STAG: Smart Tools and Applications in Graphics (2018)
M. Livesu, G. Pintore and A. Signoroni (Editors)

An optimized marker layout for 3D facial motion capture

A. D. Will[1,2], J. M. De Martino[1] and J. Bezerra[1]

[1]School of Electrical and Computer Engineering/University of Campinas, Brazil
[2]Adventist University Center of São Paulo, UNASP-HT, Brazil

Abstract

Facial motion capture (Facial Mocap) has gained increasing attention from researchers and professionals from different areas of interest, including entertainment, face-to-face communication, and training. Facial Mocap allows straightforward capture of dynamic behavior of the face from live action, providing data that can be used to drive realistic animation of a 3D virtual face. Facial Mocap is an advantageous alternative to the direct and laborious manipulation of the face model. In particular, marker-based mocap technique acquires three-dimensional facial points trajectories by tracking markers fixed on the face of an actor. However, despite the existence of several empirical facial marker layouts, the ideal positioning of the markers is still an open question. This paper presents an optimization technique to calculate the quantity and positioning facial markers and establish their influences on the polygon mesh based on the correlation of markers in a dense layout. The technique generates an optimized marker layout discarding unnecessary markers and positioning the remaining ones.

CCS Concepts
•***Computing methodologies*** → ***Motion capture;*** *Animation;*

1. Introduction

Realistic facial animation of characters is used in many applications, such as animations for movies, video games, educational multimedia, medicine, among others. The generation of realistic, convincing facial animation is chalenging, especially because we have an accurate sense to perceive even the subtlest facial expressions [Par72]. The manual generation of realistic 3D facial animation requires skilled animators with artistic sensibility and knowledge of facial physiology. Moreover, the manual manipulation of the face model by an animator is laborious and time consuming. Facial Mocap is an alternative that facilitates the generation of realistic, convincing facial animation, mitigating the burden imposed upon the animator. Currently, there are several types of mocap systems, such as magnetic, mechanical and optical. The optical motion capture system is the most commonly used one due to factors, such as flexibility and economy [LZD13]. Optical sytems can be categorized into two different techniques: *markerless* and *marker based*. The *markerless* system allows a more detailed capture of the surface of the subject, but it is more sensitive to the environment and the tracking often is not reliable [GLLG07]. On the other hand, in *marker-based* systems the markers are triangulated by more than one camera, and the correspondence finding across multiple angles can be trivial because the configuration of the markers is sparse, which makes this system typically more robust [LZD13]. Additionaly, marker-based motion capture can be done simultaneously with the capture of the actor body [LZD13], [RGL15].

In body motion captures, the positioning of markers on the actor‘s body is more intuitive and inherently guided by the joints and bones of the human body. However, since facial expressions are mainly produced by muscular movements, the determination of the number of markers and their positioning is neither intuitive nor trivial.

One approach to obtain a detailed capture is to use as many markers as possible to cover most of the actor's face. However, a high number of markers in a dense layout increases the demand for processing power and the probability of marker identification swaps that can occur when the markers are too close to each other. On the other hand, using too few or poorly-placed markers may fail to capture significant movements associated with facial expressions.

Empirical marker layouts have been created and used by researchers and industry [LZD13]. However, these ad-hoc solutions lack solid rationale, hindering generalization and many of them need standardization.

The technique presented in our work, uses data from the positions of the markers during some captures from a dense layout. Through the positioning data we compute the correlations among them and create an optimized marker layout.

The paper is organized as follows: Section 2 reviews related works. Section 3 presents the methodology to establish an optimized facial marker layout. Section 4 shows the optimized layout results. Section 5 demonstrates the application of the results ob-

DOI: 10.2312/stag.20181303

tained and shows how they were introduced in the face of a character. Finally, section 6 concludes and discusses our work.

2. Related works

Marker-based facial motion captures are usually done by optical systems due to the ability to capture even the subtlest movements [KW08] [Men10]. [KW08] argues that we should not use more markers than necessary, as often happens in body captures. He warns that extra markers in facial captures tend to be discarded, and when closed with each other, they can be viewed as a single marker by the capture system cameras. However, the use of few and poorly placed markers results in incomplete and low facial capture quality. Based on these issues, the search for an optimized layout becomes necessary [LZD13], [RGL15].

