ENHANCED IMAGE-BASED ALGORITHM FOR CONSTRUCTION OF THREE DIMENSIONAL ZAPIN ANIMATION FROM MONOCULAR VIDEO

NIK MOHAMMAD WAFIY BIN AZMI

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

School of Computing
Faculty of Engineering
Universiti Teknologi Malaysia

JANUARY 2022

DEDICATION

To my family and my teachers, Thank you.

ACKNOWLEDGEMENT

In preparing this thesis, I was in contact with many people, researchers, academicians, and practitioners. They have contributed towards my understanding and thoughts. In particular, I wish to express my sincere appreciation to my main thesis supervisor, Dr. Norhaida Binti Mohd Suaib, for encouragement, guidance, critics and friendship.

I am also indebted to Universiti Teknologi Malaysia (UTM) and also Bournemouth University for giving me an opportunity and guidance especially in terms of research facilities and related support. Next my appreciation and gratitude extended to all fellow lecturers that continuous supporting my journey in finishing this thesis

My fellow postgraduate student should also be recognised for their support. My sincere appreciation also extends to all my colleagues and others who have aided at various occasions. Their views and tips are useful indeed. Unfortunately, it is not possible to list all of them in this limited space. I am grateful to all my family member.

ABSTRACT

Extracting human motion data from body posture or movements has gained popularity in computer and multimedia industry nowadays. Industries such as entertainment, gaming and Artificial Intelligence demand the technology that can provide them with the most fast and accurate motion capture technology available. This is due to the high needs of the motion data which can be used to increase the naturalness of the graphic contents and also the automation process involving human interaction. However, the technology of the motion capture is very expensive due to its requirement for proper space and high-end equipment. In addition, the motion capture technology available today require substantial amount of time in order to complete one production of the motion data this is not entirely because of the limitation of the devices and system but mostly because the technology needs the help of specific experts to be operational. Therefore, this research addressed the issue by introducing an improved algorithm in image-based process to extract the motion data as an alternative to the motion capture technology. The approach of the research is to use an enhanced image-based algorithm to improve previous image-based processing technique in generating 3D parameters which can be used to create 3D animation. The research was also done with the intention of proposing a new methodology in preserving the Intangible Cultural Heritage in Malaysia. The general framework of the research includes three main phases namely features/keypoints extraction, visualization of 3D skeleton and lastly evaluation of research's results. Outcome from these processes is a 3D skeleton along with motion information extracted from monocular video. The evaluation of the research consists of qualitative and quantitative analysis which involved the videos and results comparison, keypoints evaluation and expert testing. All these evaluations were made in order to testify the accuracy and the satisfaction towards the results of the research. The research managed to produce an accurate 3D skeleton animation based on the movements of the Zapin dance.

ABSTRAK

Mengekstrak data pergerakan manusia daripada postur badan atau pergerakan telah mendapat populariti dalam industri komputer dan multimedia pada masa kini. Industri seperti hiburan, permainan dan Kecerdasan Buatan menuntut teknologi yang boleh memberikan mereka kaedah tangkapan gerakan paling pantas dan tepat yang ada. Ini disebabkan oleh keperluan data gerakan yang tinggi yang boleh digunakan untuk meningkatkan keaslian kandungan grafik dan juga proses automasi yang melibatkan interaksi manusia. Walau bagaimanapun, teknologi tangkapan gerakan adalah sangat mahal kerana keperluannya untuk ruang yang sesuai dan peralatan mewah. Di samping itu, teknologi tangkapan gerakan yang ada hari ini memerlukan masa yang banyak untuk menyelesaikan satu pengeluaran data gerakan, ini bukan sepenuhnya disebabkan oleh keterbatasan peranti dan sistem tetapi kebanyakannya kerana teknologi memerlukan bantuan pakar khusus untuk beroperasi. Oleh itu, penyelidikan ini menangani isu tersebut dengan memperkenalkan algoritma yang dipertingkatkan dalam proses berasaskan imej untuk mengekstrak data gerakan sebagai alternatif kepada teknologi tangkapan gerakan. Pendekatan kajian adalah menggunakan algoritma berasaskan imej yang dipertingkatkan untuk menambah baik teknik pemprosesan berasaskan imej sebelum ini dalam menghasilkan parameter 3D yang boleh digunakan untuk mencipta animasi 3D. Penyelidikan juga dilakukan dengan tujuan untuk mencadangkan satu metodologi baharu dalam memelihara warisan budaya di Malaysia. Rangka kerja am penyelidikan merangkumi 3 fasa utama iaitu pengekstrakan ciri/titik kekunci, visualisasi rangka 3D dan terakhir penilaian hasil penyelidikan. Hasil daripada proses ini ialah rangka 3D bersama dengan maklumat gerakan yang diekstrak daripada video monokular. Penilaian penyelidikan terdiri daripada analisis kualitatif dan kuantitatif yang melibatkan perbandingan video dan keputusan, penilaian titik utama dan ujian pakar. Semua penilaian ini dibuat untuk membuktikan ketepatan dan kepuasan terhadap hasil penyelidikan. Penyelidikan akhirnya berjaya menghasilkan animasi rangka 3D yang tepat berdasarkan pergerakan tarian Zapin.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

