Review on facial expression modeling

Noor Adibah Najihah Mat Noor[1], Norhaida Mohd Suaib[2]
[1]School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia, Skudai, Johor, Malaysia
[2]UTM-BDC, Ibnu Sina Institute for Scientific and Industrial Research, Universiti Teknologi Malaysia, Skudai, Johor, Malaysia

Article Info

Article history:

Received Jan 13, 2020
Revised Nov 12, 2021
Accepted Jan 31, 2022

Keywords:

3D model
Computer graphics
Face model
Facial expression
Facial reconstruction

ABSTRACT

Facial modeling has been an ongoing research for many years and still shows research trend due to its relevance to current technology. Many applications incorporate facial expression modeling with the help of facial tracking, facial animation and facial recognition. The existing performance of the modeling method faces the challenges to perform well due to many factors. Currently, the use of 2D images and videos as inputs for modeling process are gaining popularity. However, current technologies and development had extended the trends towards acquiring 3D human data. This paper provides an overview on variety of modeling techniques based on human facial model that can lead to future research.

Corresponding Author:

Noor Adibah Najihah Mat Noor
School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia
Sultan Ibrahim Chancellery Building, Jalan Iman, 81310 Skudai, Johor, Malaysia
Email: nooradibahnajihah@gmail.com

1. INTRODUCTION

Facial expression is an important aspect in non-verbal communication. Facial expression research not only addresses the areas of psychology and humanities, but it is also studied by computer science researchers. Machine learning, for example, has been used to recognize facial features and facial expressions, while in computer graphics facial expressions modeling add realism to animation, computer games and CGI films. Facial modeling is studied extensively in computer graphics since it provides the base of constructing different facial expressions and acts as a mechanism to analyze and synthesize facial expressions. It is commonly used for face recognition, human emotional analysis, drowsiness detection (such as in [1]), face tracking and 3D facial animation.

Facial expression modeling provides the base of different facial expressions and construction mechanism to analyze and synthesize the facial expression. The process of modeling the facial modeling includes the face existence detection and implementation of face extraction mechanism such as face identification and expression from sequences of captured images. This modeling approach constructed standard form for different facial expression. Some of highlighted issues in facial modeling from 2D facial images are as follows: difficulties due to head pose, scale variation, pose scale, occlusion and incomplete geometric features. 3D scene that contains facial landmark can be used as reference for facial data in generating facial expression. However, the construction of 3D face model involves high level of difficulties in fundamental computer vision problem.

Facial features and face parameters play important roles in the construction of 3D face model. Both facial features and facial parameter are needed to give detail appearance on the face model such as the position of the eyes, mouth and nose. The parameters also help in reflecting the emotional state and also describing the facial expression on the face model. Generally, facial expression modeling involves the use of energy function. The energy function takes both poses of head and scales with consideration of landmark fitting to complement the traditional 2D facial expression reconstruction. Different energy optimization methods can be used to compute the counterpart of 3D facial expression from 2D facial images such as least square and gradient descent. The increasing level of energy function complexity due to degree of freedom such as identity, expression, head pose, and scale in fitting may lead to higher computation cost and probability of reaching local optimum that eventually limit the modeling result. Other than facial feature points, the images also require depth analysis since 3D model reconstruction requires depth to obtain the three dimensional view of the model.

This paper presents a review on facial modeling. The following sections will discuss on related previous work involving few types of approaches including 2D-based modeling, 3D-based modeling (geometric, static and dynamic) and also deep learning which is the base of facial expression construction. In the next section, discussion on facial modeling approaches will be presented and finally, the conclusion will be given in last section.

2. PREVIOUS WORK

Facial modeling had been an on going research for many years. There were various works that have been presented by previous researchers. These modeling approaches are related to 2D-based modeling, 3D geometric based modeling, 3D static based modeling, 3D dynamic based modeling and modeling through deep learning which lead to better result of facial expression modeling.

