Visual Perception Evaluation from The Synthesis of Realistic Facial Expressions Based On Feature Point Clusters

Arif Sulistiyono
Dept. of Animation
Institut Seni Indonesia Yogyakarta
Yogyakarta, Indonesia
arif_sulistiyono@isi.ac.id

Samuel Gandang Gunanto
Dept. of Animation
Institut Seni Indonesia Yogyakarta
Yogyakarta, Indonesia
gandang@isi.ac.id

Agnes Widyasmoro
Dept. of Television and Film
Institut Seni Indonesia Yogyakarta
Yogyakarta, Indonesia
agnesegad@yahoo.com

Lucia Dwi Krisnawati
Dept. of Informatics
Duta Wacana Christian University
Yogyakarta, Indonesia
krisna@staff.ukdw.ac.id

Aditya Wikan Mahastama
Dept. of Informatics
Duta Wacana Christian University
Yogyakarta, Indonesia
mahas@staff.ukdw.ac.id

Abstract— **The increasing demand for animated movies by production houses and television stations needs a significant change in the animation production process. Computer facial animation research on the process of rigging and expression transfer is growing. The traditional approach of facial animation is highly dependent on the animator in making the key and the sequence of facial expression movements. This causes the production of facial animation for one face can not be reused directly for the other face because of its uniqueness. Therefore, the process of automating the formation of weighted areas on the 3D face model with clustering approach and adaptive motion transfer process to face shape is very important to shorten the production process of animation. The principle of animation is seen as one of the solutions and guidelines for the creation of animated facial expression expressively. The synthesis of realistic facial expression can be made on the basis of a feature-point cluster using a radial basis function(RBF). Automation process for formatting the motion area in the face by clustering process based on the location of the feature-point and retargeting process using RBF to perform synthesis of realistic facial expression is the novelty of this research. Based on all experimentation stages, it can be concluded that the synthesis of realistic facial expression based on a feature-point cluster using RBF can be applied to various 3D face models and can be adaptively sensitive to the facial shape of each 3D model which has the same number of marker features. The results of visual perception evaluation from the synthesis of realistic facial expression show that surprise expression has the highest percentage and easily recognizable, 89,32%. Happy expression: 84,63%, sad expression: 77,32%, angry expression: 76,64%, disgust expression: 76,45%, and a fear expression: 76,44%. The average percentage of faces is easily recognizable at 80,13%.**

Keywords— visual perception, expression synthesis, feature point cluster, animation, facial

I. Introduction

The scarcity of existing animation resources is now a major obstacle if the desired acceleration in the animation production process, especially because the process of making movements in animation recently still uses manual techniques by relying on frame-by-frame changes. This will take a lot of time and requires large human resources.

Demand for high and fast animation productivity by production houses and television stations demands significant changes in the animation production process. This is a major problem faced by Indonesian animation studios. Animation production speed is directly proportional to high production costs, so only a few animation products can be produced.

Computer vision technology that plays an important role in the field of animation and games as a motion capture engine of human motion. This technology adopts the human eyes to recognize the phenomenon of camera capture. The captured human movement can be mapped into a skeletal figure model. This also applies to movements on the face by relying on marker features. The location of the marker features is placed in the joints of facial and muscle movements which have a significant movement in expression. This is used so that the capture of motion on the face can be optimal and the mapping on the virtual model can resemble the original. The main principle of human motion capture is the generation of 3D motion animations by real human models through camera capture [1]. The reliability of this system is determined by the accuracy of the estimation of the posed model so that the determination of each segment of the human body in the initial stages is the key to its success. The more accurate the object detection process, the more reliable the system is built [2] [3]. However, due to the high cost of implementation and operations in the production sector [4] [2], this technology is rarely used and is not owned by animation studios in Indonesia which are still classified as small studio criteria.