The optimization problem formulated in this work has similarities to existing works [LZD13], [RGL15]. [LZD13] solve an optimization problem of characteristic control points based on the vertices of the polygon mesh from high resolution facial pose samples. [RGL15] used the vertex positions of the facial mesh of all the frames of a training sequence and through a clustering technique, calculated the optimized layout of markers for a given number of clusters. Both studies compute an optimized layout from the positions of the vertices of the face mesh when facial expressions are applied to them. Differently, in our work, the optimized layout determination is performed using captured data made by real actors.

We also defined a main area of influence in the vertex mesh for each marker of the optimized layout which, through a direct mapping, produces the facial expressions in the character. This differs from many works [RGL15], [DCFN06], [SL14], [PL06] that use a nontrivial process of mapping the mocap data with blendshape parameters.

3. Our Method

The facial marker layout optimization aims to determine a minimum number of markers and their respective positions while maintain a satisfy high degree of realism and intelligibility of the facial expression. To compute the optimized layout, We used a dense marker layout and made motion capture a wide range of facial movements including the following facial regions: eyebrows and forehead, eyes, nose, cheeks, mouth and chin. The positions of each marker along the capture were analyzed a clustered according to their behaviors. For each cluster, a marker was chosen and determined as a final marker of the optimized layout.

3.1. Corpus

For each facial region, a different mocap was performed with the purpose of covering a wide range of movements. Facial expressions containing, for each facial region, the following movements:

Eyebrow and forehead: raised, frowned, raised to the center
Eyes: brow lowered, upper lid raised, cheek raised, lid tightened, slit, eyes closed, blink and wink.
Nose: nose wrinkle, crooked to right and crooked to left.
Mouth: Closed smile, open smile, right cheek inflated, left cheek inflated, blowing, cheeks suck, upper lip raised, lip corner pulled, cheek puffed, lip corner lowered, lower lip lowered, chin raised, lip puckered, lip stretched, lip funneled, lip tightened, lip pressed, lips parts, jaw drop, mouth stretch and lip suck.
Cheeks: Inflated cheeks, inflated right cheek, open smile, closed smile, cheeks sucked, left inner tongue, right inner tongue and blowing.
Chin: Same as mouth.

3.2. Dense marker layout

In a regular facial motion capture we should not use markers close to each other [KW08]. However, in our approach we created a dense layout of markers about one centimeter apart achieving a better coverage of the surface of the face.

In our experiments, we assume that most facial expressions are symmetric, and therefore we placed the markers only on one side and the central line of the face. On the opposite side, we fixed some markers in a non-systematic way only for face identification (Figure 2-a).

As for the markers, we used adhesive artifacts with half-sphere geometry fixed on the actor's face with hypo-allergenic glue. The markers were painted with white matte paint giving a more uniform and diffuse reflection.

3.3. Facial regions

For a more focused analysis, we chose to restrict the investigation of the markers according to the facial region of their location. The facial regions determined in our work were based, but not restricted, to the the regions presented by [SPP*01] (Figure 1) which leans over the facial musculature. A representation of the regions on the face of the actor and their markers can be seen in Figure 2-b, and the numbers of markers in each region have been listed in Table 1.

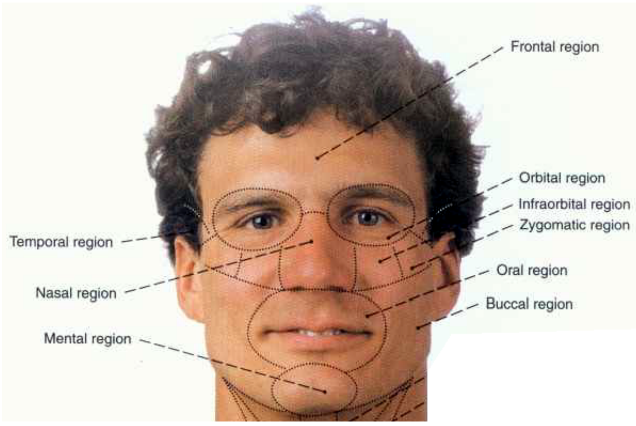

Figure 1: *Regions of head. Adapted from [SPP*01].*

3.4. Motion capturing and processing

For our motion capture we use the commercial *Vicon Cara* system which tracks white or black markers with four 720p HD cameras

