Estimation of marker placement based on fiducial points for automatic facial animation

Damian Pęszor, Andrzej Polański and Konrad Wojciechowski

Polish-Japanese Institute of Information Technology

Abstract.
Nowadays, facial animation is achieved using performance capture techniques, of which many is based on the concept of tracking reflective markers placed on human face. Currently available systems either recognize actions which are used to propel final animation or provide information about artificial skeleton which animator is using to animate character. Creation of fully automated system for facial animation requires a way to tie surface on which marker is placed directly to surface of animated model representing different face, so actor's facial expressions could be applied to animated model without need for adjustment. This paper presents a way of estimating marker placement on realistic model of human face obtained using structured light-based 3D scanner. Preprocessed mesh is consequently aligned to neutral model using Trimmed and classic version of Iterative Closest Point algorithm, while fiducial points are found using Anthroface 3D or similar methods related toface recognition until finally marker placement is found on basis of fiducial points and warped neutral model. Found placement estimates are not dependant on actor's face structure and neither are they limited by structure of animated model. Points on animated mesh related to markers' positions show promise of creating a way of automatic animation on basis of surface morphing directly related to movement of markers themselves rather than indirectly through artificial skeleton.

Keywords: face animation marker placement fiducial points
PACS: 89.20.Ff

INTRODUCTION

Performance capture is a set of technologies that enable to perform facial animation using face recorded during facial expressions performed by an actor. Most technologies use reflective markers to track parts of skin surface in order to find actions or transformations of virtual bones which control facial animation.

Given the complexity of human face and possible expressions of it, one of most important parts of successful performance capture is to place reflective markers on face's surface in a way that could effectively represent deformation of face. Oftentimes this potentially crucial step is neglected and markers are placed in possibly dense grid, which is thought to best represent entire surface of face. Practice shows, however, that there are a number of factors that result in issues for even best performance capture systems, and certainly those with lower parameters. Each such system must struggle to distuinguish small markers used for performance capture from other reflections on face which are result of noise and non-perfect lighting conditions. Also, there is a need to distuingish one marker from another if they are close to each other, which limits the density of arrangement. Fast facial expressions as well as overall head movement result in system being unable to track marker and will recognize it as entire new one or even as mistake it for another thus giving significant errors. Some facial expressions may occlude some markers, which again results in tracking problems. All this leads to need of manual repair of recorded data, which in turn becomes less and less reliable as density of the grid grows.

Unlike many other surfaces animated using motion capture techniques, face has structure which restricts possible expressions thus allowing the use of more sophisticated marker placement methods than simple grid. This results in finer details, better modeling of distortion, lower number of markers and finally less time spent on correcting errors due to more analytical approach to facial animation.

It is typical that animated character resembles actor whom animates it. While there can be a lot of difference in terms of texture and details, the anthropometrical structure of face, however, is often quite similar. This is because change in antropometrical distances would require scaling and moving some of triangles. Using fiducial points as a basis for marker placement provides model-independent data by assuming positions of anthropometrical markers instead of unreliable scaling based on grid arrangement.

While marker placement based on characteristic of human face is not uncommon, it is mostly based on intuition and not automatically transferrable to another model leading to further manual labor related to applying distortion found using marker trajectories. The aim of this work is to present a way of finding fiducial points on scanned face, which can be used to place markers for performance capture that can be independent from structure of actor's face, not be dense enough to cause recognition errors, while provide meaningful data.

MATERIALS AND METHODS

Model acquisition and preprocessing

While facial animation can be used to propel many abstract faces, it is impractical to present such algorithms on anything different than real face, since human possibly could take an error in animation for a feature of abstract anatomy. Also, some fiducial points might be detected incorrectly on an abstract face. In case of real face, however, a slight error is clearly visible and anthropometrical data applies.

Facial models used in this work were obtained by 3D 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 actors were scanned for a total of forty models representing neutral facial expression and some emotions.