2D	-	Two-Dimension
3D	-	Three-Dimension
CNN	-	Convolutional Neural Network
HIK	-	Human Inverse Kinematic
PAF	-	Part Affinity Field
PCA	-	Principle Component Analysis
PC	-	Personal Computer
ICH	-	Intagible Cultural Heritage
MOCAP		Motion Capture
RNN	-	Recurrent Neural Network

LIST OF SYMBOLS

δ	-	Minimal error
D, d	-	Diameter
F	-	Force
v	-	Velocity
p	-	Pressure
I	-	Moment of Inersia
r	-	Radius
Re	-	Reynold Number

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Overview

Cultural heritage can be categorized in two parts, first is material culture and second is immaterial culture. Material culture can be viewed primarily in the form of monuments, historic sites, and artifacts (Kitsikidis *et al.*, 2017). While immaterial culture can be referred as the cultural expressions of activity composed by gesture, language, values, norms sanctions and folkways. Intangible Cultural Heritage (ICH) can be referred as a combination of immaterial manifestation which represent the variety of living heritage of humanity as well as the most important vehicle of cultural diversity (Lenzerini *et al.*, 2017).

The topic on safeguarding the ICH has become one of the primary concern of UNESCO (United Nations Educational, Scientific and Cultural Organisation), on 32th Convention for the Safeguarding of the Intangible Cultural Heritage 2003, most of the UNESCO community agreed that ICH materials such as music, crafts and dances are equally important to the tangible heritage, which needs and deserve international safeguarding due to structured model designed by the 1972 world heritage convention has become inadequate with the current situations in preserving and safeguarding the specificities of ICH (Lenzerini *et al.*, 2017). The model only focused on tangible heritage such as artifacts, buildings, and post-war monuments, and does not include amounts of immaterial components which considered by some community as part of their cultural identity. Due to this issue, the governmental and non-governmental organizations along with academic institutions around the world have been involved

in acquiring and documenting the data of ICH where the process can lead to multiple breakthroughs in ICH preservation techniques.

Dance is a type of interactive activity that requires the movement of several limbs of human body. When human dance, it will alleviate the sense of good feelings, this much explained why dancing is serves as the medium for entertainment by the people from the past. Almost all of the countries in this world have their own unique type of dances which acts as their own identities, their own cultural heritage. This heritage is the pure forms of evidence that shows the countries are not far behind the civilizations. For certain countries, their folk dances are used as the ritual for their cultural worship, celebrations, and ceremonies. There also an artistic dance that is well rehearsed and performed to meets the eye of the audience.

In the perspective of ICH, traditional dance is an autonomous form of art and expression (Kico *et al.*, 2018). Through the ages, some cultures have been representing their feeling and believes in forms of dancing, this shows that dancing has become one of the important parts in ICH, which directly connected to local identity and culture (Voulodimos *et al.*, 2016). In South-east Asia, there is Zapin dance which originated from the Middle East and introduced by the Arab missionaries during fourteenth century. According to Nor (1993), the word "Zapin" itself is referred to the term in Arabic root words "Zaffa" which means to lead a bride to her groom in a wedding ceremony.

In Malaysia, Zapin is famously known as the part of cultural dance practiced by the Malay people in the state of Johor. At the earlier stage, Zapin has only been performed by the male dancer as permitted by the Islamic code for entertainment and it is often performed at the wedding rather than any other occasions (Daud. 2000). After the second World War, the mixed-gender participation of the Zapin dance has become the norm. However, there is assurance of no physical contact during the performance thus, the traditional values, social decency, and propriety of the traditional Zapin dance is still intact.