Currently controlling facial motion animation still depends on the animator or the imitation of the actual actor's actions. Facial Action Coding System or FACS [5] is a useful system to help analyze and simulate the expression as realistic as possible. The unique aspect of the face shape is also a challenge for the animators in forming facial expressions. The area affected by motion on the face of each model is very diverse, this is also a problem for animators for

ISBN 978-1-5386-7421-5/18/USB ©2018 IEEE

the process of automating the transfer of expression between face models. Automation in determining the area of motion must be able to minimize human intervention in the process. Adaptive aspects are assessed by the ability of algorithms to produce motion areas based on the shape of 3D face models, human faces or cartoon characters. This research will evaluate the visual perception of a realistic facial expression synthesis that uses a cluster approach and has a center of several points of rigging features so as to classify members automatically and sensitive to the vertex position. This was considered important because this evaluation stage affected the reliability of the realistic facial expression synthesis that was built on the facial animation automation process. The majority industry still performs the facial animation process manually so that this process is very time consuming and consumes too many human resources because each character's movements are based on their respective expressions and uniqueness. Therefore with this test, the reliability of the realistic facial expression synthesis based on the feature point cluster is expected to be better and can significantly shorten the facial animation production process.

II. State of The Arts

The development of facial animation systems can be seen in two interdependent activities, namely the process of developing control parameters and the user interface, and the development of facial animation implementation techniques based on their parameters [6]. The use of motion capture data as the data source of the research is based on reference to the movement of facial muscles and joints in the bones of human skulls. Motion capture data is seen as having standard similarities and can be easily implemented into 3D face models. This is the reason for Dutreve, Meyer, and Bouakaz to use data from motion capture [7]. Curio et al.'s research confirm the importance of point feature tracking process in each frame, especially when processing point feature data by introducing the Iterative Closest Point (ICP) algorithm process on motion capture data [8].

Orvalho, Zacur, and Susin in 2008 published their research on the transfer of rigging and animation of faces from one character to another face model with limitations on landmark-based human face models and mesh deformations [9]. While Li, Weise & Pauly in 2010 tried to transfer 3D model face rigging with an intermediary sample base. The existence of an intermediary model serves as a regulator of reference changes so that it is reliable. In addition to that in the same year, Dutreve et al. also conducted research that processed rigging removal using the principle of automatic registration and removal of skinning parameters based on facial feature points [10].

Dutreve, Meyer & Bouakaz Research in 2009 introduced the use of radial base functions algorithm to transform coordinate points of 3D features from one 3D face model as a source to another 3D face model as a target. In 2012, the same thing was also done by Umenhoffer et al. By adding the testing of its application to cases of instant processing and cases of processing facial expressions [11].

Research that processed realistic facial expression data from motion capture data sources was initiated by Ju and Lee's research in 2008. This study tried to generate synchronized facial expressions with the actor's speech input using the Markov random fields method approach [12]. Lazzeri et al. in 2015 conducted a preliminary study to examine the face validation of expressive humanoid character models that both reflect positive expressions and negative expressions. In this study, it was concluded that positive expression was easier to recognize compared to negative expression [13].

Kwon and Lee in 2008 also tried to process motion capture data to produce character movements that fulfill the principle of exaggeration using the sub-joint hierarchy. But the motion data used is still limited to body movement data [14]. Utsugi et al. in 2011 tried to make a rendering by utilizing a systematic camera arrangement to produce an interesting action image because of the rendering process of exaggeration. This process is carried out by combining perspectives from multiple cameras combined in perspective [15]. Kwon and Lee in 2012 then continued their research by combining the principle of exaggeration with the principle of squash-and-stretch to produce more expressive movements. This development is able to generate the squash-and-stretch effect for character body movements by combining the general form of exaggeration from posing characters with stretch parameters resulting from the time-wrapping method [16].

The introduction of dynamic changes in the facial skin, especially in the area of wrinkles on the forehead was studied by Dutreve, Meyer, and Bouakaz in 2009 using data on skull skeletal poses and facial wrinkle maps. Deformation of the mesh model is simulated using Large-scale Deformation methods and reference poses that are implemented on 3D human face models [17]. In 2011, Weise, Bouaziz, Li, and Pauly used 2D image data and 3D depth maps captured by stereo Kinect cameras trying to do mesh deformation using blendshape weights instantly on 3D human face models [18].

In 2016 Gunanto began introducing a combination of clustering methods to segment the motion area on the face by modifying it with the position of the rigging feature point as the center of the cluster [19]. The synthesis of human facial expressions is done using a camera with a face source that has been marked with a marker point. The transformation process from 2D coordinate space to 3D coordinate is done by the retargeting method using radial basis function technique abbreviated as RBF [20]. The combination of the two methods, clustering with retargeting, is then referred to as a realistic facial expression synthesis system based on feature-point clustering [21].

