Preprocessing of 3D scanned images for facial animation on the basis of realistic acquisition

Damian Pęszor*, Andrzej Polański† and Konrad Wojciechowski†

*Institute of Informatics, The Silesian University of Technology
†Polish-Japanese Institute of Information Technology

Abstract. Facial animation done by performance capture techniques is a subject of growing importance while the quality of graphics and animation in entertainment industry is constantly improving. Most realistic results in facial animation can be obtained using realistic acquisition of human face by various three-dimensional scanning devices. Such registration, however, is prone to errors related to specifity of scanned area, therefore some preprocessing is needed so the obtained model could be used. The aim of this paper is to present typical, face-specific issues as well as solutions related to preprocessing of mesh constructed by scanning techniques. Mesh traversing is applied to reduce number of noisy data related to light dispersion and reflection. Non-manifold edges and vertices are corrected on basis of specifity of studied area, strips of noisy triangles typical for hair are removed and holes typical for chin-neck part of model are filled. Resulting mesh represents single, continuous surface without non-manifold edges or vertices and with hair-related noise removed. Although model after preprocessing is ready for animation, future study to minimize data not related to facial expressions might be needed.

Keywords: facial animation face scans preprocessing performance capture
PACS: 89.20.Ff

INTRODUCTION

Performance capture is a way of acquiring information about deformation of surface of human face during facial expressions by tracking reflective markers placed on actor's skin. This technique, however, is still refined so it could became fully automatic instead of just providing clues for animators who use skeleton system to propel the animation.

To achieve the most realistic results it is essential to use an appropriate model of human face. Although there are artists capable of creating a very realistic model of the face, using a model acquired from real person is still undeniably an approach that guarantees the highest degree of realism. Additionally, in case of performance capture being done by same actor as the one used to obtain the model, his mimicry will be always best represented by his own face.

Obtaining an accurate model of human face is mostly done using 3D scanning techniques employed by 3D scanners, depth cameras or similar devices. The technology used to acquire mesh of scanned object is not perfect and neither are the conditions under which acquisition is performed, therefore this approach is prone to some specific errors that are either non present or negligible in case of different type of surface. Humans can easily notice any error of face representing mesh, especially if it changes over time due to animation, therefore special care needs to be employed to remove erroneous data from acquired model. This work presents approach used for scanned mesh correction and rationale behind it based on year-long study of facial animation in Polish-Japanese Institute of Information Technology.

MATERIALS AND METHODS

Obtaining three-dimensional mesh of human face is mostly done using one or more of three major techniques; stereo vision, structured light, time of flight (TOF). Each of those is prone to different kind of errors which will later affect the construction of mesh. Stereo vision is very sensitive to lens distortion, lighting differences between cameras and ambiguous matching which due to low number of sharp features in face image can lead to errors, especially in areas in which size of observed object is comparable with size of pixel, such as narrow band of hair. Structured light often suffers similar problems. Additionaly, surfaces such as hair, glossy skin or eyes are highly reflective, which leads to incorrect determination of depth. Unwanted light intensifies reflections on some areas, while not on others, translucency of skin results in sub-surface scattering all lead to erroneous mesh. While widely considered as best option, even TOF method is prone to errors from interference from another camera or background lighting.

Models of human face used in this work were obtained by three-dimensional scanning using structured light based scanner, namely 3dMDface System at The Institute of Theoretical and Applied Informatics of Polish Academy of Sciences in Gliwice. Eight different actors were scanned for a total of forty models representing neutral facial expression and some emotions, samples of which are presented in Fig. 1.

FIGURE 1. Meshes as acquired by 3dMDface System. Left: Without texture Right: With texture

Number of different errors is clearly visible as well as parts of scene that are not strictly related to human face mimicry and for purposes of animations are effectively just a noise. This is different problem from having noisy, but consistent structure of mesh (as solved by e.g. [1]) It is worth to note that most of below methods can be used together to minimize amount of traversing through vertices or facets. They are, however, presented sequentially for simplicity.

A major issue with any retrieved model is topological continuity. The specificity of 3D mesh acquisition for the purpose of facial animation indicates that face is central object on the scene and is not occluded by anything different than actor's hair. Based on that, it is clear that major part of acquired mesh will represent face itself, while smaller fragments will be either erroneous or will represent hair or ear, which should not be affected by performance capture. Any algorithm, either one related to preprocessing or animation itself, will be affected through decrease of quality of results or performance in case of topologically discontinued model having additional parts which will not be animated.

