ENHANCED FACIAL EXPRESSION ANIMATION FOR ZAPIN DANCE USING RAY CASTING AND POSITION-BASED ESTIMATION

MUHAMMAD ANWAR BIN AHMAD

A thesis submitted in fulfilment of the
requirements for the examination of the degree of
Master of Philosophy

School of Computing,
Faculty of Engineering
Universiti Teknologi Malaysia

OCTOBER 2019

ABSTRACT

Currently, there are many efforts to preserve traditional dances digitally as part of the Intangible Cultural Heritage (ICH) preservation efforts. Such efforts include digital scans, videos, and animations in either 2D or 3D. Motion capture data is one of the common methods to animate the 3D dancer model. Nevertheless, they are usually limited to movement only and there is no facial expression data. For 3D animation, facial expression animation can be added to the digitalised 3D dancer model to increase the authenticity of the dance. Similar to body language, facial expressions in some dances are important to effectively communicate with the audience. Awkward facial expressions will make the performance look unnatural. This issue has led to this research's purpose, which is to enhance the process of adding facial expressions onto a virtual 3D dancer model using an algorithm that combines ray casting and position-aware concepts. The proposed algorithm was used to map the facial expressions according to the dancer's current position in certain segments of the dance. The primary goal of the algorithm is to enhance the process of animating facial expressions on a 3D Zapin dance motion data. Consequently, the animator does not need to manually animate the face on every keyframe in the animation or to capture expressions using facial motion capture systems due to high cost and complexity. In this scope of research, only the eyes and mouth were animated since they are the main focus of facial movements when the dancers perform. The algorithm and output were respectively evaluated using algorithm complexity and user evaluation tests. The complexity test was conducted using the Big O Notation method and it was found that the algorithm has the O(N) complexity, which makes it an efficient algorithm. The user evaluation test was performed by interviewing three Zapin experts from the Johor Heritage Foundation and they were mostly satisfied with the results.

ABSTRAK

Ketika ini, terdapat banyak usaha untuk memelihara tarian tradisional secara digital sebagai satu daripada Warisan Kebudayaan Tidak Ketara (ICH) seperti imbasan digital, video dan animasi 2D / 3D. Namun, pada kebiasaannya terhad kepada pergerakan sahaja dan tiada data ekspresi muka. Untuk animasi 3D, animasi ekspresi wajah boleh ditambah kepada model penari 3D digital yang meningkatkan keaslian tarian. Sama seperti bahasa badan, ekspresi muka dalam tarian adalah penting untuk agar mesejnya sampai dan difahami oleh penonton. Ekspresi muka yang janggal akan menjadikan persembahan kelihatan tidak semula jadi. Isu ini telah membawa kepada tujuan kajian ini, iaitu untuk menambahbaik proses penambahan ekspresi muka ke atas model 3D maya penari dengan menggunakan algoritma yang menggabungkan *ray casting* dan konsep anggaran kedudukan. Algoritma yang dicadangkan ini telah digunakan untuk memadankan ekspresi wajah mengikut kedudukan penari di dalam suatu segmen tarian. Matlamat utama algoritma ini adalah untuk menambahbaik proses penganimasian ekspresi muka untuk data pergerakan penari Zapin 3D. Hasilnya, juruanimasi tidak perlu memasukkan ekspresi muka pada setiap *keyframes* dalam animasi secara manual atau menggunakan sistem *Motion Capture* muka disebabkan oleh kos yang tinggi dan memerlukan pemasangan yang kompleks. Dalam skop kajian ini, hanya mata dan mulut sahaja yang digunakan untuk animasi kerana ia adalah fokus utama semasa penari membuat persembahan. Hasil kajian diuji menggunakan ujian kerumitan algoritma dan penilaian pengguna. Ujian kerumitan dijalankan menggunakan kaedah Tatatanda O Besar dan hasil ujian ini memberikan kerumitan O(N) yang mana menunjukkan ia adalah algoritma yang efisien. Penilaian pengguna pula telah dijalankan dengan menemubual beberapa pakar tarian Zapin dari Yayasan Warisan Johor dan secara umumnya mereka berpuas hati dengan dapatan kajian.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVATIONS

