel and are not universally applicable to any user. Therefore, this paper proposes to use AR-Kit technology to achieve real-time markerless face capture in the ue4 platform. The advantage of this solution is that the software used requires only minimal cost. And it does not rely on the experience of technicians, so it can be applied to most users.

4. Conclusion

For real-time markerless facial motion capture animation, there is currently no satisfactory industrial process. For this, this paper proposes a new workflow, using Wrap for topology and correction of 3D models, Avatary's AutoFace for generating blend deformation files, two The automation of this software greatly reduces the time cost and technical cost of the work, and the virtual real human face model made in this way can realize real-time markerless facial motion capture in standard game engines and VR frameworks. The workflow proposed in this paper is suitable for the vast majority of ordinary users and provides help for scenarios that require real-time markerless facial motion capture of 3D real humans, such as VR, virtual video conferences, virtual concerts. This study hopes that this work will open up a new avenue for future research in
this field. The goals achieved in this thesis include technologies, software, and workflows associated with scanning to create 3D digital models of real humans and markerless facial animation capture that can be inserted into the production pipeline of virtual reality characters in a sustainable manner. Therefore, this paper aims to identify the ability to significantly reduce the time and skilled personnel required for real-time facial animation capture of 3D real humans, thereby reducing production costs. Although the process proposed in this paper has achieved clear advantages in time cost and technical personnel cost, there are still some limitations. This process does not always capture the animation of the eye part well, because the 3D face model obtained by photogrammetry cannot generate the eye part, Therefore, the eyeball part needs to be produced separately, and we will continue to focus on the generation of eyeball animation in the future.

References