III. Experimental Methods

Feature-point clusters are built by applying clustering method with feature-point value (*fp*) related to the position of marker points in the marker-based motion capture system as its center. This substitution facilitates the grouping process because the center or centroid is known so that it can be directly focused on the membership selection process of each cluster by calculating the distance value of each vertex to the registered centroid. After all cluster areas are formed, a realistic facial expression retargeting technique is carried out by transforming the marker point from the source in the form of camera capture to the point of motion feature or rigging in the target 3D model coordinates using a radial basis function to obtain visual expression of results.

Transformation using the radial basis function method is usually implemented into retargeting motion transfer techniques. Retargeting brings the idea of reusing similar character animations so that it can lighten the work of the animator. This technique is presented attractively and uses low-quality 2D-based visual motion data to animate faces with good quality 3D motion capture face calculation data [22]. Radial Basis Functions (RBF) [23] is used to adapt the motion vector of a mesh to another [24]. The RBF method can also be used to transform an animation obtained from 3D face animation or 2D point-based video recording [7], see Fig. 1.

The RBF equation used is as follows:

$$F(x) = \sum_{i=1}^{n} \alpha_i \,.\, \phi(\|x\text{-}x_i\|) \tag{1}$$

The value ϕ depends only on the distance from the center and thus is called radial (2). This radial is the distance between the feature point on the 2D face image.

$$\phi(\|xy_i\text{-}xy_j\|) = \sqrt{(\|xy_i\text{-}xy_j\|)^2 + r^2} \tag{2}$$

Equation (2) is RBF Multiquadric, the variation evaluated in this study is by variation of the application of radial function(2). The value of *xy* is the position of the marker point at 2D coordinates. Using the Pythagoras theorem, the distance between the marker points can be known. The *r* value is determined by the shortest distance of all marker points on the source-animation face (3).

$$r = min_{i \neq j}(\|xy_i\text{-}xy_j\|) \tag{3}$$

The value of $\phi(\|xy_i \text{ -}xy_j\|)$ is used to construct the matrix *H*. Then the weight value (*α*) for each coordinate (*x, y, z*) of the 3D model face is obtained (4):

$$T_x = H.\alpha_x, T_y = H.\alpha_y, T_z = H.\alpha_z \tag{4}$$

then by applying the Gauss elimination is obtained:

$$\alpha_x = H^{-1}.T_x;\ \alpha_y = H^{-1}.T_y;\ \alpha_z = H^{-1}.T_z \tag{5}$$

After the matrix *H*(2) and the weight of each coordinate (*x, y, z*) (5) are obtained, the feature point mappings can be calculated rapidly for each position of the marker point based on the animation face movement of the source using Eq. (6).

$$\begin{aligned} F(x) &= \textstyle\sum_{i=1}^{n} \alpha_i^x \,.\, \phi(\|xy\text{-}xy_i\|) \\ F(y) &= \textstyle\sum_{i=1}^{n} \alpha_i^y \,.\, \phi(\|xy\text{-}xy_i\|) \\ F(z) &= \textstyle\sum_{i=1}^{n} \alpha_i^z \,.\, \phi(\|xy\text{-}xy_i\|) \end{aligned} \tag{6}$$

In the testing phase, consecutive 2D images are extracted to find the location of the marker point on the source face. After obtaining the radial distance value then the weight value (*α*) and the matrix *H* (*ϕ*) of the learning process, the RBF space transformation will estimate the location (*x, y, z*) from the face point of the 3D model. So the face of 3D models still have depth and maintain its 3D form.