ICH	-	Intangible Cultural Heritage
FACS	-	Facial Action Coding System
AU	-	Action Units
PCA	-	Principal Component Analysis
RBF	-	Radial Basis Functions
XML	-	Extensible Markup Language
VHML	-	Virtual Human Markup Language
TH	-	Talking Head
DMML	-	Dialogue Manager Markup Language
FAML	-	Facial Animation Markup Languag
BAML	-	Body Animation Markup Language
SML	-	Speech Markup Language
EML	-	Emotion Markup Language
GML	-	Gesture Markup Language
FML	-	Face Modelling Language
FAP	-	Facial Animation Parameter
BEAT	-	Behavior Expression Animation Toolkit
DEM	-	Dynamic 3D Expression Model
APA	-	American Psychological Association
AI	-	Artificial Intelligence
BVH	-	Biovision Hierarchy
FBX	-	Filmbox
PAL	-	Palmer Heuristic
CDS	-	Campbel, Dudek and Smith Heuristic
DAN	-	Dannenbring Heuristic
A1	-	A1 Heuristic

LIST OF APPENDICES

APPENDIX	TITLE	PAGE

CHAPTER 1

INTRODUCTION

1.1 Background of the Study

Facial expression is one of the most common form of communication for human beings. People can show their emotions and intentions non-verbally through facial expressions. These include but not limited to being happy, sad, angry, bored or sleepy. It also can show a certain reaction to an action performed on a person whether it is voluntary or not such as surprised or disgust. For example, when a person is suddenly inflicted with a certain damage such as being hit, he or she may show the facial expression of pain. These expressions are important to us as human beings as it is meaningful towards others in understanding each other.

Research on facial expression has been done extensively by psychology researchers. Paul Ekman is one of the most prominent name in this field. He began his research on facial expressions in 1965 and has produced a number of books and articles since then. He also has successfully classified human facial expressions objectively through his Facial Action Coding System (FACS) project. It is developed alongside Wallace Friesen in 1976. It functions by describing the face muscles' movement. FACS is useful because it does not assign any emotional meaning to an expression and purely objective by measuring only the frequency and intensity. It categorizes an expression by breaking it down to the smallest visible movement called Action Units

(AU) (Prince, Martin, & Messinger, 2015). A facial expression can contain few or many AU depending on the complexity of the expression. As of 2002, the 370 pages FACS manual can be purchased from the Paul Ekman Group, LLC which includes 27 face AU, 25 head and eye position AU and 28 miscellaneous AU.

Animating facial expression on a 3D dance animation can be a time-consuming and tedious process for animators. The challenge to achieve compelling and believable realism when animating face models with minimal complexity is still an ongoing hot topic in the computer graphics community (Deng et al, 2006). Many researchers had proposed their toolkits, software and algorithms on how to properly imitate a real human's facial expression in 3D world space. There were a lot of significant results; however there are still more rooms for improvements. One area that has the potential to be researched is the facial expressions on dancers. Similar to body language, facial expressions are important to communicate with the audience. Awkward facial expressions will create an uncomfortable experience to the dancer and thus making the performance look unnatural (Bedinghaus, 2017).

The same principles apply to this research's domain, which is dance. Dancers utilize a lot of area in the dance floor. In some type of dances such as Bharatanatyam, they also make different facial expressions when they move to certain positions (O'Shea & Verne, 2007). Therefore, this research's purpose is to enhance the process of adding facial expressions onto a virtual 3D model of a dancer that contains dance motion data without any facial expressions using ray casting position-aware algorithm. The algorithm can map facial expressions according to the current position of the dancer from start to finish of the dance. This will enhance the animation process by removing the need to animate the face on every key frames in the animation. Eye movement and smiling are the main components of a facial expression. Therefore, this research focuses on these two areas only.

1.2 Problem Statement

Facial expression is a complex part of the human to be studied. A simple smile requires the movement of several face muscles. Translating these movements into 3D world space requires extensive research. To date, there are still many parts of the facial expression research field that can be explored such as dance facial expression.

Facial expression generally plays an important role in making a dance performance interesting. It is a form of communication to the audience to show as if they are conversing with them keeps the audience involved in the performance. Non-verbal cues relay the dancer’s emotion to the audience. Some dances such as the Indian dance Bharatanatyam even have facial expressions as a major part of the dance (O’Shea & Verne, 2007). Ballet dancing also makes use of emotional expressions. In dance competitions, judges can easily notice if the dancer is stressed or uncomfortable just by looking at their facial expressions (Vickers, 2015). Therefore, digital dance representations need to include facial expressions whenever it is possible.

Dance motion data collection has been conducted by many researchers on various types of dance performances. The use of motion capture system for capturing dance motion version of a dance has been a norm in the computer graphics industry. Some motion data are free to access but usually they need to be purchased. However, usually they just record the body movement only while leaving the face expressionless. A survey on several dance motion databases such as DanceDB's website, Hacettepe University's Midas Human Motion Database website, and CMU Graphics Lab Motion Database show that they captured only the body motion of the dances. The lack of facial expressions on these motion data is an issue that needs to be addressed.