FIGURE 1. Mesh acquired from 3dMDface System

Due to specifity of structured light-based scanning, few preprocessing steps had to be done in order to improve quality of the model. First, surfaces that were not connected to main part of the mesh were removed, then non-manifold edges and vertices were repaired and strips of erroneous data were deleted. Any holes present due to hair dispersing the light were filled with preservation of overall shape of surrounding area. Normals around changed regions were recalculated, since they were based on erroneous data.

Transformation to canonical pose

Most existing algorithms based on anthropometrical data are used to solve the issue of facial recognition rather than automatic animation. One of most promising algorithms that can be applied to animation, called Anthroface 3D, was developed by Gupta et al. [1]. A basic step of bringing human face to canonical pose is assumed, since scanned face could be rotated impairing proper determination of directions used for finding fiducial points. This step also reduces the area searched for fiducial point, due to assumption of anthropometrical restrictions. This enables to find fiducial points using features which are not globally unique, but are strong enough to be found if search area will be reduced.

[1] suggests aligning mesh to one in canonical pose, since differences between faces are not big enough to prevent that. While Gupta et al. propose iterative closest point (ICP) [2], it had been improved for robustness, since it can easily find local minimum of it's error function instead of global one. For a case of face alignment, it is rewarding to use more robust approach of Trimmed ICP [3]. This approach minimizes the error in case of two aligned models not being exactly same, but sharing overall shape.

As reported by it's developers, TrICP works properly for initial relative rotations of up to 30°, therefore to fully automatize the process, there is a need to cover the rotation space with enough density. In our approach, this problem was solved using standard icosphere algorithm with level of refinement being a parameter. Each vector from icosphere's centre in (0,0,0) to vertex is used as a basis to find rotation used to rotate model before running TrICP algorithm. In case of residual error being not small enough, part of icosphere with biggest errors is rejected, while other part is refined. This process is repeated until acceptable residual error is achieved. TrICP seems to focus on case of two models being partial representations of same object and therefore have same scale, in case of face alignment it is rewarding to also include scale aspect. Both TrICP and ICP use unit quaternion method [4] to obtain absolute rotation and translation between transformed model and original one. Apart from rotation and translation, [4] also includes a way to calculate scale. Using scale improves quality of consequent iterations of TrICP. Although Horn suggests that in case of one model being more accurate than other, either scale calculated using rotation matrix or it's inverse should be used. In this case, however, symmetrical scale have yielded better results. TrICP might cause some problems in case of models being of different resolution because of finding correspondences which in one model represent small part of surface, while in the other majority of vertices is used. Neutral mesh should therefore be prepared with resolution comparable to the one achieved by source of models to be animated.

Fiducial points estimation

Few changes are proposed to [1]. Manual detection of noise tip in neutral model was extended to every fiducial point for which estimation has been implemented as well as further points in between. Search in initial region of 96x96 mm about ICP estimate of nose tip seems to be based on view of canonical pose of human face projected to 2D creating sort of depth map rather than 3D model. This approach can be misleading in case of bulbous, rounded or pinched nasal tip (see [5]) which might make curvature assumptions invalid, especialy in case of high resolution, when curvature estimated on basis of one-ring neighbourhood can be much lower than expected. Different approach was therefore assumed. Region searched for nose tip was determined by placing sphere of radius r with center in position of neutral face's nose tip p_{nt} in aligned mesh composed of set of vertices $V = (v_1, v_2, ..., v_n)$. The radius is calculated as:

$$r = \|p_{nt} - (\sum_{i=1}^{n} v_i)/n\|/2 \quad (1)$$

Similarly, all regions used in searching for fiducial points were defined as a spheres. Next, a gradient based method was used to traverse from vertices closest to centroid of the model to farthest. This allowed to differentiate nose tip (farthest from centroid) from nose bridge and sides. Due to model's centroid position depending on quality of TrICP as well as bordering vertices of the model, it is necessary to establish smaller (5 times, which corresponds to radius of ca. 1cm) sphere in newly found nose tip. Nose tip estimation based on curvature (as suggested in [1]) is then much less prone to errors. Still, however, it is rewarding to use curvature estimation based on more than one-ring neighbourhood. Among methods presented in [6] paraboidal fitting was used to obtain curvatures with two-ring neighbourhood. Obtaining further fiducial points follows similar procedures. Once all points were found using curvature-based or other methods (a range of two-dimensional methods employing texture information as well as three-dimensional algorithms for facial recognition have already been developed and can be used here), neutral model is once again used.