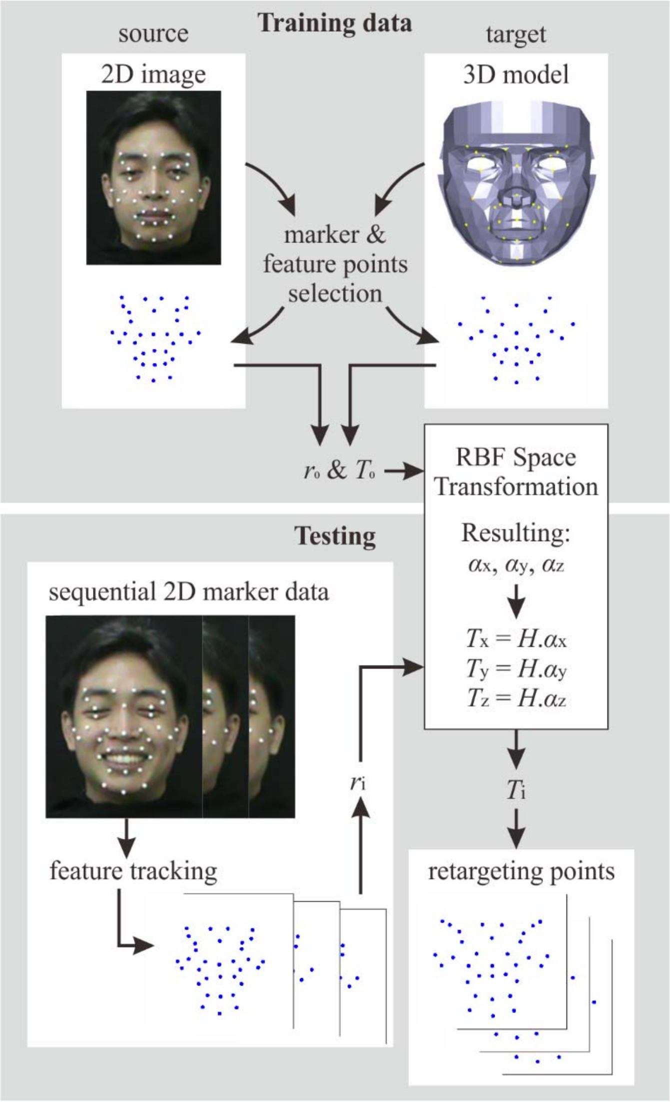

Fig. 1. RBF-based retargeting flow [7] [25]

This research is complete following data retargeting linkages with the results of feature-point clustering calculations to produce synthesis of facial expressions into 3D models starting with the extraction and feature-point registration process for source data (*S*) in the form of video files to perform marker extraction from video files, parameterizing markers as feature-point source values (*So*), and doing marker track by maintaining the coherence of marker parameters in each frame. As for the target data (*T*), it functions to mark the position of the motion feature point or feature-point on the 3D face model (*To*) which will be used in the feature-point clustering and retargeting process using a radial base function transformation that will determine the visualization of realistic facial expression synthesis through the mesh deformation process on the face of the 3D model. The complete flow of this research includes the stages in Fig. 2.

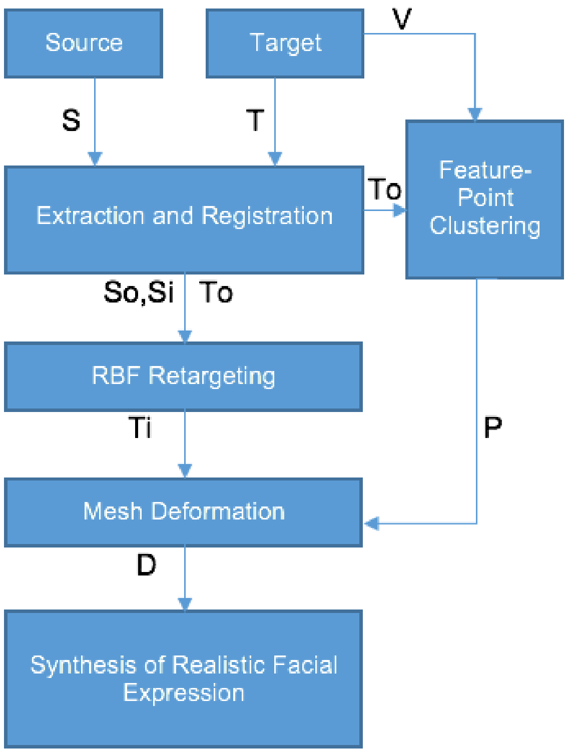

Fig. 2. The flow of research to produce realistic facial expression synthesis with one of the stages in the form of retargeting using a radial basis function.

The term clustering features are introduced to name the clustering process that uses feature-point values that have been determined as the center of the grouping. The clustering process that is done does not need to do an iteration calculation to get the center of the group or cluster that will be formed because the feature-point value chosen is considered as the center or in the discussion about k-means clustering known as the centroid.