While it is possible to capture facial expressions using motion capture systems, usually it is not easily accessible to the masses. For example, the Vicon MoCap system has only one distributor in Malaysia. The studios that have Vicon setups are also limited. Currently, UPM, Kolej Kemahiran Tinggi MARA, UTM and Akademi Seni Budaya dan Warisan Kebangsaan (ASWARA) offers Vicon system for rental either at high cost or strictly just for research only. Furthermore, sometimes they are under maintenance so it is not always available for access. There is also an issue on the mobility of dancers when using this kind of motion capture system since they constantly need to move around during the capture process.

In addition, according to a dance expert representative from Johor Cultural Heritage Foundation, all dance definitely need facial expressions to make the performance livelier. As the goal of preserving Intangible Cultural Heritage (ICH) is to make the digital dance look as realistic and authentic as possible, then the facial expression should be included as well.

The issue of animating 3D facial expressions is that they usually need to be added manually by animators. This process takes a lot of time and labour because the animator needs to animate it frame by frame. Therefore, by enhancing the mapping process of facial expression data to a dance motion data through an estimation algorithm, the process of creating near-realistic virtual dance performance can be enhanced.

Eyes and mouth plays the most important role in conveying the emotions in a facial expression. A study by Koch in 2006 suggests that the perception of emotions of a facial expression is equally influenced by the eyes and the mouth. This can be further backed by a research that is on a more neurological standpoint by Watanabe, Miki and Kakigi (2005). They stated that eye and mouth movements are important for the human to perceive emotions in facial expressions. Therefore, this research will mainly focus on animating the eyes and mouth.

Automatic facial expression generation in dance has been done by a previous research. This research focuses on generating expressions based on music cues. The facial expression changes based on the current emotion and mood of the input music. (Asahina et al., 2015). From further readings, there are no further enhancements on this research, therefore this research intends to explore more on this gap by using position-aware concept instead of music cues.

1.3 Aim

The aim of this research is to implement an algorithm that enhances the process of animating facial expression on a 3D Zapin dance motion data.

1.4 Objectives

The main goal of this research is to accelerate the process of mapping facial expression to a Zapin dance motion data. In order to achieve this goal, the objectives below must be completed:

1. To establish a set of facial expressions related to Zapin dance in 3D using 3D modelling software and Zapin dance segment motion data in the pre-processing step.

2. To implement a facial expression mapping algorithm that automatically maps a suitable expression on an animated dance model using a ray casting position-aware algorithm.
3. To evaluate the algorithm of facial expression mapping based on the output generated using expert evaluation and algorithm complexity test.

1.5 Scope of Research

This research focuses on proposing the method of mapping suitable facial expression to an animated 3D dancer model in a dance segment motion data. The mapping process is achieved by utilizing an algorithm to estimate the current position of the dancer model and automatically choose the corresponding facial expression from a facial expression database.

The facial expressions are prepared in a 3D modelling software. The facial expressions are limited to eye and mouth movement only. This is based on study by Koch in 2006 that proved that eyes and mouth movement plays an important role in the perception of emotion in facial expressions.

The test data used is a motion data of a Zapin dance. The motion data is attached to a sample 3D dancer model that contains a preset facial expressions called blendshapes that were prepared in the pre-processing stage. The 3D dancer model does not include costume since it does not affect the performance. This is confirmed by a preliminary interview with Zapin dance experts in Johor Heritage Foundation.

Zapin dancers smile when they look at the audience and their eyes look according to their turning movements. For example if they turn left then their eyes also look left and vice versa. Therefore these actions were selected as the triggers that allow the algorithm to apply the correct expressions.

1.6 Significances of the Research

The algorithm provides an enhanced facial expression generation for dance motion data. It is executed in a software on a dance segment and automatically applies the suitable facial movement therefore reducing the time and labor work needed to manually animate the facial movements.

Currently there is a gap in the automatic facial expression generation related to dance since the most current method uses only music and emotional cue (Asahina, 2015). Therefore, this research provides a different way to tackle the issue.

Furthermore, currently there is an effort to the preservation of Intangible Cultural Heritage (ICH). From this research, it is a step further to enhance and simplify the process of making a complete and near-realistic dance animation, which will enhance the authenticity of the dance in digital form.

1.7 Summary

The research aims to propose a method of mapping of facial expressions to a dance motion data through a ray casting position-aware algorithm. Chapter 2 discusses on the previous research on facial expressions to determine the methodologies and issues of this research field. The methodologies and implementation of the proposed algorithm are discussed in Chapter 3 and Chapter 4. The results of this research were analysed and discussed in Chapter 5. Chapter 6 discusses the achievements and contributions of this research.