Depending on quality of performance capture system, one might be able to use space between already established fiducial points to increase sampling of face surface thus increasing the quality of animation. It is important to note, that those additional fiducial points need to be distant enough from other points, so system could distinguish markers placed on those points without a chance of recognizing one as the other. Fiducial points on animated mesh should be now aligned to their known correspondences on neutral model using ICP (there is no need for TrICP due to low number of points). This serves to correct the initial pose. After this, neutral model should be warped, so each fiducial point would occupy same coordinates as it's correspondent point on second model. All additional marker points on neutral mesh should be warped so they would remain in their position relative to fiducial points. For every marker point m_j a normal N can be calculated from plane composed of it's neighbouring fiducial points. Marker placement on animated model can be estimated as a vertex v_i such that:

$$\arg\min_i \frac{\|v_i - m_j\|((v_i - m_j) \cdot N)}{\|v_i - m_j\|\|N\|} \quad (2)$$

RESULTS

Marker placement was estimated on face acquired using structured light-based scanning by 3dMDface System. In presented case, positions of 58 markers were estimated, with average distance from manually selected points being equal to 1,3% of the distance from mesh centroid to manually selected point and maximum one being as big as 2,7%. On models acquired during this research, an average error of 2,2% was achieved in case of 58 points. Figure 2 presents found estimates as n-ring neighbourhood of faces with area big enough to be noticeable.

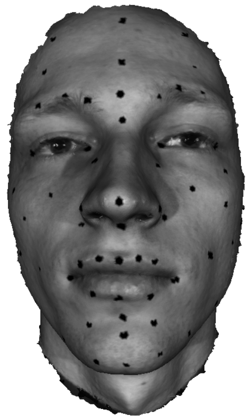

FIGURE 2. Mesh with estimated marker placement

DISCUSSION

Estimates of marker placement found using described approach seems to indicate, that the overall structure of the method shows promise of great results, which seem to be better the more points is used, since the search region becomes smaller estimate becomes more precise. In this study, information about colour as given by texture coordinates for every vertex was not used, however given the number of developments about facial recognition based on images, it is reasonable to assume that smaller errors could be achieved if that knowledge was employed. Future study about deformation based on found characteristic points is also needed to fully utilize results of this method. This approach might however be used as an initial step toward non-skeleton and non-action based fully automatic facial animation.

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. Shalini Gupta, Mia K. Markey and Alan C. Bovik "Anthropometric 3D Face Recognition" in *International Journal of Computer Vision*, Springer US, 2010, Volume 90, Issue 3, pp. 331-349.
2. P. J. Besl, H. D. McKay "A method for registration of 3-D shapes" in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, IEEE, Volume 14 Issue 2, February 1992, pp. 239-256.
3. D. Chetverikov, D. Svirko, D. Stepanov, P. Krsek "The Trimmed Iterative Closest Point algorithm" in *Proceedings of 16th International Conference on Pattern Recognition*, Volume 3, 2002, pp. 545-548.
4. Berthold K. P. Horn "Closed-form solution of absolute orientation using unit quaternions" in *Journal of the Optical Society of America A*, Volume 4, Issue 4, 1987, pp. 629-642.
5. Dean M. Toriumi, Mark A. Checcone "New Concepts in Nasal Tip Contouring" in *Facial Plastic Surgery Clinics of North America*, Volume 17, Issue 1, 2006, pp. 55-90.
6. T. Surazhsky, E. Magid, O. Soldea, G. Elber, E. Rivlin "A comparison of Gaussian and mean curvatures estimation methods on triangular meshes" in *Proceedings of IEEE International Conference on Robotics and Automation*, Vol. 1, 2003, pp. 1021-1026.