2.1. 2D-based modeling

2D-based face model construction are based on 2D image features such as face contour, facial feature landmark and texture. Active shape model [2] and active appearance model [3] are classical model that were used to represent the facial feature by facial landmark while [4] improved the algorithm to reach out better analysis of facial expression. Other than that, facial feature can also be presented by facial contour [5]. The facial image analysis and facial expression clustering can be done by using face space that was constructed based on principal component analysis [6], [7]. The modeling process based on face space was able to help face detection [8]. 2D facial images usually include information of 2D data that can help in 3D estimation of shape, features layer (including landmark points) and occluding boundaries or texture edges called contour. The 3D morphable model (3DMM) fitting algorithm to 2D geometric information was also presented [9].

2.2. 3D Geometric based modeling

Other than that, the construction of facial model also can be based on geometric features extraction from multi-facial images or 3D facial data. Facial model can be built using anthropometric facial features [10] and multi views of facial data [11]. Face modeling also can be done based on surface of B-Spline reconstruction [12] or sparse-iterative closest point [13]. Hybrid facial geometry algorithm (HFGA) can also be used without initialized model, as it can automatically extracted facial features and classifying the facial expression by the geometric values of facial features [14].

2.3. 3D Static based modeling

The static facial feature space construction is one of the available methods for face modeling. A classical facial modeling method called 3D morphable model (3DMM) was used to reconstruct 3D facial data with special characteristic [15]. The 3DMM used fitting method and the method considers the edge information to improve the fitting speed [16]. The static modeling constructed the face space using facial feature analysis [17]. The different faces were transferred to regular representation and face space. The performance was limited since the method was depending to facial samples. For modeling method, more computations are required to ensure better result and performance.

2.4. 3D Dynamic modeling

A variety of facial samples with different expressions can also be used to construct facial model. There are past works that implemented the method. Deformation modeling was suggested to be applied for 3D face matching [18]. Reconstruction of high fidelity 3D facial model using facial data which includes the texture, wrinkles, facial frontal and side images [19], [20]. Others suggested the usage of core tensor space to reduce computation of facial reconstruction, tensor subspace analysis for face modeling and face recognition [21], and local tensor for face recognition [22]. Multilinear models are used for facial expression modeling by [23] and also being constructed by Facewarehouse with facial surface fitting that was implemented for face tracking and facial animation [24]. Reconstruction process usually needs the energy equation optimization but it will affect and limit the quality of facial modeling by local optimization.

2.5. Deep learning

One of the approaches in facial modeling is by using variation of facial features in deep learning and manifold learning. Multiple processing layers are applied in deep learning to learn data representation with different level of feature extraction [25] which eventually improves the performance such as The Deepface [26]; VGGFace [27] and FaceNet [28]. Some proposed methods for facial expression analysis such as restricted boltzmann machines [29] and deep fusion convolutional neural network for multi model are based on manifold framework [30]. Other than that, convolutional neural networks was also proposed for extended based method of facial expression to tackle data accuracy issues in machine learning [31]. Multi facial images deep neural network was suggested for facial expression recognition [32] as well as the 3D facial data flow that was mapped into grassman manifold [33] where combination of manifold learning and static facial modeling were introduced [34]. Shape space was also one of the features that can be used in this approach. Shape space can be used for facial data analysis [35] and elastic measure in shape space enables 3D faces differences to be analyzed [36]. Kendall shape space was proposed in reconstruction of 3D facial model in facial modeling for higher accuracy rate [37]. Taxonomy of state-of-the-art-image-base 3D object [38] through potraying the deep learning and manifold learning approach may lead to better result of facial expression recognition with complicated framework for facial expression modeling.

3. DISCUSSION

Each modeling method takes on a variety of challenges in producing the best performance. 3D reconstruction for 2D images is one of the fundamental issues of modeling and computer vision. By these previews of methods, the most preferable choice is reconstruction of the scene automatically by little or no user intervention. 3D scene reconstruction from 2D image is very challenging as 2D image contains less information from loss of depth information due to the nature of image forming process [39]. In order to obtain points of the scene, there are some image requirements to reconstruct 3D model: points projection of the scene on the image, matching ability to recognize and associate points, 3D coordinate determination based on association and calibration parameters, and obtaining mesh by finding surface own by the point to build model.

