FACIAL EXTRACTION AND LIP TRACKING USING FACIAL POINTS

Hlaing Htake Khaung Tin

University of Computer Studies, Yangon, Myanmar
hlainghtakekhaungtin@gmail.com

ABSTRACT

Automatic facial feature extraction is one of the most important and attempted problems in computer vision. It is a necessary step in face recognition, facial image compression. There are many methods have been proposed in the literature for the facial feature extraction task. However, all of them have still disadvantage such as not complete reflection about face structure, face texture. In this paper, we propose a method for fast and accurate extraction of feature points such as eyes, nose, mouth, eyebrows and the like from dynamic images with the purpose of face recognition. These methods are far from satisfactory in terms of extraction accuracy and processing speed. The proposed method achieves high position accuracy at a low computing cost by combining shape extraction with geometric features of facial images like eyes, nose, mouth etc. In this paper, a facial expressions synthesis system, based on the facial points tracking in the frontal image sequences. Selected facial points are automatically tracked using a cross-correlation based optical flow. The proposed synthesis system uses a simple facial features model with a few set of control points that can be tracked in original facial image sequences.

KEYWORDS

Facial feature point extraction; face recognition; geometric features.

1. INTRODUCTION

One of the most elusive goals in computer animation is the realistic animation of the human face. Human face modelling and animation are very complex tasks because of the physical structure of the face and the dynamics involving its psychological and behavioural aspects [11]. The objective of this paper is to study a statistical model for human face aging, which is then used for face aging simulation and age estimation. Face aging simulation and prediction is an interesting task with many applications in digital entertainment [1]. A problem of personal verification and identification is an actively growing area of research. Face, voice, lip movements, hand geometry, odor, gait, iris, retina, fingerprint are the most commonly used authentication methods. All of these psychological and behavioural characteristics of a person are called biometric. The driving force of the progress in this field is due to the growing role of the Internet and electronic transfers in modern society. Therefore, considerable number of applications is concentrated in the area of electronic commerce and electronic banking systems. The biometrics have a significant advantage over traditional authentication techniques due to the biometric characteristics of the individual are not easily transferable, are unique of every person and cannot be lost, stolen or broken. The biometrics is a measurable physiological or behavioural characteristic of an individual used in personal identification and verification and the choice of the biometric solutions depends on user acceptance, level of security, accuracy, and cost and implementation time [2].

2. RELATED WORK

In Facial feature extraction, local features on face such as nose, and then eyes are extracted and then used as input data. And it has been the central step for several applications. Various approaches have been proposed in this chapter to extract these facial points from images or video sequences of faces. The basically of approaches are geometry-based, template-based, colour segmentation techniques, appearance-based approaches [5]. The ability for a computer system to sense the user's emotions opens a wide range of applications in different research areas, including security, law enforcement, medicine, education, and telecommunications [6]. However , it is important not to confuse human emotion recognition from facial expression recognition: the latter is merely a classification of facial deformations into a set of abstract classes, solely based on visual information. Indeed, human emotions can only be inferred from context, self-report, physiological indicators, and expressive behaviour which may or may not include facial expressions [7]. Zhilin Wu et al. [8] have presented a combined method to accurately track the outer lips by using the Gradient Vector Flow (GVF) snakes with parabolic templates as an additional external force. This combination needs fewer requirements of both salient boundaries and accuracy of templates. Erol [9] described a facial modelling and animation system that used muscle-based generic face model and deformed it using deformation techniques to model the individualized faces. Two orthogonal photos of the real faces were used for this purpose. Image processing techniques were employed to extract certain features on the photographs, which were then refined manually by the user through the facilities of the user interface of the system. The feature points located on the frontal and side views of a real face were used to deform the generic model. Then, the muscle vectors in the individualized face model were arranged accordingly. The individualized models produced in this manner were animated using the parametric interpolation techniques. Worral et al. [10] examined the performance of the low bit rate 3-D , "Talking Heads", in order to determine the appropriate delivery scheme for the encoded data.

2. IMAGE DATABASE

In this work, we used a database that prepared at University of Computer Studies, Yangon for this research. There are 500 subjects in the database. Subjects sat directly in front of the camera, for each person there are on average 5 frames for frontal view. Image sequences for the frontal views were digitized into 640*490 pixel array with 8 bits greyscale. Table 1 shows database specifications.