First step of preparing acquired model for the purpose of facial animation is removal of surfaces not continuous with main part of mesh. Topologically continuous surface S_i is obtained using any of mesh-traversing algorithms, such as breadth-first search (BFS). The traversing algorithm starts with any triangle t_j of the mesh and traverses through adjacent facets while grouping them into a set representing continuous surface. When no adjacent facets are present, algorithm should start again for any non-grouped facet and continue to do so until every facet of the mesh is grouped. After the grouping is done, triangle count $n(S_i)$ in each group will determine which group is major part of acquired facial model and therefore should be kept while other should be removed, as seen in eq. 1.

$$\arg\max_i f(i) := n(S_i) \tag{1}$$

In case of scanned mesh, regions like eyes, eyelashes, eyebrows, mustache, beard or hair are prone to generate non-manifold edges. This type of error can significantly impair any mesh-based algorithm. Non-manifold edge, can be found by traversing through each facet of the model, while counting every facet adjacent to every edge. Adjacent facet's count for each edge can either be 1 - in case of model's edge, 2 - which indicates correctnes, or more in which case this edge is non-manifold. BFS algorithm is used to obtain sets A of facets connected to each facet of non-manifold edge. Only facets from which sets indexed as shown in eq. 2 are retrieved, should be preserved.

$$\arg\max_i f(i) := n(A_i), \arg\max_j f(j \neq i) := n(A_j) \tag{2}$$

In some cases it is possible, that this strip of triangles will again be connected to main surface. If on both ends this there are non-manifold edges, while other triangles adjacent to original non-manifold edge are correctly adjacent to another triangles at same depth of search, consistency of this strip with model is highly doubtful and entire strip should be removed. Another technique is to check the difference between normals of facets around non-manifold edge. In most cases, two of them will have little angle between each other, while high angle compared to normal from erroneous facet. As a last resort, one should remove facet adjacent to non-manifold edge to which adjacent path is shorter until it can not find another triangle that was not visited earlier, which will prevent long strip of hair being attached to face.

Even with facets related to non-manifold edges removed from the model, some artifacts will still be present. Some can be removed by removing non-manifold vertices. In most cases, this error will be seen as a pyramid representing few recognized vertices of hair strip attached to main surface of the model. In this case it is reasonable to remove set of facets that has smaller area. In case of areas being comparable - set with smaller area of polygon connecting outer edges of it's triangle will need to be cut, so smoother side of non-manifold vertex will be preserved.

Some noisy data around edges of the model will remain intact after removing non-manifolds. Those strips of triangles representing hair or parts of ear might be continuous part of mesh with topologically correct edges and therefore are difficult to find. First method to do so is based on the breadth of connection to main surface. For this purpose the edges of the model are found (see below) and number of vertices each edge is composed of is calculated. For every vertex of an edge, Euclidean distance to every another in given range is calculated. Experiments show that square root of number of vertices is a good range of search for second vertex. Vertices that have smaller distance to base vertex than those with fewer vertices in between, is likely to be a starting point for unwanted strip. Starting from the most distant (in terms of vertices in between) vertex having this characteristic, the shortest path through facets' edges is calculated. Then, the part of mesh that contains edge in between selected vertices can be removed.

Another type of error is related to single vertex being placed with an offset to original model (this is mostly reason of either a reflection or correctly estimated position of hair surface, which was not affecting surrounding facial surface due to it's small size) which creates a triangle connected to others under large angle. To prevent that, one is bound to calculate normals for triangles near the edge of the model (with correction methods explained above, a depth of 3 triangles proves to be sufficient) and compare them. An estimation of curvature can be calculated or simply for each step of depth, average of normals can be used to extrapolate next depth's normal. Our experiments prove that triangles with normals being off from estimate by more than 30° can be considered erroneous and removed.

It is quite common that scanned model contains some gaps due to light dispersion, device's error recognition or part of actor's head not being clearly visible by camera. Simple filling this region with triangles will be perceived as artificial. Many algorithms were developed that deal with this issue, of which [4] proved to provide great results in case of facial hair and chin area, which due to it's size is most demanding. It is important to note, that this algorithm assumes continuous, manifold surface, therefore previous error removal is necessary. [4] assumes that every identified hole is going to be filled, this is not really the case when model is to be used for facial animation. Any additional data that is not strictly related to mimicry will only harm further attempt of animating face and glueing it to skeleton-based model. Therefore after identifying all gaps of the model, one should calculate the length of edge of each gap and resign from filling the hole which has longest border. Next, [4] refers to 3D adaptation of advancing front mesh algorithm [5] which creates a basic patch of triangles to fill the gap, however it's results do not preserve shape of surrounding region. Next, the main part of [4] contribution takes place as desirable normals are computed, triangles are rotated and patch is reconstructed based on Poisson equation. Authors suggest using Harmonic-based desirable normal computing for more planar surfaces since it is faster than geodesic-based approach, in case of relatively high curvature of chin and not so planar curvature of chin-neck area, it is preferable to use geodesic-based normal computing, which will not be as expensive in case of other holes present in the model due to it's small sizes. After successful reconstruction of vertices positions, it is worthwhile to calculate texture coordinates for those vertices as well because assuming any colour not based on texture will result in unrealistic appearance. This can be done using spherical interpolation based on linear combination of vertices surrounding the hole.