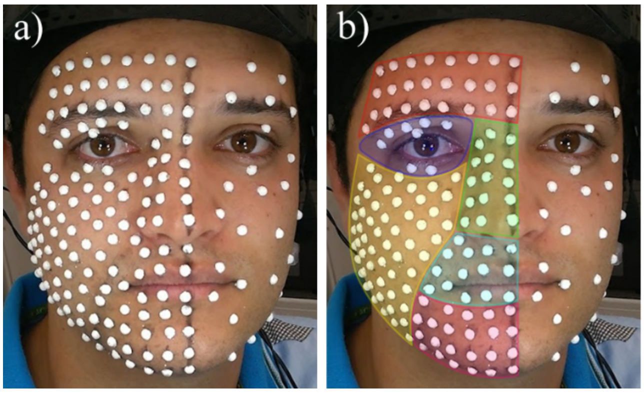

Figure 2: *a) Dense capture layout. b) Representation of facial regions.*

Face region	number of markers
Eyebrow and forehead	25
Eye	8
Nose	15
Cheek	56
Mouth	14
Chin	23

Table 1: *Facial regions and number of markers in each*

with up to 60 fps [vic] (Figure 3). In our captures we use only white markers.

During the sequence of region captures, execution between target expressions and neutral expression (rest) was alternated with a pause of about two seconds in each capture.

Through the *Vicon Cara Post* software, the capture data were verified and processed resulting in the three-dimensional trajectory of each marker along the captures.

3.5. Position matrices

From the processed captured data, the position of the X, Y and Z coordinates for each marker on every 5 frames is obtained creating three different position matrices (PX, PY and PZ) on every facial region.

$$PX = \begin{bmatrix} m_1 px_1 & m_2 px_1 & \ldots & m_M px_1 \\ m_1 px_2 & m_2 px_2 & \ldots & m_M px_2 \\ \vdots & \vdots & \ddots & \vdots \\ m_1 px_N & m_2 px_N & \ldots & m_M px_N \end{bmatrix}$$

$$PY = \begin{bmatrix} m1py1 & m2py1 & \ldots & m_M py1 \\ m1py2 & m2py2 & \ldots & m_M py2 \\ \vdots & \vdots & \ddots & \vdots \\ m1py_N & m2py_N & \ldots & m_M py_N \end{bmatrix}$$

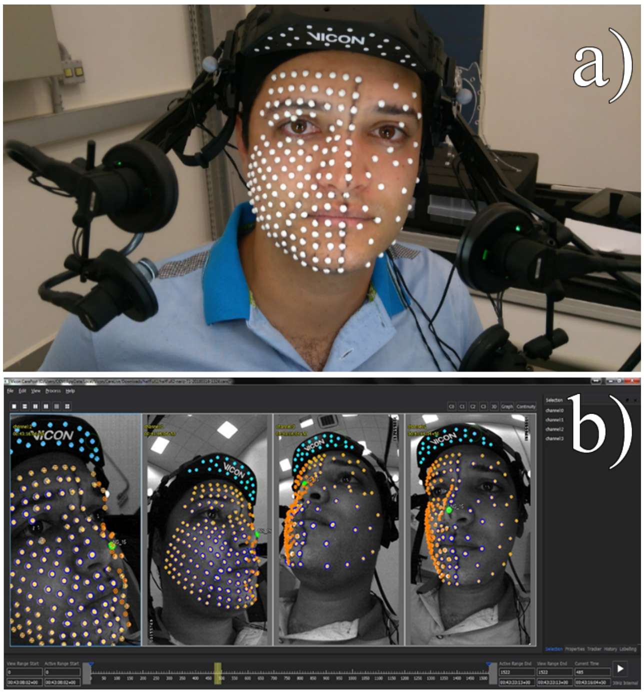

Figure 3: *a) Actor using the Vicon Cara Motion Capture Equipment. b) View of the cameras during the processing of the captured data.*

$$PZ = \begin{bmatrix} m1pz1 & m2pz1 & \ldots & m_M pz1 \\ m1pz2 & m2pz2 & \ldots & m_M pz2 \\ \vdots & \vdots & \ddots & \vdots \\ m1pz_N & m2pz_N & \ldots & m_M pz_N \end{bmatrix}$$

Where m_i is the i^{th} marker and px_j the x-axis coordinate at the j^{th} analyzed frame. M is the number of markers in a particular facial region, and N is the number of observed frames.