The evolution that takes place in the Zapin has made the dance to be a performance deemed for viewing instead of participated by the people. This is because, the contemporary Zapin will need a rehearsed choreography to ensure the perfection of the performance, and it is no longer served for the entertainment of peoples during wedding occasion only where all the people can also join the dance, but more focused on a larger crowded in the special events.

As with other forms of ICH, Zapin also faces the risk of extinction in terms of the continuation of the legacy itself as there is only few practitioners left to hold the duty of preserving the heritage since there is a lack of participation from the young generations. Learning Zapin dance is not an easy task as it needs a full commitment of the dancer. Dancers need to abide all the training sessions which costs more than hours of training, moreover, an introductory learning programme to our cultural dance heritage does not include in the national educational system which making it less familiar among Malaysian people especially younger generations. It is also hard for people to learn these dances as the classes are only limited to certain club and organization.

With the emergence of technologies in the field of Computer Graphics and Computer Vision, intangible heritage resources such as Zapin dance can be possibly preserved by digitizing the content into the form that can be processed and enhanced by the computer. For example, the introduction of motion capture technology (mocap) has helped lots of researcher to speed up the process of digitizing any motion and gesture by creating the computer animation based on recorded and real-time dataset. In the prospect of Computer-Generated Animation (CGI) nowadays, mocap can be regarded as one of the fastest growth technologies available right now which capable of tracking the motion of the human and changed the data into human motion character animation.

Animation is known by many as one of the fastest growth industries nowadays. The foundation of the animation is through the concept called 'persistence of vision' which comes of the process where the retina in human eyes retains an image in a brief split-second after the actual image. In Malaysia, animation has been regarded as one of the storytelling mediums for generations (Azmi, 2014). For example, the animation of *Hikayat Sang kancil* (1983), and *Silat Lagenda* (1997) is the examples of animation products which have been used as folklore storyteller by the national broadcasting before. The animation which was an animal-based story were used as an approach to re-tell the stories concerning human behaviour (Norafizah *et al.*, 2018).

With the excellent acceptance of animation among the Malaysians, the research believes that with the advancement of animation technologies available nowadays, the method of digitizing the ICH materials such as Zapin dance into 3D animation can be achieve. Furthermore, the traditional dance like Zapin is in verge of threat and forgotten by the communities due to the modernization and changes in cultures and belief, most of the younger generations today does not have a knowledge about our own cultural dance and more drawn to modern dance. This is an issue that must be addressed for the benefits of our cultural heritage, and in line with this, the research on producing a 3D skeletal animation on Zapin dance was initiated.

1.2 Background of The Study

There were several efforts based on the mocap technology were done by previous researchers to use animation as a method of preserving their own cultural heritage. For example, Stavrakis *et al.* (2012) in his effort on digitization of "Cypriot Folk" dances has introduced the use of Phasespace motion capture with the materials of the dance video held by the cultural institutions, to record and preserved the motion data of the dance. Meanwhile in Japan, Oshita (2012) with his research developed the animation authoring tool for the Noh traditional Japanese dance. These Noh dances are defined by the traditional notations called "Katasuke" which is the important reference for the motion unit composition process to form the true traditional Noh

dance. The research had shown the importance of understanding the nature of the dance in assisting the development process. Mandery *et al.* (2015) with his research had developed a KIT Whole Body Human Motion Database that stored the whole-body actions of human. The recorded motion data are classified within the Motion Description Tree, the hierarchical tags that describes motion properties such as speed, movement type and direction during the dance.

However, most of the mocap technologies nowadays are not easily accessible and expensive (Tung *et al.*, 2017). This is due to the technology demands numbers of expertise to perform device calibration and there is huge list of hardware requirements that need to be fulfilled in order to perform the tracking and capturing process. Also, the huge cost on performing the mocap has become one of the major reasons why the technology is not on the first list of many researchers to work with. In addition, to cover all the expenses from the mocap process, most of the works produced will be issued as private database/archives by the corresponding organization, and any party that want to access the data need to pay huge amount of money. This will lead to limited dataset for researcher as most of the researchers does not have sufficient fund or large research grant to spend on dataset. With all the issues found in using the mocap, alternative methods were suggested in order to capture the motion and gesture of the dance, first is by using keyframe animation technique and second by using image-based technique which usually done in machine learning fields.