It is typical for any 3D model to contain data about normals, however one needs to remember, that those normals are based on same data that led to development of erroneous mesh. It is therefore rewarding to estimate normals for parts of model affected by error removal. There are number of algorithms for normal estimation, many of them focused on preserving sharp features. In case of face, however, proper selection of algorithm for normal estimation should focus on minimizing error in low noise environment, since most of noise is already removed from model, and high-noise methods could damage the quality of fine details like wrinkles.

RESULTS

Meshes of faces acquired using modern scanning techniques were searched for errors resulting from light dispersion, reflection and differences in hair registration caused by dissimilarity in perspective of cameras. An attempt to remove discontinuous surfaces, non-manifolds and strips of noise-influenced triangles was made. Also vertices, texture coordinates and normals in gaps and noisy areas were reconstructed. Depending on quality of acquired mesh, removal of erroneous facets have resulted in up to 13% of vertices and facets of original mesh being deleted of which none were proper parts of face's expressive area. Sample meshes obtained using described approach are shown in Fig. 2.

FIGURE 2. Meshes after removal of erroneous data. Left: Without texture Right: With texture

DISCUSSION

Models subjected to preprocessing are clearly of higher quality. Strips of hair are removed, holes are filled and shape of face is preserved while resulting model is topologically correct. It is worth to note that further work is needed in order to reduce amount of data that is not related to facial expressions, since final model will be subject to continues change during animation and therefore should not have irrelevant data. In this, one-time preprocessing, a complexity of traversing method used in each step can be assumed, with exception of hair strip removal and offseted vertices, which have linear complexity. In case of most models, this focuses on removing parts of neck which should be animated using entire body motion capture. Presence of this additional parts of body may cause problems when face mesh is being connected with body model. Employing a multi-resolution approach for different levels of detail as described in [6] could also be useful in terms of facial animation. This work is however, a step toward proper facial animation with minimized imperfections, which are so clearly visible to the naked eye. While custom models do not have any imperfections related to lighting conditions and surfaces represented, they do induce the feel of artificiality in viewer, therefore it is important to be able to animate models which are burdened with realistic conditions of their acquirement.

ACKNOWLEDGMENTS

This project has been supported by the National Centre for Research and Development, Poland.(project INNOTECH In-Tech ID 182645 "Nowe technologie wysokorozdzielczej akwizycji i animacji mimiki twarzy.")

REFERENCES

1. Fleishman S., Drori I., Cohen-Or D. "Bilateral mesh denoising" in *ACM SIGGRAPH 2003 Papers*, 2003, pp. 950-953.
2. Dey T.K., Li G., Sun J. "Normal Estimation for Point Clouds: A Comparison Study for a Voronoi Based Method" in *Proceedings of the Second Eurographics / IEEE VGTC conference on Point-Based Graphics*, Eurographics Association Aire-la-Ville, Switzerland, Switzerland, 2005, pp. 39-46.
3. Pauly M., Keiser R., Kobbelt L., Gross M. "Shape Modeling with Point-Sampled Geometry" in *Proceedings of ACM SIGGRAPH 2003*, ACM Press, 2003, pp. 641-650.
4. Zhao W., Gao S., Lin H. "A Robust Hole-Filling Algorithm for Triangular Mesh" in *The Visual Computer: International Journal of Computer Graphics*, Springer-Verlag New York, Inc. Secaucus, USA, Vol. 23 Issue 12, November 2007, pp. 987-997.
5. George L.P., Seveno E. "The advancing-front mesh generation method revisited" in *International Journal for Numerical Methods in Engineering*, Volume 37 Issue 21, November 1994, pp. 3605-3619.
6. Szczęsna A. "The Multiresolution Analysis of Triangle Surface Meshes with Lifting Scheme" in *Computer Vision/Computer Graphics Collaboration Techniques, Proceedings of MIRAGE, LNCS*, Springer, Vol. 4418, 2007, pp. 274-282.