3.6. Correlation matrix

To calculate the correlation matrices, first we calculate the covariance from the position matrix of each facial region implementing the following steps [GW00]: extraction of the mean vector, subtraction of the mean, product of each vector by its transpose and extraction of the mean of the matrices. Then, from the covariance matrices, we compute the correlation matrices [cor].

For each facial region, the correlation matrices of each axis are consolidated to a final matrix through the sum of the matrices illustrated in Figure 4.

3.7. Marker clustering

The optimization process is based on the clustering of the markers that have similar behaviors pointed out by the correlation that they have with each other according to their data of movements along

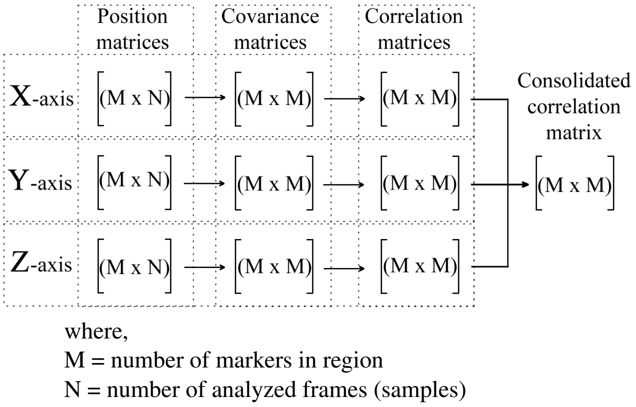

Figure 4: *Illustration of the process of creating the consolidated correlation matrix*

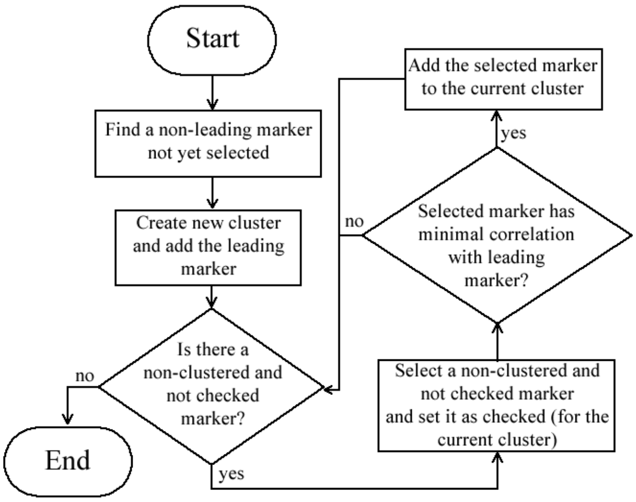

Figure 5: *Clustering algorithm flowchart*

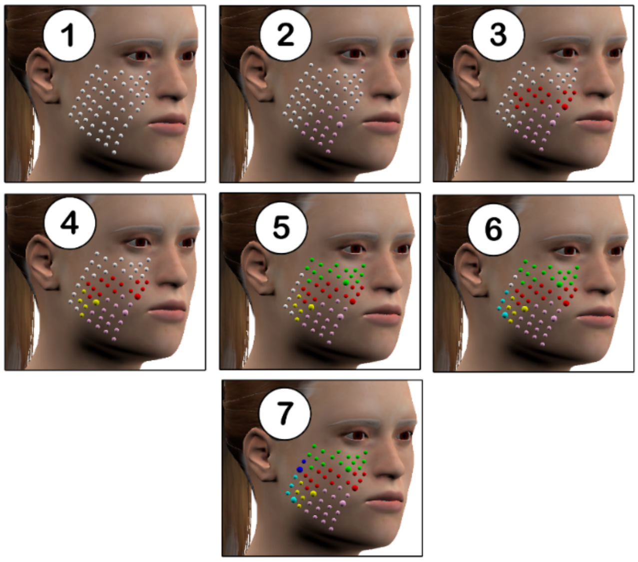

Figure 6: *Example of a clustering sequence*

the captures. For each cluster a representative marker is selected to compose the optimized layout.

The clustering and determination of the representative marker in a given facial region are performed according to the following steps:

1. Create a new cluster and set as its representative marker, a marker which is not yet clustered, with the largest displacement observed throughout the capture.
2. Inclusion of the markers whose correlation coefficient among the representative marker previously defined is below the stipulated threshold based on the mean of the correlations among the markers.
3. If all markers are clustered the process ends, otherwise, go to Step 1.