REFERENCES

Anjali. (2008). Navarasa - The Nine Moods. Retrieved June 1, 2017, from https://onlinebharatanatyam.com/2008/03/10/navarasa-the-nine-moods/

Arya, A., & DiPaola, S. (2007). Face modeling and animation language for MPEG-4 XMT framework. *IEEE Transactions on Multimedia*, *9*(6), 1137–1146. https://doi.org/10.1109/TMM.2007.902862

Asahina, W., Okada, N., Iwamoto, N., Masuda, T., Fukusato, T., & Morishima, S. (2015). Automatic facial animation generation system of dancing characters considering emotion in dance and music. *SIGGRAPH Asia 2015 Posters on - SA '15*, 1–1. https://doi.org/10.1145/2820926.2820935

Bao, J., Zheng, Y., & Mokbel, M. F. (2012). Location-based and preference-aware recommendation using sparse geo-social networking data. *Proceedings of the 20th International Conference on Advances in Geographic Information Systems - SIGSPATIAL '12*, 199. https://doi.org/10.1145/2424321.2424348

Bedinghaus, T. (2017). Ballet Dance Expressions. Retrieved June 1, 2017, from https://www.thoughtco.com/dance-expressions-1007476

Bell, R. (2009). A beginner's guide to Big O notation. Retrieved from https://rob-bell.net/2009/06/a-beginners-guide-to-big-o-notation/

Bouaziz, S., Wang, Y., & Pauly, M. (2013). Online modeling for realtime facial animation. *ACM Transactions on Graphics*, *32*(4), 1. https://doi.org/10.1145/2461912.2461976

Cao, C., Weng, Y., Zhou, S., Tong, Y., & Zhou, K. (n.d.). FaceWarehouse: a 3D Facial Expression Database for Visual Computing, 1–11. Retrieved from http://www.kunzhou.net/2012/facewarehouse-tr.pdf

Carnegie Mellon University. (n.d.). CMU Graphics Lab Motion Capture Database. Retrieved from http://mocap.cs.cmu.edu/search.php?maincat=4&subcat=2

Cassell, J., Vilhjálmsson, H. H., & Bickmore, T. (2004). BEAT: the Behavior

Expression Animation Toolkit. *Life-Like Characters*, *137*(August), 163–185. https://doi.org/10.1007/978-3-662-08373-4_8

Chai, J., Xiao, J., & Hodgins, J. (2003). Vision-based Control of 3D Facial Animation. *Proc. of ACM SIGGRAPH/Eurographics Symp. on Comput. Anim.*, 193–206.

Chi, J., Gao, S., & Zhang, C. (2017). Interactive facial expression editing based on spatio-temporal coherency. *The Visual Computer*, *33*(6–8), 981–991. https://doi.org/10.1007/s00371-017-1387-4

Deng, Z., & Pei-Ying Chiang USC Pamela Fox USC Ulrich Neumann USC, U. (2006). Animating Blendshape Faces by Cross-Mapping Motion Capture Data. *I3D*, *2006*(March), 7. https://doi.org/10.1145/1111411.1111419

DiPaola, S., & Arya, A. (2006). Emotional remapping of music to facial animation. *Proceedings of the 2006 ACM SIGGRAPH Symposium on Videogames - Sandbox '06*, *1*(212), 143–149. https://doi.org/10.1145/1183316.1183337

Dyck, E. Van, Burger, B., & Orlandatou, K. (2017). THE COMMUNICATION OF EMOTIONS IN DANCE, 122–130.

Ekman, P. (1993). Facial Expression and Emotion. American Psychologist, 384-392.

Ekman, P., & Friesen, W. V. (1976). Measuring Facial Movement. *Environmental Psychology and Nonverbal Behavior 1(1), Fall 1976*. https://doi.org/10.1109/RE.2006.19

Hacettepe University. (n.d.). Human Motion Database. Retrieved from http://www.motion.hacettepe.edu.tr/?c=dance

Homans, J. (2010). *Apollo's angels : a history of ballet*. New York: Random House.

Iwamoto, N., Kato, T., & Shum, H. P. H. (2017). DanceDJ : A 3D Dance Animation Authoring System for Live Performance.

Koch, C. (2006). The Role of the Eyes and Mouth in Facial Emotions.