There were various previous works that used 2D face modeling and 3D face modeling which include geometry modeling, static modeling and dynamic modeling. However, deep learning had become the current trend in solving problem on computer vision issues as it demonstrates good performance in identification, classification and target detection. Apart from that, it shows the ablity to lower the reliability on image preprocessing, feature extraction and shows better result compared to conventional approaches.

Clearly, it is difficult to perform facial expression model due to head pose changes and facial changes based on expression. The occlusion and illumination that happened may cause information loss. Extracting the features as facial expression representation can be done by traditional methods but these approaches may not be able to handle it well due to the large pose variation. Pose normalization for mapping using facial features can be completed by exploiting the correlation of different poses by regression model in small face poses. Large face pose drop their performance due to accuracy of facial landmark detection. 3D geometrical transformation suggested by aligning 2D facial image with 3D model [40] tackled the normalization pose but is inconvenient for facial expression modeling. Most methods learn multiple classifiers for each pose which eventually will be time-consuming. Recovering 3D coordinate of an image using features matching across images based on a variety of view and 3D construction by segmented 2D silhouettes of the same object with multiple images that captured by well-calibrated cameras. Quality of 3D model made by these models is sensible but not feasible in many situations.

Comparison made by [41] which is made up of illumination changes, head motion, aging, facial make-up, acquisition of data, availability of data, computational cost, measurement of face surface, action units (AU)s low-intensity performance, acquisition and recognition, neutral scanning, AUs and facial sample database availability and real-time basically shows the pro and cons of using 2D and 3D. However, both 2D and 3D are having also similar problems such as lack of spontaneous dataset, manually marked features and facial points, and equal accuracy needed for all expression.

Recently, several researchers revealed the deep learning in computer vision and visual task inspired by deep network where they managed to get good result in identifying, classifying, and detecting. Deep convolutional neural network (CNN) leads in processing images, videos, speech and tools while current NN is leaning more towards the sequential data like text and speech where both success rate are depending on input data [42]. 3D database in CNN architecture for training data will affect the rate of performance. Generative adversarial network (GAN) is also a popular model due to its ability through the adversarial training with the ability to produce sharp images and is widely used for face-related tasks. This encourages the usage of GAN into the training set. Other than that, accuracy in deep learning can be boosted up using comprehensive parameter turning as the result of facial expression based on deep learning that utilizes the transfer learning of available models with results of more than 70% accurate rate [43].

4. CONCLUSION

This paper focuses on methods of facial expression modeling. Reconstructing 3D facial expression based on 2D face modeling also deals with similar problems as in the area of static and dynamic modeling, geometric modeling and deep learning. Based on preview of previous work, deep learning shows improvement of result based on accuracy compared to other method of facial modeling. 3D facial expression modeling accuracy rate is influenced by the shape representation; either single image, multiple images or sequences of images. Architecture of network implies the output representation and eventually affected the computational efficiency and reconstruction quality. Training mechanism applied include variation of datasets, loss functions that can affect the output quality and training procedure.

Based on current review, deep learning method are showing the high percentage of accuracy result for facial model however the process is mainly time-consuming. Facial modeling using deep learning method with the enhancement of time consumption due to work load process would be a recommendation for future research area.

ACKNOWLEDGEMENTS

We would like to express our gratitude to Universiti Teknologi Malaysia (UTM) - research grant UTMER 19J10 in particular, and Ministry of Higher Education (MOHE), of which direct and indirect supports make this research a success.

REFERENCES

[1] B. Vijayalaxmi, K. Sekaran, N. Neelima, P. Chandana, N. M. Meqdad and S. Kadry, "Implementation of face and eye detection on DM6437 board using simulink model," *Bulletin of Electrical Engineering and Informatics*, vol. 9, no. 2, pp. 785-791, 2020, doi: 10.11591/eei.v9i2.1703.