Figure 5 shows a flowchart of the process of clustering and determining the representative markers with more details.

It is also possible to change the density of the optimized layout according to the need for a greater or lesser degree of capture detail by multiplying the correlation coefficient by a numerical factor.

Only to have a more visual reference, we presented the clusters through spheres on the face of a character delimited by different hues, the markers with the largest diameters represent the representative markers. Figure 6 shows an example of a clustering sequence applied to the cheek region can be seen.

3.8. Final cluster adjustment

Because member markers are clustered to the first representative marker that has minimal correlation with it, there may be situations where this marker can have a higher correlation with a further representative marker created later. For these cases, reallocation adjustments are made. Figure 7 shows the reallocation of markers between groups represented by different hues.

In this way each member marker is analyzed and moved to the group whose representative marker has a higher correlation with it, as shown in Figure 8.

4. Markers layout and clustering results

As shown in Section 3, the captures were taken only on the right side of the face. However, for the presentation of the results, the markers displayed on the character face are mirrored to its opposite side resulting in a symmetrical layout (Figure 9).

4.1. Optimized marker layout

The optimized marker layout is composed of the representative markers of the clusters in each facial region. As already stated, the clusters are created through a Threshold based on the mean of

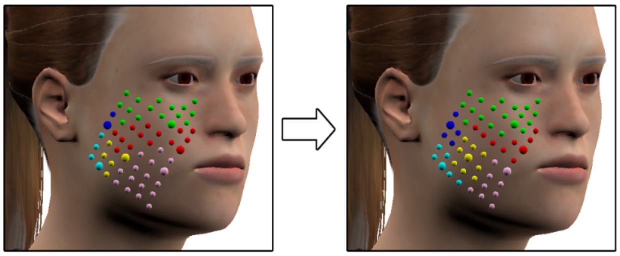

Figure 7: *Example of the clustering adjustments of the cheek region*

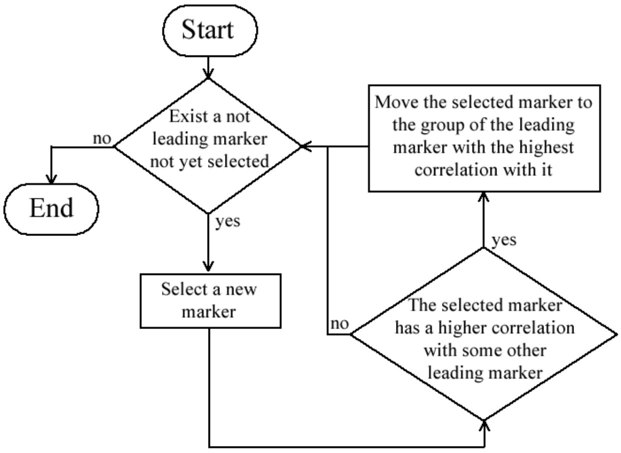

Figure 8: *Clustering adjustment algorithm flowchart*

correlations between markers. However, this threshold can be manipulated by multiplying a factor to get layouts with different refinements for different applications and needs. Figure 10 presents different layouts obtained through the application of different refinement factors.

4.2. Region of main influence

To obtain a representation of the facial expression with the optimized layout, each representative marker influences, with varying

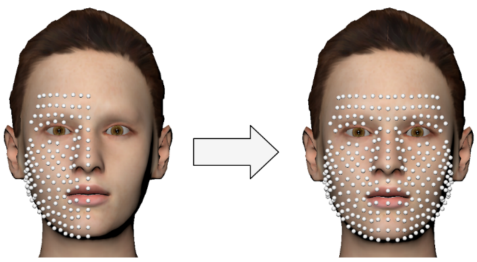

Figure 9: *Markers displayed on the face of the character. Left: original layout captured. Right: Mirrored layout (symmetrical).*

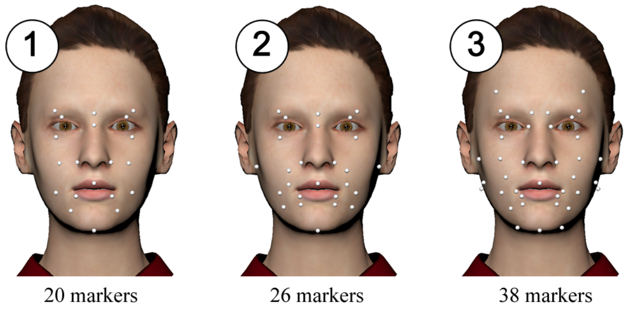

Figure 10: *Optimized marker layouts obtained through refinement factors equals to 0.9, 1.0 and 1.1 respectively*

degrees, a set of vertices on the face of the character according to its cluster area.