3D keyframe animation is an approach that exist in the field of computer animation which can be used in digitizing the contents of ICH. Keyframe is a method which is used to create animation by altering the key or condition of the object's transition from early state to end state. It is one of the simplest and most famous technique to create an animation nowadays. There were numbers of past research which focused on 3D keyframe animation techniques to produce the motion data of the human body. Even though the research objectives do not directly address the ICH preservation issues, such as Zhou *et al.* (2020) with their research on animation transfer, Ali-Hamadi *et al.* (2013) and Chang *et al.* (2006) with their research on joints retargeting attributes, the used of keyframe technique does produces a decent result

with less cost compared to mocap approach. However, the problems found in keyframe animation technique is, the method consumed lots of time to produce an animation, normally a full equipped studio with several animators are needed, and the animation produced cannot be recycled to save the time (Holmqvist et al., 2017).

Several initiatives to overcome the issues mentioned above has been initiated by numbers of researchers, the main idea is to use less-expensive data such as recorded video data produced by monocular device like camera phone and standard digital camera to replace the 360^0 camera setup used in mocap environment in order to obtain the motion data (Li *et al.*, 2018). Second, Sminchisescu *et al.* (2002) presents a non-deep learning approach to extract the human silhouette from an image sequence to acquire the pose estimation, pose estimation is a process of analysing movements or transformation of an object from the model. The continuation of the process is done by Bureniu *et al.* (2013) where his research proposed to replace the model based used by Sminchisescu *et al.* (2012) to pictorial structure based. Third, the approach on using deep learning method has been developed to improve the detection and tracking performance on previous approach. Toshev *et al.* (2014) used DeepPose to gather the pose of the single person using Dynamic Neural Network (DNN). The research based on deep learning approach then was expanded with another breakthrough of using Convolutional Neural Network (CNN) which improves the accuracy of object classification significantly. The earlier used of CNN can be seen done by Wei *et al.* (2016) with his approach on using heatmap prediction to produce keypoints. Keypoints is the total numbers of interest points that can be found in the whole sequence of images, it consists of the spatial locations or points in the image which considered as interesting. The acquisition of keypoints is vital in the process image-based approach as it helps with the fine-grained classification Guo and Farell. (2019) and also re-identification of the dataset (Zhu *et al.*, 2020). Next, Cao *et al.* (2018) proposed the real-time multi person pose estimation which can be used to generate multiple keypoints on multiple persons in single image by applying Part Affinity Fields (PAFs).

Based on the discussion above, lots of research managed to find the alternatives method on getting the motion data by using single image from sequences of image captured by monocular device. However, the complexity in estimating the human pose

estimation is high especially when the process involves in taking potentially fast movement and considerable non-rigid deformation like dancing steps (Xu *et al.*, 2018).

1.3 Problem Statement

With the abundance of the videos available on the internet platform such as YouTube, Facebook and other open-source websites, the videos of the traditional dance such as Zapin can easily be found, and these videos can be used as a dataset in acquiring the motion data via image-based processing technique. The process of recording a video using standard recording device also does not generate high cost as it can be done by using standard device like smartphone or video recorder. A research on capturing the motion data based on image processing by acquiring the human pose estimation on dance using deep learning technique has been done by several researchers, such as He *et al.* (2018) with his research on applying the Recurrent-Neural Networ (R-NN) in order to acquire the motion data of solo dancing sequence and Hou *et al.*, (2017) with his research on estimating the pose in the dance video using Convolutional Pose Machines (CPMs) to create a real-time dance guidance application. However, the image-based method only focused on gathering the information of the interest points in the video called as keypoints which usually used in detection and recognition application and does not cover the production of 3D mesh to contain these keypoints. In order to address this issue, there is a need to propose an improved algorithms to extract and construct a 3D skeletal animation based on image-based approach such as Convolutional Neural Network (CNN), so that the motion data extracted from the videos can be visualize. The research believes that, by an improved creative method on digitizing and visualization of Zapin, the legacy of the Zapin dance can be preserved for future generations.

1.4 Aim

This research proposed an improved algorithm to extract a 3D skeleton motion data using monocular video.

1.5 Research Objectives

Following are the objectives of this research:

- To study and identify a suitable image-based processing approach in extracting Zapin traditional dance's motion data.
- To develop and to test an enhanced image-based processing algorithm using extracted dataset from monocular video.
- To validate the effectiveness of enhanced image-based processing algorithm in constructing 3D skeleton animation from monocular video.

1.6 Scope of The Research

To achieve the objectives as stated above, this research is limited to the following scope:

I. **Data:** Malaysia Traditional Zapin dance as a main dataset as an effort to preserve the originality of the dance from extinction. The videos used in the research were recorded using High Definition (HD) Video Camera.