[2] T. F Cootes, C. J. Taylor, D. H. Cooper and J. Graham, "Active shape models-their training and application," *Computer vision and image understanding*, vol. 61, no. 1, pp. 38-59, 1995, doi: 10.1006/cviu.1995.1004.

[3] G. J. Edwards, T. F Cootes and C. J Taylor, "Face recognition using active appearance models," *In European conference on computer vision. Springer*, Berlin, Heidelberg, pp. 581-595, 1998.

[4] M Iqtait. F. S Mohamad and M. Mamat, "Feature extraction for face recognition via Active Shape Model (ASM) and Active Appearance Model (AAM)," *IOP Conference Series: Materials Science and Engineering. IOP Publishing*, vol. 332, no. 1, pp. 1-9, 2018, doi: 10.1088/1757-899X/332/1/012032.

[5] C. L Huang and Y. M Huang, "Facial expression recognition using model-based feature extraction and action parameters classification," *Journal of visual communication and image representation*, vol. 8, no. 3, pp. 278-290, 1997, doi: 10.1006/jvci.1997.0359.

[6] P. S. Penev and L. Sirovich, "The global dimensionality of face space," *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580)*, 2000, pp. 264-270, doi: 10.1109/AFGR.2000.840645.

[7] X. W. Chen and T. Huang, "Facial expression recognition: a clustering-based approach," *Pattern Recognition Letters*, vol. 24, no. 9-10, pp. 1295-1302, 2003, doi: 10.1016/S0167-8655(02)00371-9.

[8] P. P. Paul and M. Gavrilova, "PCA Based Geometric Modeling for Automatic Face Detection," *International Conference on Computational Science and Its Applications*, 2011, pp. 33-38, doi: 10.1109/ICCSA.2011.69.

[9] A. Bas and W. A. Smith, "What does 2D geometric information really tell us about 3D face shape?," *International Journal of Computer Vision*, vol. 127, pp. 1455-1473, 2019, doi: 10.1007/s11263-019-01197-x.

[10] D. Decarlo, D. Metaxas and M. Stone, "An anthropometric face model using variational techniques," *SIGGRAPH '98: Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, pp. 67-74, 1998, doi: 10.1145/280814.280823.

[11] K. S. Lee, K. H. Wong, S. H. Or and Y. F. Fung, "3D face modeling from perspective-views and contour-based generic-model," *Real-Time Imaging*, vol. 7, no. 2, pp. 173-182, 2001, doi: 10.1006/rtim.2000.0241.

[12] W. Peng, C. Xu and Z. Feng, "3D face modeling based on structure optimization and surface reconstruction with B-Spline," *Neurocomputing*, vol. 179, pp. 228-237, 2016, doi: 10.1016/j.neucom.2015.11.090.

[13] S. Zhan. L. Chang, J. Zhou, T. Kurihara, H. Du, Y. Taang and J. Cheng, "Real-time 3d face modeling based on 3d face imaging," *Neurocomputing*, vol. 252, pp. 42-48, 2017, doi: 10.1016/j.neucom.2016.10.091.

[14] S. P. Khandait, R. C. Thool and P. D. Khandait, "ANFIS and BPNN based Expression Recognition using HFGA for Feature Extraction," *Bulletin of Electrical Engineering and Informatics*, vol. 2, no. 1, pp. 11-22, 2013.

[15] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3D faces," *SIGGRAPH '99: Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, pp. 187-194, 1999, doi: 10.1145/311535.311556.

[16] A. Bas, W. A. P. Smith, T. Bolkart and S. Wuhrer, "Fitting a 3D Morphable Model to Edges: A Comparison Between Hard and Soft Correspondences," *Asian Conference on Computer Vision*, pp. 377-391, 2017.