Kohata, S. (2018). An Interactive Sound Dystopia: Real-Time Audio Processing in "NieR:Automata." Retrieved April 12, 2018, from https://www.gdcvault.com/play/1025401/An-Interactive-Sound-Dystopia-Real

Lab, U.C.G.(n.d.) .DanceDB. Retrieved from https://dancedb.cs.ucy.ac.cy/main/performances

Lewis, J., Anjyo, K., Rhee, T., Zhang, M., Pighin, F., & Deng, Z. (2014). Practice and Theory of Blendshape Facial Models. *Eurographics*, 199–218. https://doi.org/10.2312/egst.20141042

Lewis, J. P., Mooser, J., Deng, Z., & Neumann, U. (2005). Reducing blendshape interference by selected motion attenuation. *Proceedings of the 2005 Symposium on Interactive 3D Graphics and Games - SI3D '05*, 25. https://doi.org/10.1145/1053427.1053431

Vandevenne, L. (2004). Raycasting. Retrieved February 21, 2018, from http://lodev.org/cgtutor/raycasting.html

Macquarie Motion Capture Manual. (n.d.). Retrieved September 3, 2019, from http://psy.mq.edu.au/me2/mocap_v1/introduction.html

Marriott, A. (2001). VHML–Virtual Human Markup Language. *Online), Http://Www. Interface. Domputing. Edu. Au/Document/VHML*, (Xml), 2002. Retrieved from http://www.talkingheads.computing.edu.au/documents/workshops/TalkingHeadTechnologyWorkshop/workshop/marriott/vhml_workshop.pdf

Matsumoto, D., & Hwang, H. S. (2011). Reading facial expressions of emotion. Retrieved November 2, 2018, from https://www.apa.org/science/about/psa/2011/05/facial-expressions.aspx

Narang, S., Besty, A., Randhavanez, T., Shapirox, A., & Manocha, D. (2016). PedVR: Simulating gaze-based interactions between a real user and virtual crowds. *Proceedings of the ACM Symposium on Virtual Reality Software and Technology, VRST, 02–04–Nove*, 91–100. https://doi.org/10.1145/2993369.2993378

Nor, M. A. M. (1993). *Zapin, Folk Dance of the Malay World*. Oxford University Press.

O'Shea, J., & Verne, J. (2007). *At Home in the World: Bharata Natyam on the Global Stage*. Wesleyan University Press.

Ondřej, J., Pettré, J., Olivier, A.-H., & Donikian, S. (2010). A synthetic-vision based steering approach for crowd simulation. *ACM Transactions on Graphics, 29*(4), 1. https://doi.org/10.1145/1833351.1778860

Oyetunji, E. O., & Oluleye, A. E. (2009). Comparison of the Running Time of Some

Flow Shop Scheduling Algorithms using the Big O Notation Method, *1*(1), 4–8.

Paier, W., Kettern, M., Hilsmann, A., & Eisert, P. (2015). Video-based Facial Re-animation. *Proceedings of the 12th European Conference on Visual Media Production*, 4:1--4:10. https://doi.org/10.1145/2824840.2824843

Pandzic, I. S., & Forcheimer, R. (2002). MPEG4 Facial Animation - The standard, implementations and applications.

Prince, E. B., Martin, K. B., & Messinger, D. S. (2015). Facial Action Coding System.

Saragih, J. M., Lucey, S., & Cohn, J. F. (2011). Real-time avatar animation from a single image. *2011 IEEE International Conference on Automatic Face and Gesture Recognition and Workshops, FG 2011*, 117–124. https://doi.org/10.1109/FG.2011.5771383

Struijk, S. Van Der. (2018). FACSvatar : An Open Source Modular Framework for Real-Time FACS based Facial Animation. *IVA '18*, 159–164.

Umenhoffer, T., & Tóth, B. (2012). Facial animation retargeting framework using radial basis functions. *Sixth Hungarian Conference on Computer Graphics and Geometry*, 64–69.

Weise, T., Bouaziz, S., Li, H., & Pauly, M. (2011). Realtime performance-based facial animation. *ACM Transactions on Graphics*, *30*(4), 1. https://doi.org/10.1145/2010324.1964972

Yin, L., Wei, X., Sun, Y., Wang, J., & Rosato, M. J. (2006). A 3D facial expression database for facial behavior research. *FGR 2006: Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition*, *2006*, 211–216. https://doi.org/10.1109/FGR.2006.6

Zhang, X., Yin, L., Cohn, J. F., Canavan, S., Reale, M., Horowitz, A., Girard, J. M. (2014). BP4D-Spontaneous: A high-resolution spontaneous 3D dynamic facial expression database. *Image and Vision Computing*, *32*(10), 692–706. https://doi.org/10.1016/j.imavis.2014.06.002

Zhang, Z. (2012). Microsoft kinect sensor and its effect. *IEEE Multimedia*, *19*(2), 4–10. https://doi.org/10.1109/MMUL.2012.24