II. **Data:** The mapping of the joints detection was represented by MPI model skeleton. The data are provided by MPII human pose dataset.

III. **Data:** The mapping of the joints validation was represented by COCO model skeleton. The data are provided by COCO dataset.org.

IV. **Method:** The proposed method highlights the enhancement of keypoints manipulation framework only and does not cover the accuracy of the generated keypoints between another approach.

V. **Method:** The projection of 3D matrices generated from 2D keypoints were encoded in Human Inverse Kinematic Architecture (HIK).

VI. **Evaluation:** The qualitative evaluation of the motion data were conducted through joints validation of the motion data.

VII. **Evaluation:** The quantitative evaluation of the data was conducted through the expert testing questionnaire.

1.7 Significance of The Research

As mentioned in previous section, in the emergence of Computer Graphics and technology, the preservation of intangible cultural heritage dance such as Zapin can be done by digitizing all the datasets into processing-enabled form. The justification of this approach is due to the fact that the digital dataset is much more future-proof and can be archived in much more safe and accurate manner compared to traditional archive such as video clips and verbal recording from the instructors. In the research, the digitization of Malaysia Traditional Zapin dance was done through the process of motion extraction using keypoints acquisition from the dance videos. By extracting all the keypoints from the videos, the research managed to contain all available motion data from the video and used it to produce another form of digital process such as 3D animation reconstruction process.

Next, the research also managed to successfully propose and implemented an alternative method of extracting the motion data without relying on motion capture (mocap) technology. By extracting the motion data using only monocular videos, the cost of generating the motion data can be reduced significantly in terms of money and time in order to get the datasets. This alternative approach of motion extraction process achieved by the research will notably help numbers of under-budget researchers and animators to finish their works.

Third, the research also managed to highlight the framework of enhancing the usage of the keypoints. The implemented framework in the research has successfully showed that the generated keypoints which were usually used in the process of classification of the human pose estimation can be processed and enhanced to create a full 3D skeleton animation instead of only being used widely in detection and recognition process only. This will encourage more of researchers to focus on generating other sets of keypoints apart from Zapin dance, and these datasets can be shared among another party so that a big data archive regarding traditional dance keypoints can be created.

Lastly, the research also managed to show that the generated 3D skeleton animation produced using proposed framework moves according to the real Zapin dance accurately. By producing accurate animation movement, it is proven that the framework proposed by the research can be implemented in real-world situation involving dance movements.

REFERENCES

Kitsikidis, A., Alivizatou-Barakou, M., Tsalakinidou, F., Dimitropoulos, K., and Nikolopoulos, S, (2017) Intangible Cultural Heritage and New Technologies: Challenges and Opportunities for Cultural Preservation and Development. *Mixed Reality and Gamification for Cultural Heritage.* doi: 10.1007/978-3-319-49607-8_5.

Lenzerini, F., Francioni, F., (2017) The Obligation to Prevent and Avoid Destruction of Cultural Heritage: From Bamiyan to Iraq. *Cultural Heritage Rights.* ISBN: 9781315258737.

Kico, I., Grammalidis, N., Christidis, Y., Liarokapis, F, (2018) Digitization and Visualization of Folk Dances in Cultural Heritage: A review, *Innovation in Machine Intelligence for Critical Infrastructures.* https://doi.org/10.3390/inventions3040072

Voulodimos, A., Doulamis, N., Fritsch, D., Makantasis, K., Doulamis, A., and Klein, M, (2016) Four Dimensional Reconstruction of Cultural Heritage Sites Based on Photogrammetry and Clustering , *J Electron Imaging.* 26:011013.

Nur Azmi, Rahimah A. Hamid, Rohani Hashim (2014) Four Dimensional Reconstruction of Cultural Heritage Sites Based on Photogrammetry and Clustering , *J Electron Imaging.* 26:011013.

Norafizah M, B., Julina Ismail, K., Nurliana Yusri, (2018) Hikayat Sang Kancil and Buaya: an Interactive animation, *Advances in Social Sciences, Education and Humanities Research, Volume 207.* 3rd International Conference on Creative Media, Design and technology (REKA).

Stavrakis, E., Aristidou, A., Savva, M., Himona, S., and Chrysanthou, Y. (2012) Digitization of Cypriot Folk Dances, *Euromed.* doi:10.1007/978-3-642-34234-9_41.

Oshita, M., Yamanaka, R., Iwatsuki, M. (2012) Development of Easy-To-Use Authoring System for Noh (Japanese Traditional) Dance Animation, *IEE. 2012 International Conference on Cyberworlds.* doi: 10.1109/CW.2012.14.