However, to achieve a smoother transition, the area of its influence may extend to other clusters and facial regions when analyzed individually.

The cluster refinement (as well as optimized layout) can be modified by changing the refinement factor. Figure 11 shows the main influence clusters obtained through the application of different refinement factors.

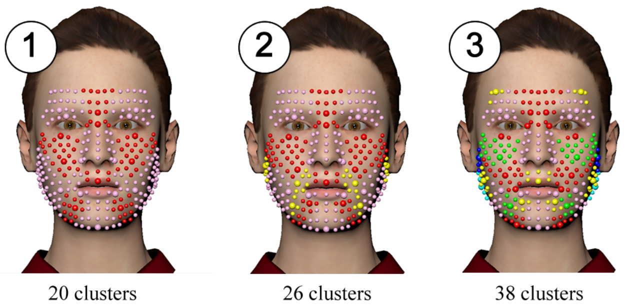

Figure 11: *Main influence regions obtained through refinement factors equals to 0.9, 1.0 and 1.1 respectively*

5. Application example

As an example of application, we inserted some rigs into the face of a character according to the positioning of the markers in the optimized layout obtained through the refinement factor equal to 1.1 (Figure 10-3). In addition to the 38 bones defined by the markers of this layout, other bones were inserted to aid in the influence of deformations as shown in Figure 12.

Using the optimized layout, we executed motion capture of various facial expressions through the Vicon Cara system. The data of the captures were then processed and positioned next to the character face as can be seen in Figure 13.

Joints were defined for the character face and placed at the same location of the Mocap controllers which followed the optimized marker layout positions. These joints were bonded with character

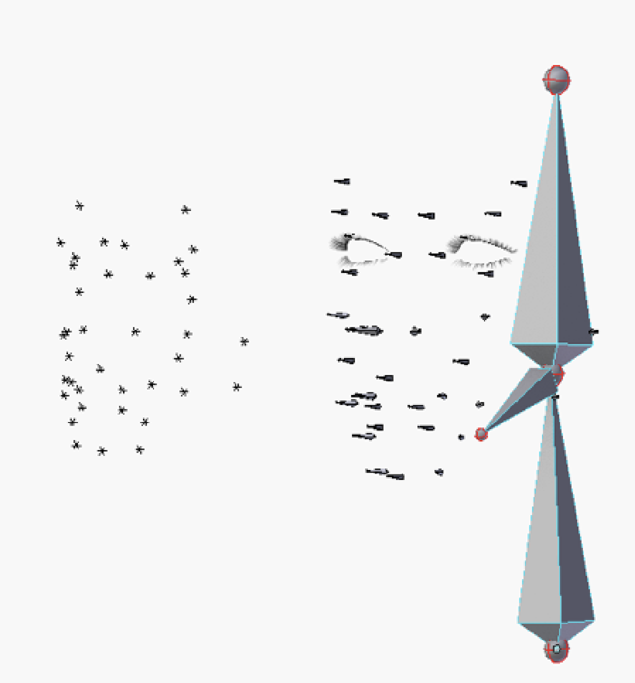

Figure 12: *Main Bones and Auxiliaries*

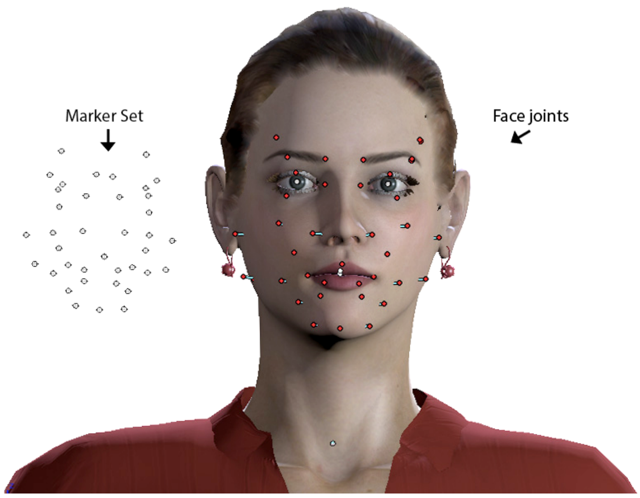

Figure 13: *Optimized marker layout and face joints*

face geometry so that it can deform a specific area on the deformation map. Each joint influence has a similar region of influence of the marker layout.