So that $X \in R^3, X = \{x_1, x_2, \cdots, x_n\}$ for the point set and $F \in X, F = \{f_1, f_2, \cdots, f_m\}$is any rigging point selected as a feature-point X with $m<n$, the Feature Point Cluster (P) is defined as the set of points on X which the minimum or closest distance is f_m where $P \in X$ with $\{P_1, P_2, \cdots, P_m\}$.

The cluster center point is determined by the F value, the group formation process based on the feature-point values and positions is done by first separating the set X members which are not feature-points as the center point and holding them in the set X_s (7).

$$Xs = \{x \in X \mid \neg(X \cap F)\} \tag{7}$$

The group membership process is determined by first calculating the distance of each X_s member to each feature-point f_m using the euclidean distance, so that each xs_i point will have a distance value for all feature-point f_m(8).

$$D_i = [d_{i,1}, d_{i,2}, d_{i,3}, \ldots, d_{i,m}] \tag{8}$$

Group membership with the feature-point f_m is determined by the minimum or closest distance value (*min* (D_i)) to the center so that all members of the X scattered point data set can be ascertained with a membership value and it's cluster center at the feature-point f_m.

IV. Result and Discussion

This experiment was an evaluation of the 3D facial expression system synthesis framework by describing the results of visual perception questionnaires from the audience to assess the results of the process of transferring emotional expressions from humans to 3D animation characters. The output from the results of this experiment is expected to contribute to an evaluation of the results of the implementation of the 3D realistic facial expression synthesis that has been tested for visual quality.

This evaluation is done after all the terms of the expression formation have fulfilled the FACS theory. For example, the happy expression has an AU criterion: the corner of the lip (22, 24) is drawn wide and rises (Figure 3a). At the shift of the marker point, 2D face image (Fig. 3b) is indicated by the marker point at the edge of the lip (22, 24) which widens and rises. Similarly, the results of the RBF space transformation (Figure 3c) feature a shift in feature points at the lip end (22, 24) that widen and rise.

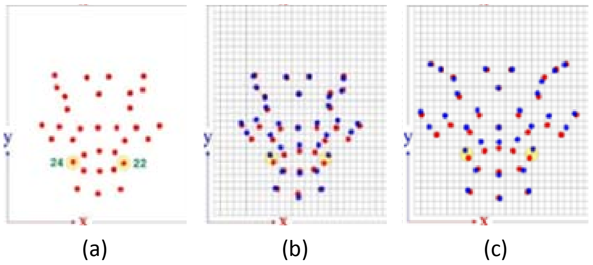

Fig. 3. The position of the FACS marker points that shifted on the happy expression: (a) according to the FACS theory; (b) Extraction results on human faces; (c) RBF retargeting results on 3D model face targets.

Presentation of the results of changes in face shape to each expression can be seen in Fig 4. The six expressions are then displayed in the form of a written questionnaire in the presentation of static and electronic images in the form of a video display. There are 3 3D face models that are targeted for RBF retargeting, namely the human face model, the goose character face model, and the ape character model. Respondents of 33 people had witnessed the results of the implementation while filling out the questionnaire given 10 times to minimize the occurrence of visual perception errors.

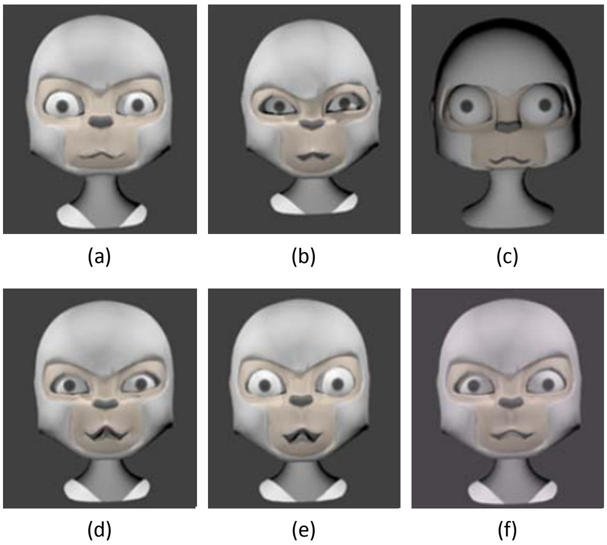

Fig. 4. Visualization of the results of the expression transfer to 3D face models: a) Disgust; b) Angry; c) Fear; d) Happy; e) Surprised; f) Sad.