[17] F. Duan, D. Huang, Y. Tian, K. Lu, Z. Wu and M. Zhou, "3D face reconstruction from skull by regression modeling in shape parameter spaces," *Neurocomputing*, vol. 151, pp. 674-682, 2015, doi: 10.1016/j.neucom.2014.04.089.

[18] X. Lu and A. Jain, "Deformation Modeling for Robust 3D Face Matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 8, pp. 1346-1357, 2008, doi: 10.1109/TPAMI.2007.70784.
[19] A. F. Ichim, S. Buoaziz and M. Pauly, "Dynamic 3D avatar creation from hand-held video input," *ACM Transactions on Graphics (ToG)*, vol. 4, pp. 1-14, 2015, doi: 10.1145/2766974.
[20] H. Jin, X. Wang, Z. Zhong and J. Hua, "Robust 3D face modeling and reconstruction from frontal and side images," *Computer Aided Geometric Design*, vol. 50, pp. 1-13, 2017, doi: 10.1016/j.cagd.2016.11.001.
[21] X. Gao and C. Tian, "Multi-view face recognition based on tensor subspace analysis and view manifold modeling," *Neurocomputing*, vol. 72, no. 16-18, pp. 3742-3750, 2009, doi: 10.1016/j.neucom.2009.06.001.
[22] X. Song, Z. H. Feng, X. Yang, X. Wu and J. Yang, "Towards multi-scale fuzzy sparse discriminant analysis using local third-order tensor model of face images," *Neurocomputing*, vol. 185, pp. 53-63, 2016, doi: 10.1016/j.neucom.2015.12.019.
[23] D. Vlasic, M. Brand, H. Pfister and J. Popovic, "Face transfer with multilinear models," *ACM transactions on graphics (TOG) ACM*, vol 24, no. 3, pp. 426-433, 20056, doi: 10.1145/1185657.1185864.
[24] C. Cao, Y. Weng, S. Zhou, Y. Tong and K. Zhou, "FaceWarehouse: A 3D Facial Expression Database for Visual Computing," *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 3, pp. 413-425, 2014, doi: 10.1109/TVCG.2013.249.
[25] M. Wang and W. Deng, "Deep face recognition: A survey," *Neurocomputing*, vol. 429, pp. 215-244, 2021, doi: 10.1016/j.neucom.2020.10.081.
[26] Y. Taigman, M. Yang, M. A. Ranzato and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 1701-1708.
[27] O. M. Parkhi, A. Vedaldi and A. Zisserman, "Deep face recognition," *Distributed unchanged freely in print or electronic forms*, 2015.
[28] F. Schroff, D. Klaenichenko and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 815-823.
[29] S. Elaiwat, M. Bennamoun and F. Boussaid, "A spatio-temporal RBM-based model for facial expression recognition," *Pattern Recognition*, vol. 49, pp. 152-161, 2016, doi: 10.1016/j.patcog.2015.07.006.
[30] H. Li, J. Sun, Z. Xu and L. Chen, "Multimodal 2D+3D Facial Expression Recognition With Deep Fusion Convolutional Neural Network," *IEEE Transactions on Multimedia*, vol. 19, no. 12, pp. 2816-2831, 2017, doi: 10.1109/TMM.2017.2713408.
[31] A. T. Lopes, E. de Agujar, A. F. DeSouza and T.s Oliveira-Santos, "Facial expression recognition with convolutional neural networks: coping with few data and the training sample order," *Pattern Recognition*, vol. 61, pp. 610-628, 2017, doi: 10.1016/j.patcog.2016.07.026.
[32] T. Zhang, W. Zheng, Z. Cui, Y. Zong, J. Yan and K. Yan, "A Deep Neural Network-Driven Feature Learning Method for Multi-view Facial Expression Recognition," *IEEE Transactions on Multimedia*, vol. 18, no. 12, pp. 2528-2536, 2016, doi: 10.1109/TMM.2016.2598092.
[33] T. Alashkar, B. B. Amor, M. Daoudi and S. Berretti, "A Grassmann framework for 4D facial shape analysis," *Pattern Recognition*, vol. 57, pp. 21-30, 2016, doi: 10.1016/j.patcog.2016.03.013.
[34] A. Patel and W. A. Smith, "Manifold-based constraints for operations in face space," *Pattern recognition*, vol. 52, pp. 206-217, 2016, doi: 10.1016/j.patcog.2015.10.003.
[35] A. Brunton, A. Salazar, T. Bolkart and S. Wuhrer, "Review of statistical shape spaces for 3D data with comparative analysis for human faces," *Computer Vision and Image Understanding*, vol. 128, pp. 1-17, 2014, doi: 10.1016/j.cviu.2014.05.005.
[36] S. Kurtek and H. Drira, "A comprehensive statistical framework for elastic shape analysis of 3D faces," *Computers & Graphics*, vol. 51, pp. 52-59, 2015, doi: 10.1016/j.cag.2015.05.027.
[37] C. Lv, Z. Wu, X. Wang, and M. Zhou, "3D facial expression modeling based on facial landmarks in single image," *Neurocomputing*, vol. 355, pp. 155-167, 2019, doi: 10.1016/j.neucom.2019.04.050.
[38] X. -F. Han, H. Laga and M. Bennamoun, "Image-Based 3D Object Reconstruction: State-of-the-Art and Trends in the Deep Learning Era," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 5, pp. 1578-1604, 2021, doi: 10.1109/TPAMI.2019.2954885.
[39] M. AharchI and M. A. Kbir, "Review on 3D Reconstruction Techniques from 2D Images," *The Proceedings of the Third International Conference on Smart City Applications. Springer*, Cham, 2019, pp. 510-552.
[40] F. Zhang, T. Zhang, Q. Mao and C. Xu, "Geometry Guided Pose-Invariant Facial Expression Recognition," *IEEE Transactions on Image Processing*, vol. 29, pp. 4445-4460, 2020, doi: 10.1109/TIP.2020.2972114.
[41] F. Nonis, N. Dagnes, F. Marcolin and E. Vezzetti, "3D approaches and challenges in facial expression recognition algorithms—A Literature Review," *Applied Sciences*, vol. 9, no. 18, pp. 1-33, 2019, doi: 10.3390/app9183904.
[42] Y. LeCun, Y. Bengjo and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 463-444, 2015.
[43] H. Kusuma, M. Attamimi and H. Fahrudin, "Deep learning based facial expressions recognition system for assisting visually impaired persons," *Bulletin of Electrical Engineering and Informatics*, vol. 9, no. 3, pp. 1208-1219, 2020, doi: 10.11591/eei.v9i3.2030.