The regions of influence of each marker on the character face were guided by the position and area of the clusters obtained from the optimized layout shown in Figure 11-3. The capture markers were then connected to the rigs through a *Child of* constraint, so that the rigs might perform the same translations determined by their associated markers.

The regions of influence are based on a set of colors, where each color represents the percentage of deformation when the joint is moved (Figure 14). Each joint of the rig generates a map of deformation in the geometry of the character and that region is based on the main influence regions as seen in Figure 11. Each created cluster indicates the position and the reference for the influence painting in the character. The 38 optimized markers (figure 10), generate the same number of motion capture points used to record the animation. These markers are used to move the joints face in 3d characters and to paint influences in geometry head also.

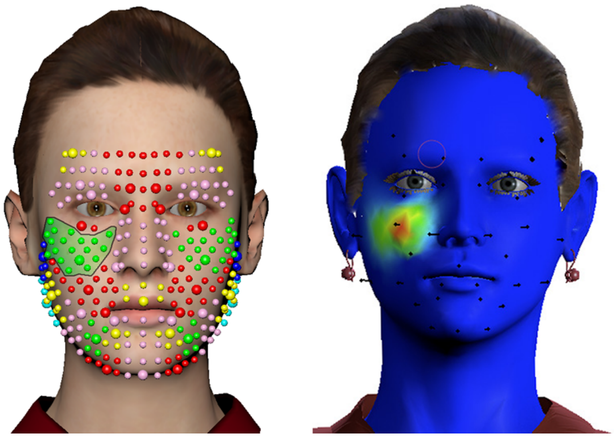

Figure 14: *Marker set comparison influence deformation map*

Figure 15 shows some comparisons of facial expressions captured by an actor and produced by the character according to the results and procedures presented in our work.

6. Discussion and Conclusions

We propose an optimization approach of markers layout for 3D facial motion capture based on the correlation of markers through captures using a dense layout. For this purpose, we analyzed the positions of the markers along specific captures for different facial regions and clustered the markers that showed similar behaviors pointed out by correlation calculations among themselves.

The markers, originally from the dense layout, in each cluster formed in this process, are replaced by a single marker at the position of the one that showed the greatest displacement observed throughout the capture. In addition to the optimized layout, the clusters formed can help in determining the region of influence in the mesh of vertices of the character face.

We also presented the possibility of manipulating the level of detail of the optimized layout by changing a refinement factor in the clustering algorithm. For further work, it is planned to create an automatic polygon mesh manipulation system considering the position of the markers and their main area of influence.

References

[cor] Numpy Manual v1.13 numpy.corrcoef. https://docs.scipy.org/doc/numpy-1.13.0/reference/generated/numpy.corrcoef.html. Accessed: 2018-04-16. 3