The results obtained generally show that out of the six expressions, the expression of surprise has the highest percentage easily recognizable, namely: 89.32%. Happy expression: 84.63%, sad expression: 77.32%, angry expression: 76.64%, disgusted expression: 76.45%, and fearful expression: 76.44%. The mean percentage of faces is easily recognizable by 80.13%.

TABLE I. ANSWER TABULATION OF VISUAL PERCEPTION QUESTIONNAIRE ON EXPRESSION SYNTHESIS RESULTS

Expression Synthesis	**3D Head Model**			**Average**
	Human	***Goose***	***Ape***	
Surprise	89,58	89,58	88,79	89,32
Happy	85,42	83,33	85,15	84,63
Disgust	77,08	75,00	77,27	76,45
Angry	79,17	75,00	75,76	76,64
Fear	77,08	77,08	75,15	76,44
Sad	75,00	83,33	73,64	77,32
Average of Accuracy	**80,56**	**80,55**	**79,29**	**80,13**

V. CONCLUSION

Calculation of changes in facial expression of three-dimensional models using linear deformation method can be done on the basis of feature-point clustering segmentation and linear changes using 33 point features as a result of retargeting transformation using the radial basis function. This calculation has resulted in changes in vertex points as a result of changes in centroid position or movement. Deformation that occurs on three-dimensional faces is formed in a linear motion based on changes in feature points in each frame that form every facial animation movement that reflects six basic expressions, namely happy, angry, sad, scared, shocked, and disgusted.

Based on the evaluation results on the six basic expressions used in the testing phase, namely: sad, happy, angry, scared, disgusted and shocked, the retargeting results can remap feature points on the face of the 3D model according to the movement of the marker point on the 2D face image that becomes the face source of animation. Changes in expression that occur are in accordance with the minimum requirements for changes in FACS, so that the expression formed is in accordance with the FACS theory.

The percentage of the results of the audience's visual perception questionnaire which is generally obtained about the implementation of the synthesis of facial expressions into 3D models shows clearly that even though visualization of facial expressions has fulfilled theoretical requirements, it turns out that the implementation is not always able to perfectly describe the desired conditions, namely the percentage of facial expressions easily recognizable amounting to 80.13%. The expression of surprise has the highest percentage easily recognizable, namely: 89.32%. Happy expression: 84.63%, sad expression: 77.32%, angry expression: 76.64%, disgusted expression: 76.45%, and fearful expression: 76.44%. Therefore, the influence of animators in the control of micro expression improvement or the addition of the principle of exaggeration in the making of facial expressions is very important to produce facial expressions that are easily recognized by the audience.

Based on all the stages of experimentation, it can be concluded that the synthesis of realistic facial expressions on the basis of feature-point clusters using radial base functions can be applied to various 3D face models and can be adaptively sensitive to the face shape of each 3D model that has a number of marker features same. This shows that this system can help shorten the process of duplicating similar movements in various forms of 3D face models.

ACKNOWLEDGMENT

This research was sponsored by National Strategics Research Grant 2018 from Ministry of Research, Technology and Higher Education under contract no: 005/SP2H/LT/DRPM/2018.

REFERENCES

[1] J. Aggarwal and Q. Cai, "Human Motion Analysis: A review," in *Computer Vision and Image Understanding, Vol. 73 No.3*, 1999.

[2] F. Perales, "Human Motion Analysis & Synthesis using Computer Vision and Graphics Techniques: State of Art and Applications," in *Workshop on Centre of Computer Graphics and Data Visualisation*, Czech Republic, 2002.

[3] T. Moeslund, "The Analysis-by-Synthesis Approach in Human Motion Capture: A Review," in *The 8th Danish conference on pattern recognition and image analysis*, Denmark, 1999.

[4] T. F. Shipley and J. S. Brumberg, "Markerless Motion-capture for Point-light Displays," Biological Motion Project, Department of Psychology, Temple University, Philadelphia, 2005.

[5] P. Ekman and W. V. Friesen, Facial Action Coding System: a technique for the measurement of facial movement, Palo Alto: Consulting Psychologists Press, 1978.