Mandery, C., Terlemez, O., Do., M., Vahrenkamp, N., Asfour, T. (2015) The KIT Whole-Body Human Motion Database, *International Conference on Advanced Robotics (ICAR)*. doi: 10.1109/ICAR.2015.7251476.

Tung, H.-Y., Tung, H.-W., Yumer, E. and Fragkiadaki, K. (2017). Self-supervised learning of motion capture. *Advances in Neural Information Processing Systems*. 5236-5246.

Zhou, Yi., Lu, J., Barnes, C., Yang, J., Xiang, S., Li, H.,. (2020). Generative Tweening: Long-term Inbetweening of 3D Human Motions. *Computer Vision and Pattern Recognition (cs.CV); Graphics (cs.GR)*. arXiv:2005.08891.

Ali-Hamadi, D., Liu, T., Gilles, B., Kavan, L., Faure, F., Palombi, O., Paule Cani, M. (2013). Anatomy Transfer. *ACM Transactions on Graphics (Proceedings of ACM SIGGRAPH Asia 2013)*. doi:10.1145/2508363.2508415.

Chang, Y.-T., Chen, B.-Y., Luo, W.-C. and Huang, J.-B. (2006). Skeleton-driven animation transfer based on consistent volume parameterization. *Computer Graphics International Conference*. Springer, 78-89.

Holmqvist, L. and Ahlström. E. (2017). Comparing Traditional KeyFrame. *Animation Approach and Hybrid Animation Approach of Humanoid Characters*.

Sminchisescu, C., (2003) 'Kinematic jump processes for monocular 3d human tracking', *Computer Vision and Pattern Recognition, proceeding, IEEE Computer Society Conference, Vol 1, pp. 1-69*.

Burenius, M., Sullivan, J., and Carlsson, S. (2014) '3d pictorial structures for multiple view articulated pose estimation', in *CVPR 2013*.

Toshev, A., Szegedy, C.,. (2014). Deep Pose: Human Pose Estimation via Deep Learning. *IEEE Conference on Computer Vision and Pattern Recognition*. doi:10.1109/CVPR.2014.214.

Wei, S.E., Ramakrishna.,Kanade. (2016) 'Convolutional Pose Machine', *proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 4724-4732.

Wei, S.E., Ramakrishna.,Kanade. (2016) 'Convolutional Pose Machine', *proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 4724-4732.

Guo, P., Farrell, R., (2019). Aligned to The Object, Not to The Image: A Unified Pose-Aligned Representation For Fine-Grained Recognition. *IEEE Winter Conference on Applications of Computer Vision (WACV)*. 1876-1885.

Zhu, X., Wu, G., and Gong, S. (2020). Tracklet Self-Supervised Learning for Unsupervised Person Re-Identification. *Association for The Advancement of Artificial Intelligence*. Vol.34 No. 07: AAAI-20 Technical Tracks 7.

Cao, Z., Hidalgo, G., Simon, T., Wei, S., and Sheikh, Y. (2018) OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields, *CoRR.* abs/1812.0.

Xu, W., Chatterjee, A., Zollhöfer, M., Rhodin, H., Mehta, D., Seidel, H. P. and Theobalt, C.. (2018) MonoPerfCap: Human Performance Capture From Monocular Video, *ACM Transaction on Graphics.* 37(2).1-15. ISSN 15577368. doi: 10.1145/3181973.

He., k., Gkioxari., G.Dollar and Girshick. R. (2018) 'Mask R-CNN', *IEEE Trans Pattern Anal Mach Intell.*

Hou, Y., Yao, H., Li, H., and Sun, X. (2018) Dancing like a Superstar: Action guidance Based on Pose Estimation and Conditional Pose Alignment, *IEEE International Conference on Image Processing (ICIP).* 37(2).1-15. ISSN 15577368. doi: 10.1109/ICIP.2017.8296494.

Triyani, R., Masunah, J., and Nugraheni, T. (2020) The Uniqueness of Malay Zapin Dance Choreography, *Advances in Social Science, Education and Humanities Research.* Volume 519.

Sari, D. (2015) Reconstruction of The Zapin 12 Traditional Dance of Kuala Kampar, *Universiti Pendidikan Indonesia.*

Nor, M.A.M. (1993). *Zapin, folk dance of the Malay world.* Oxford University Press. USA.