Age	15 to 70
Female	70%
Male	30%
Myanmar	80%
Other	20%
Resolution	640 * 490
	Greyscale

Table 1. Database Specifications

3. Facial Feature Point Tracking Using Optical Flow

In the first digitized frame, 21 key feature points were manually marked with a computer-mouse around facial features such as eyes, eyebrows and mouth in Figure 1. That is why the information expressing movement of eyes, eyebrows and mouth is desirable for machine recognition of the facial expressions. We are confined in these which are representative of the boundary between these components and then determine the facial feature points which are representative of the boundary between these components and skin. Some of this points used only for specifying the model features in the first frame. Other points were automatically tracked in the subsequent frames using cross-correlation based optical flow.

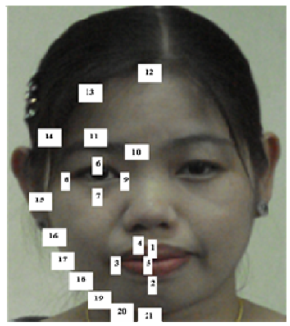


Figure 1. Selected 21 facial feature points

Each point is the centre of a 11*11 flow window that includes horizontal and vertical flows. Figure.2 shows the implementation of this method in two subsequent frames.

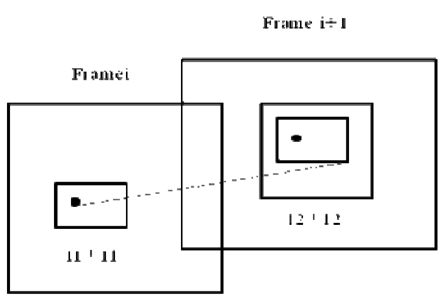


Figure 2. Cross- correlation optical flow calculation

Cross-correlation of 11*11 windows in the first frame, and a 21*21 window at the next frame were calculated and the position with maximum cross-correlation of two windows, were estimated as the position of the feature point at the next frame. In this paper, we developed a face model for synthesizing the facial expressions, by tracking some feature points that were described in the Figure.1. In the subsequent section we describe the specifications of the proposed face model.

4. Face Model

We are confined in the three facial features: mouth, eyes and eyebrows. We used a muscle-based triangular patch object model that the vertices of the triangles were determined from the feature points tracking results [3]. Deformed triangles gave the ability of shape tracking to the model. In this section we describe the mouth, eyes and eyebrows model.

4.1. Mouse Model

Figure 3 shows the selected mouth feature points and face features that were extracted from these points.

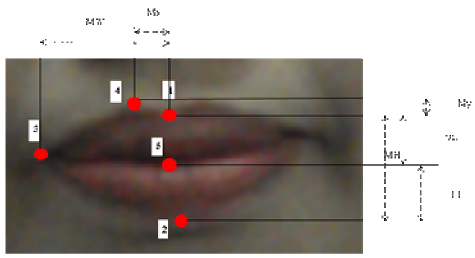


Figure 3. Mouth features and feature points

Points 1, 2 and 3 were tracked in the subsequent frames and their position was transferred to the mouth model. But other points were only used in the mouth frame for determining some features such as mouth wide, upper and lower lip thickness. Figure 4 shows the proposed mouth model. The coordinates of the vertices 1, 8 and 15 directly determined from the normalized position of feature points in the Figure 3 (UL were normalized to one). Coordinates of the other vertices were determined from the features MW, Mx, My, UL, LL and MH relative to 1, 8 and 15 vertices. Right hand side of the model is the symmetry of the left hand side.

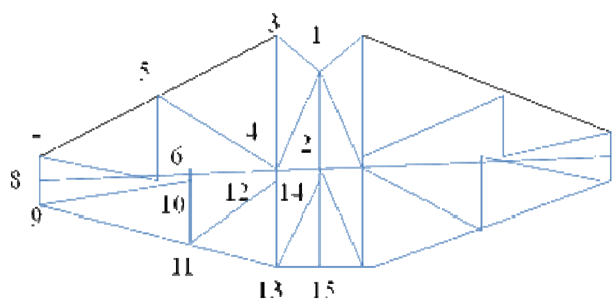


Figure 4. Eye Model

4.2. Eye Model and Eyebrow Model

Figure 5. shows the selected eye feature points and the features that were extracted from these points.