BIOGRAPHIES OF AUTHORS

Noor Adibah Najihah Mat Noor from Universiti Teknologi Malaysia (UTM) is currently pursuing Ph.D in Computer Science, previously graduated with master degree in 2017 in computer science and also received bachelor degree in computer science (graphic and multimedia software) in 2015. Her interest is mainly related to computer graphics research – her research during previous study in master degree, was related to non-photorealistic rendering techniques. Current research interest for PhD study is on facial recognition and also including facial modeling and facial animation. She can be contacted at email: nooradibahnajihah@gmail.com.

Norhaida Mohd Suaib specializes in the field of computer graphics, particularly interactive computer graphics, collision detection, physics-based model and virtual reality. As a lecturer at School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia (UTM), she taught many computer graphics-related subjects in the undergraduate and postgraduate levels. She is an active member in the UTM ViCubeLab Research Group - a research group dedicated on virtual (virtual reality/environment), Visualization and Vision. Among her current research focus is on using computer graphics and computer games technology in the preservation of both tangible and intangible cultural heritage; where face modeling, recognition and reconstruction play important part to increase the realism of computer-generated scenes. Currently she is attached to UTM Big Data Center (UTM BDC), Ibnu Sina Institute for Scientific and Industrial Research (ISI-SIR), Universiti Teknologi Malaysia as a research fellow. She can be contacted at email: haida@utm.my.