Daud, T.R. (2000) Dasar Langkah Tari Zapin Riau dan Sekelumit Perkembangannya, *Zapin Melayu di Nusantara.* Pp 125-150.

Liu, Y., and Stoll, C. (2011) Markerless Motion Capture of Interacting Characters Using Multi-View Image Segmentation, *Proceeding / CVPR, IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society Conference on Computer Vision and Pattern Recognition.* doi:10.1109/CVPR2011.5995424.

Loper, M., Mahmood, N., Romero, J., And Pons-Moll, G. (2014) SMPL: A Skinned Multi-Person Linear Model, *ACM Transaction on Graphics 34(6).* ACM SIGGRAPH Asia Conference. Volume 32.

Garcia, R., Martin-Gutierrez, J., Mendoza, S.M, and Marante J.R. (2015) Open Data Motion Capture: MOCAP-ULL Database, *Procedia Computer Science 75:316-326.* doi: 10.1016/.procs.2015.12.253.

Pullen, K. and Bregler, C. (2002) Motion Capture Assisted Animation: Texturing and Synthesis, *Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques.* 501-508.

Williams, R. (2012) *The Animator's Survival Kit: A Manual of Methods, Principles and Formulas for Classical, Computer, Games, Stop Motion and Internet Animators.* Macmillan.

Roberts R. (2018). Converting Motion Capture into Editable Keyframe Animation Fast, Optimal, and Generic Keyframe Selection.

Du, Y., Wong, Y., Liu, Y., Han, F., Gui, Y., Wang, Z., Kankanhalli, M, and Geng, W. (2016) Marker-Les 3D Human Motion Capture with Monocular Image Sequence and Height Maps. *European Conference on Computer Vision ECCV 2016.* PP 20-36.

Howe N. (1990). Silhouette Lookup for Monocular 3D Pose Tracking.

Sidenbladh, H., Black, M.J. and Fleet, D.J. (2000) Stohastic Tracking of 3D Human Figures Using 2D Image Motion. *European Conference on Computer Vision.* 702-718.

Rebinth, A. and Kumar S, M. (2019) Importance of Manual Image Annotation Tools and Free Dataset for Medical Reasearch. *Journal of Advanced Research in Dynamical and Control System.* 1880-1885.

Eichner, Marcin, Jimenez, M., Zisserman, Andrew, Ferrari, and Vittorio. (2012) 2D Articulated Human Pose Estimation and Retrieval in (Almost) Unconstrained Still Images. *International Journal of Computer Vision 99(2).*

Protopapadakis, E., Grammatikopoulou, A., Doulamis, A. and Nikos, G.. (2021) Folk Dance Pattern Over Depth images Acquired Via Kinect Sensors. *International Archieves of the Photogrammetry, Remote Sensing and Spatial Information Sciences.* Vol: XLII-2/W3.

Rebinth, A. and Kumar S, M. (2019) Importance of Manual Image Annotation Tools and Free Dataset for Medical Reasearch. *Journal of Advanced Research in Dynamical and Control System.* 1880-1885.

Chalodhorn, R., MacDorman, K.F., Ayasha, M. (2005) An Algorithm That Recognizes and Reproduces Dinstinct Types of Humanoid Motion Based on Periodically-Constrained Nonlinear PCA. *Springer.* Vol. 3276, PP 370-380.

Jolliffe, Lan, T., and Cadima J. (2016) Principle Component Analysis: A Review and Recents Developments. *Philospphical Transactions of The Royal Society A: Mathematical, Physical and Engineering Sciences.* Vol. 374, Issues 2065.

Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M., Moore, R., Kipman, A. and Blake, A. (2011) 'Real-time human pose recognition in parts from single depth image', *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition, ser. CVPR, pp. 1297-1304, ISBN: 978-1-4577-0394-2.*

D. Hogg. (1983) 'Model-based vision: a program to see a walking person', *Image and Vision computing*, vol 1, no.1, pp. 5-20.

Wei, S.E., Ramakrishna.,Kanade. (2016) 'Convolutional Pose Machine', *proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 4724-4732.

Burenius, M., Sullivan, J., and Carlsson, S. (2014) '3d pictorial structures for multiple view articulated pose estimation', in *CVPR 2013.*

Dantone, M., Gall, J., Leistner, C., Van Gool, L. (2013) Human Pose Estimation From Still Images Using Body Parts Dependent Joint Regressors, *CVPR.*

Ionescu, C., Li, F., Sminchisescu, C. (2011) Latent Structured Models for Human Pose Estimation. *ICCV* PP. 2220 – 2227.