[6] F. Parke and K. Waters, Computer Facial Animation 2nd Edition, Massachusetts: AK Peters, 2008.

[7] L. Dutreve, A. Meyer and S. Bouakaz, "Feature points based facial animation retargeting," in *Proceedings of the 15th ACM symposium on virtual reality software and technology*, 2008.

[8] C. Curio, M. Breidt, M. Kleiner, Q. C. Vuong, M. A. Giese and H. H. Bulthoff, "Semantic 3D Motion Retargeting for Facial Animation," in *Proc. of the 3rd Symposium on Applied Perception in Graphics and Visualization*, 2006.

[9] V. C. Orvalho, E. Zacur and A. Susin, "Transferring the Rig and Animations from a Character to Different Face Models," in *COMPUTER GRAPHICS FORUM Volume 27 Number 8*, 2008.

[10] L. Dutreve, A. Meyer, V. Orvalho and S. Bouakaz, "Easy Rigging of Face by Automatic Registration and Transfer of Skinning Parameters," in *Proceedings of the 2010 International Conference on Computer Vision and Graphics*, 2010.

[11] T. Umenhoffer and B. Toth, "Facial Animation Retargeting Framework Using Radial Basis Functions," in *6th Hungarian Conference on Computer Graphics and Geometry*, Budapest, 2012.

[12] E. Ju and J. Lee, "Expressive Facial Gestures From Motion Capture Data," in *Eurographics*, 2008.

[13] N. Lazzeri, D. Mazzei, A. Greco, A. Lanata, D. D. Rossi and A. Rotesi, "Expressive Humanoid Face: a Preliminary Validation Study," in *The 8th International Conference on Advances in Computer-Human Interactions*, 2015.

[14] J.-Y. Kwon and I.-K. Lee, "Exaggerating Character Motions Using Sub-joint Hierarchy," *Computer Graphics Forum Vol 27,* pp. 1677-1686, 2008.

[15] K. Utsugi, T. Naemura, M. Oikawa and T. Koike, "E-IMPACT: Exaggerated Illustrations using Multi-perspective Animation Control Tree Structure," in *Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology*, 2011.

[16] J.-Y. Kwon and I.-K. Lee, "The Squash-and-Stretch Stylization for Character Motions," *IEEE Transactions on Visualization and Computer Graphics Vol 18,* pp. 488-500, 2012.

[17] L. Dutreve, A. Meyer and S. Bouakaz, "Real-time Dynamic Wrinkles of Face for Animated Skinned Mesh," in *Proceedings of 5th International Symposium Visual Computing (ISVC)*, 2009.

[18] T. Weise, S. Bouaziz, H. Li and M. Pauly, "Realtime Performance-Based Facial Animation," *Journal ACM Transactions on Graphics,* 2011.

[19] S. G. Gunanto, M. Hariadi and E. M. Yuniarno, "Generating Weight Paint Area on 3D Cartoon-Face Models," *INFORMATION - An International Interdisciplinary Journal,* vol. 19, no. 9B, pp. 4183-4190, September 2016.

[20] S. G. Gunanto, M. Hariadi and E. M. Yuniarno, "Computer Facial Animation with Synthesize Marker on 3D Faces Surface," in *2nd International Conference of Industrial, Mechanical, Electrical, Chemical Engineering (ICIMECE)*, Yogyakarta, 2016.

[21] S. G. Gunanto, M. Hariadi and E. M. Yuniarno, "Facial Animation of Life-Like Avatar based on Feature Point Cluster," *Journal of Engineering Science and Technology Review,* vol. 10, no. 1, pp. 168-172, 2017.

[22] J. Chai, J. Xiao and J. Hodgins, "Vision-based control of 3D facial animation," in *Proceedings of the 2003 ACM SIGGRAPH/Eurographics symposium on computer animation*, 2003.

[23] M. Powell, "Radial basis functions for multivariable interpolation: a review," in *Algorithms for approximation*, 1987, pp. 143-167.

[24] J. Noh and U. Neumann, "Expression cloning," in *Proceedings of the 28th annual conference on computer graphics and interactive techniques*, 2001.

[25] Troy, Pranowo and S. G. Gunanto, "2D to 3D Space Transformation for Facial Animation Based on Marker Data," in *The 6th International Annual Engineering Seminar (InAES)*, Yogyakarta, 2016.