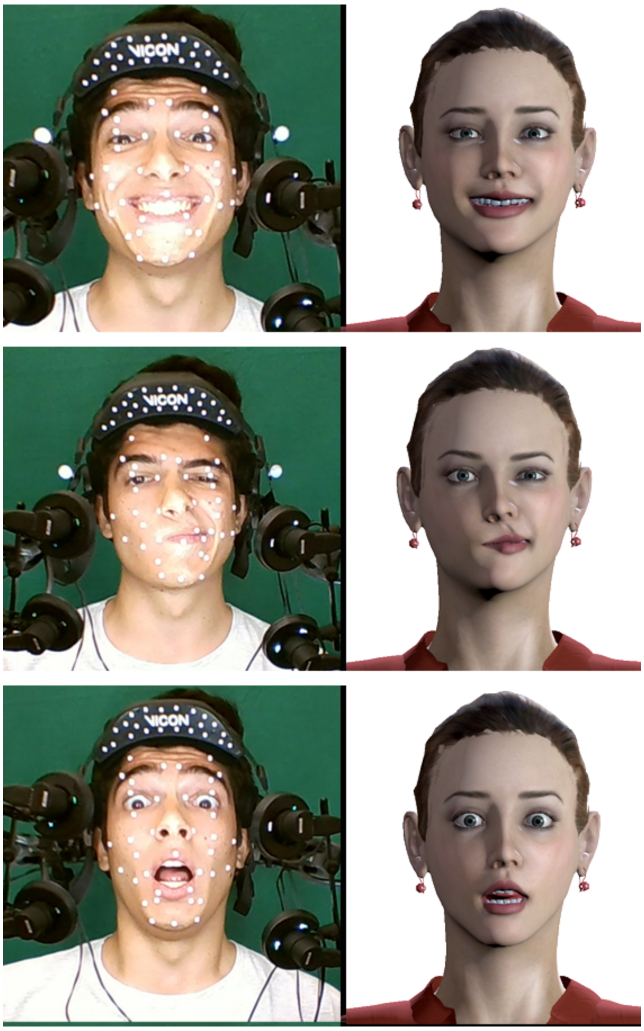

Figure 15: *Capturing and producing of facial expressions*

[DCFN06] Deng Z., Chiang P.-Y., Fox P., Neumann U.: Animating blendshape faces by cross-mapping motion capture data. In *Proceedings of the 2006 Symposium on Interactive 3D Graphics and Games* (New York, NY, USA, 2006), I3D '06, ACM, pp. 43–48. URL: http://doi.acm.org/10.1145/1111411.1111419, doi:10.1145/1111411.1111419. 2

[GLLG07] Gao H., Lim M., Lin E., Green R.: Marker based facial tracking application in communication disorder research. p. 244âĂŞ247. URL: http://hdl.handle.net/10092/754. 1

[GW00] Gonzalez R., Woods R.: *Processamento Digital De Imagens*. Edgard Blucher, 2000. 3

[KW08] Kitagawa M., Windsor B.: *MoCap for Artists: Workflow and Techniques for Motion Capture*. Focal Press, 2008. 2

[LZD13] Le B. H., Zhu M., Deng Z.: Marker optimization for facial motion acquisition and deformation. 1859–1871. URL: doi.ieeecomputersociety.org/10.1109/TVCG.2013.84, doi:10.1109/TVCG.2013.84. 1, 2

[Men10] Menache A.: *Understanding Motion Capture for Computer Animation, Second Edition*, 2nd ed. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2010. 2

[Par72] Parke F. I.: Computer generated animation of faces. In *Proceedings of the ACM Annual Conference - Volume 1* (New York, NY, USA, 1972), ACM '72, ACM, pp. 451–457. URL: http://doi.acm.org/10.1145/800193.569955, doi:10.1145/800193.569955. 1

[PL06] Pighin F., Lewis J. P.: Facial motion retargeting. In *ACM SIGGRAPH 2006 Courses* (New York, NY, USA, 2006), SIGGRAPH '06, ACM. URL: http://doi.acm.org/10.1145/1185657.1185842, doi:10.1145/1185657.1185842. 2

[RGL15] Reverdy C., Gibet S., Larboulette C.: Optimal marker set for motion capture of dynamical facial expressions. In *Proceedings of the 8th ACM SIGGRAPH Conference on Motion in Games* (New York, NY, USA, 2015), MIG '15, ACM, pp. 31–36. URL: http://doi.acm.org/10.1145/2822013.2822042, doi:10.1145/2822013.2822042. 1, 2

[SL14] Seol Y., Lewis J. P.: Tuning facial animation in a mocap pipeline. In *ACM SIGGRAPH 2014 Talks* (New York, NY, USA, 2014), SIGGRAPH '14, ACM, pp. 13:1–13:1. URL: http://doi.acm.org/10.1145/2614106.2614108, doi:10.1145/2614106.2614108. 2

[SPP*01] Sobotta J., Putz R., Pabst R., Putz R., Weiglein A.: *Sobotta Atlas of Human Anatomy*. No. v. 1. Lippincott Williams & Wilkins, 2001. 2

[vic] Vicon cara. https://www.vicon.com/products/camera-systems/cara-1. Accessed: 2018-04-16. 3