Tekin, B., Marquez-Neila, P., Salzmann, M., Fua, P. (2017) Learning to Fuse 2D and 3D Image Cues for Monocular Body Pose Estimation. *Proceedings of the IEE International Conference on Computer Vision (ICCV).* PP. 3941 – 3950.

Zheng, Yufeng, Yang, Clifford, Merkulov and Aleksey. (2018) Breast Cancer Screening Using Convolutional Neural Network and Follow up Digital Mammography. doi:10.1117/12.2304564.

Uijlings., J.R.R., Van de Sande, K.E.A., Geavers., T.,& Smeulders (2013) 'Selective Search for Object Recognition, International Journal of Computer Vision', *Image and Vision computing*, vol 1, no.1, pp. 5-20.

He, K., Zhang, X., Ren, S., Sun, J. (2015) Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition, *IEEE Transaction on Pattern*

Analysis and Machine Intelligence 37(9). Doi: 10.1109/TPAMI.2015.2389824.

Ren., S.He., k.Girshick and Sun. J. (2017) 'Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks', *IEEE Trans Pattern Anal Mach Intel, 39(6), 1137-1149*.

Lin, T-Y., Dollar´, P., Girshick, R., He, K., Hariharan, B. and Belongie, S. (2017) Feature Pyramid Networks for Object Detection, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. PP. 2117-2125.

Lucas, B. and Kanade, T. (1981) An Iterative Image Registeration Technique With an Application to Stereo Vision, *Proc. Seventh International Joint Conference on Artificial Intelligence*. PP. 674-679.

Newell, A., Yang, K. and Deng, J. (2016) Stacked Hourglass Networks for Human Pose Estimation, *Computer Vision and Pattern Recognition (cs.CV)*.

Carreira., Fragkiadaki., Agrawal and Malik (2016) 'Human Pose Estimation With Iterative Error Feedback', *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*.

Kanazawa, A., Black, M.J., Jacobs, D.W and Malik, J. (2018) End-to-end Recovery of Human Shape and Pose, *Computer Vision and Pattern Recognition (cs.CV)*.

Running, S.A. (2017) An Evaluation of Human Pose Estimation Using a Deep Convolutional Neural Network.

Bulat, A., and Tzimiropolous, G.. (2016) Human Pose Estimation via Convolutional Part Heatmap Regression, *European Conference on Computer Vision*. PP. 717-732.

Chai, J. and Hodgins, J.K. (2005). Performance Animation From Low-Dimensional Control Signals, *ACM Transactions on Graphics*. 24(3), 686-696. ISSN 07300301. doi:10.1145/1073204.1073248.

Wei, X. and Chai, J. (2010). VideoMocap: Modelling Pysically realistic Human Motion from Monocular Video Sequences, *ACM SIGGRAPH 2010 Papers, SIGGRAPH 2010*. 24(3), 1(212). 1-10. ISSN 07300301. doi:10.1145/1778765.1778779.

Xu, W., Chatterjee, A., Zollhöfer, M., Rhodin, H., Mehta, D., Seidel, H.P. and Theobalt, C. (2018). MonoPerfCap: Human Performance Capture From

Monocular Video, *ACM Transactions on Graphics*. 37(2), 1-15. ISSN 15577368. doi:10.1145/3181973.

Sinaga, A.M.H.P. (2014). Short Essay: Difference Between Qualitative And Quantitative Analysis and How It Should Be Applied In Our Research, *Academia*. (March), 1-7, Retrievable at www.academia.edu.

Van der Struijk, S., Mirzaei, M.S., Huang, H.H. and Nishida. (2018). Facsvatar: An Open Source Modular Framework for Real-Time FACS based Facial Animation. *Proceedings of the 18th International Conference on Intelligent Virtual Agents, IVA 2018*. 159-164. doi: 10.1145/3267851.3267918.

LIST OF PUBLICATIONS

Indexed conference proceedings

1. **Azmi, N. M. W**., Albakri, I F., Suaib, N.M., Rahim, M. S. S., Yu, H. (2020). 3D Motion and Skeleton Construction from Monocular Video. *In computational Science and Technology (pp. 75-84).* Springer, Singapore.
2. **Albakri, I. F.**, Wafiy, N., Suaib, N. M., Rahim, M. S. M., Yu H. (2020). 3D Keyframe Motion Extraction from Zapin Traditional Dance Videos. *In computational Science and Technology (pp. 65-74).* Springer, Singapore.