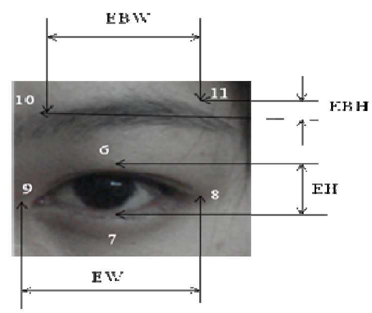


Figure 5. Eye and Eyebrow features and feature points

Points 6, 7, 10 and 11 were tracked in the subsequent frames and the other points were only used in the first frame for determining some features such as wide and height of the eye and eyebrow. Figure 6 shows the proposed eye and eyebrow model. The coordinates of the vertices 16, 17, 30 and 31 were directly determined from normalized position of the 6, 7, 10, and 11 feature points in the Figure 5. Coordinates of the other vertices were determined from features EH, EW, EBH and EBW relative to 16, 17, 30 and 31 vertices. For example wide of the iris assumed to be a half of the EW. Right hand side of the model is the symmetry of the left hand side. Face landmarks were determined by using 10 landmark points around the face in the first frame (points 12-21 in the Figure. 1). These points weren't tracked in subsequent frames, but the points around the jaw were synchronized with the movement of the point 2 in the Figure 3.

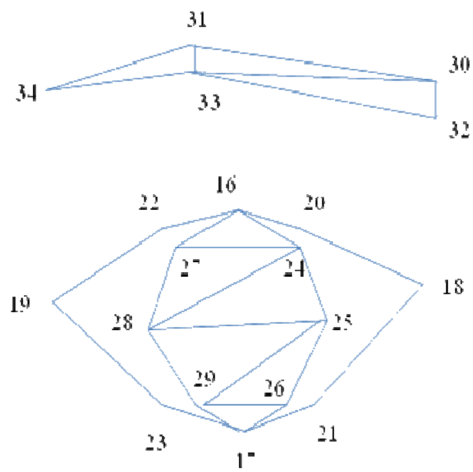
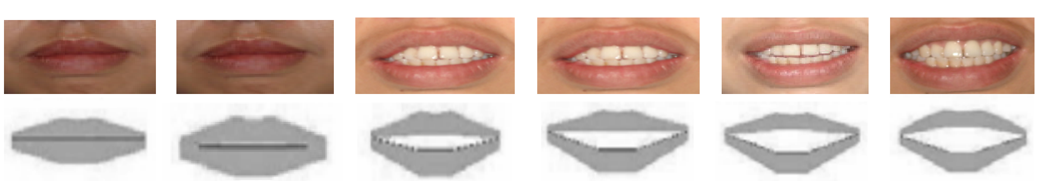


Figure.6 Eye and Eyebrow model

5. EXPERIMENTAL RESULTS

The proposed face model was evaluated with prepared database. Figure.7 shows the result of feature points tracking in the original face and the model deformation to show the same expressions. Algorithms have shown good robustness and reasonable accuracy for the photos from our test set. Tracked feature points in the original frames were also denoted in this figure. The proposed model also can be used for lip tracking and speech synchronizing facial animation system. Figure.7 shows the results of lip tracking for pronouncing the word "MinGaLarPar", that means "Hello" in Myanmar.



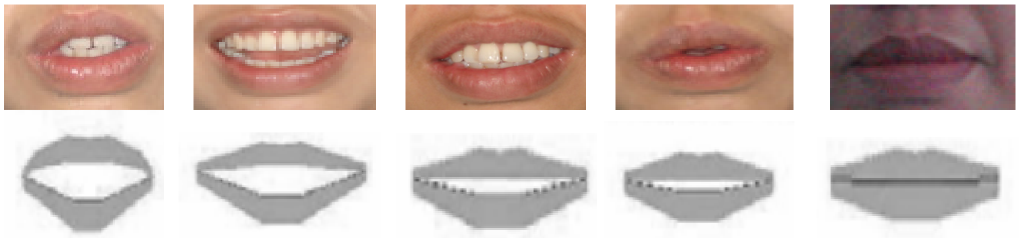


Figure 7. Lip tracking for pronouncing the word “MinGaLarPar”

6. CONCLUSION

In this paper, muscle-based face model that tracks some facial points in the real face image sequences and shows the same expressions. The proposed model had a simple structure and used a few set of control points comparing to the similar face models. The proposed model has a low complexity and is suitable for real time implementations, such as real time facial animation. Because of using the frontal images, we used a 2-D face model. By using the side images beside the frontal images and applying 3-D coordinates to the vertices of the model, 3-D face model can be obtained. The procedure is fully automatic, while user modification is also allowed if necessary. Also, the results can play a crucial role in feature-based face recognition system. Future work includes the investigation of additional features and the application of the method to the recognition of more naturalistic facial expression videos.

ACKNOWLEDGEMENTS

First of all, the author thanks Dr Nilar Thein, Rector for University of Computer Studies, Yangon, Myanmar. The author is also grateful to her family who specifically offered strong moral and physical support, care and kindness. Finally, the author is deeply thankful to Dr Myint Myint Sein, Professor, University of Computer Studies, Yangon.

REFERENCES

[1] Jinli Suo, Song-Chun Zhu, Shiguang Shan, Member, IEEE, and Xilin Chen, Member, IEEE, “A Compositional and Dynamic Model for Face Aging”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 32, No.3, March 2010.

[2] Hadi Seyedarabi, Ali Aghagolzadeh, and Sohrab Khanmohammadi, “ Facial Expressions Animation and Lip Tracking Using Facial Characteristic Points and Deformable Model”, *International Journal of Information and Communication Enginerring 1:8 2005.*

[3] Pohsiang Tsai, Tom Hintz, and Tony Jan, “Facial Behavior as Behavior Biometric? An Empirical Study”, 2007.

[4] Nabil Hewahi, Aya Olwan, Nebal Tubeel, Salha EL-Asar, Zeinab Abu-Sultan, “Age Estimation based on Neural Networks using Face Features”, *Journal of Emerging Trends in Computing and Information Science* , Vol.1 No. 2, October 2010.

[5] Elham Bagherian, Rahmita.Wirza.Rahmat and Nur Izura Udzir, “Extract of Facial Feature Point”, *International Journal of Computer Science and Network Security,* Vol.9, January 2009.

[6] Sebe, N., Sun, Y., Bakker, E., Lew, M., Cohen, I., Huang, T.: "Towards authentic emotion recognition", *International Conference on Systems, Man and Cybernetics.* (2004).

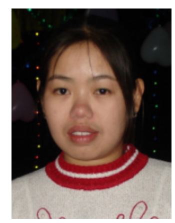
[7] Cohn, J.F , "Foundations of human computing: facial expression and emotion", *International Conference on Multimodal Interfaces*. (2006) 233-238.

[8] Z. Wu, P. S. Aleksic and A. K.Katsaggelos, " Lip tracking for MPEG-4 Facial Animation," in Proc. 4th IEEE Int. Conf. On Multimodal Interfaces, ICMI'02.

[9] F. Erol, "Modeling and Animating Personalized Faces", M.Sc. Thesis, Bilkent University, January 2001.

[10] S.T. Worral, A. H. Sadka, and A.M.Kondoz, "3-D Facial Animation for Very Low Bit Rate Mobile Video", *in Proc. IEEE Int.Conference on 3G Mobile Communication Technologies,* May 2002.

[11] Hussein Karam Hussein, "Towards Realistic Facial Modeling and Re-Rendering of Human Skin Aging Animation', *in Proc. IEEE Int. Conference of the Shape Modeling,* 2002.



Hlaing Htake Khaung Tin received the B.C.Sc and B.C.Sc (Hons:) degree from Government Computer College, Lashio, Northern Shan State, Myanmar in 2003, the Master of Computer Science degree from University of Computer Studies, Mandalay, Myanmar in 2006. She is currently attending her Ph.D at University of Computer Studies, Yangon, Myanmar. Her research interest mainly include face aging modelling, face recognition, and perception of human faces.