HUMAN POSE ESTIMATION IN IMAGE SEQUENCES

TOH JUN HAU

UNIVERSITI TEKNOLOGI MALAYSIA

HUMAN POSE ESTIMATION IN IMAGE SEQUENCES

TOH JUN HAU

A project report submitted in partial fulfilment of the
requirements for the award of the degree of
Master of Engineering (Computer and Microelectronic Systems)

Faculty of Electrical Engineering
Universiti Teknologi Malaysia

JUNE 2018

I dedicate this work to my family

My circle of sanity

ACKNOWLEDGEMENT

I want to thank Universiti Teknologi Malaysia (UTM) for providing me knowledge throughout this 2 year of part time master study. UTM has provided many useful courses such as VLSI circuits and design.

I want to thank my supervisor, Assoc. Prof. Dr Syed Abdul Rahman and those who had helped me out in this project. He always patience and guide me to the right path whenever I met some problem. The completion of this project would be difficult without his support.

Besides, I want to thank to my colleagues especially Joe, Chiang Hong Seng and my manager, Chua Boon Haw. They constantly provide support on my work and allow me having sufficient of time to complete this project.

Lastly, I want to thank my parent and family members. Their moral support and encouragement had led me to the succeed of this project.

ABSTRACT

Human action recognition (HAR) has been a popular research topic and received a huge attention for several decades due to its wide range of applications such as security and surveillance, human computer interaction, health care and video indexing. However, most research focus on either video or image sequence but very few work is done on still images. The process of estimating pose configuration in a still image is called as human pose estimation (HPE). One of the problems dealing with still image for human action recognition is that there exist many articulated human points which are difficult to be captured within a single image. Moreover, more often than not, the ability to obtain the posture adds as an extra cue to the contextual information for recognizing human action. Furthermore, excessive background elements are unnecessary and often contribute to false detection of pose estimation algorithm. The objective of this project is firstly to design an effective model in estimating human pose or structure in still images by showing skeleton line of different size depicting different parts of the human body. In order to analyze posture in still image, the low resolution video is separated into several frames and each frame is enhanced by subtracting the background for accurate detection. Then, the frame is parsed into pose estimation algorithm to capture the human structure. From the result of performance evaluation, background subtraction successfully increases the true positive rate (TPR) but not the precision. On the other hand, the introduction of region of interest (ROI) successfully increases the accuracy of HPE detection by 2.16 % in the positive rate and 16.46 % in the negative rate for proposed evaluation when threshold is equal to 25. However, the TPR of ROI enhancement (88.84 %) shows slightly lower than the original algorithm (93.39 %) due to certain frames that were unable to be detected. As a conclusion, the proposed method performed at least as good as those of the state-of-art methods in estimating the human post and subsequently in classifying the human actions.

ABSTRAK

Pengecaman aksi manusia (HAR) merupakan topik penyelidikan yang popular dan mendapat perhatian yang besar sejak beberapa dekad ini disebabkan oleh pelbagai aplikasi seperti keselamatan dan pengawasan, interaksi manusia komputer, penjagaan kesihatan dan pengindeksan video. Namun, kebanyakan penyelidikan memberikan perhatian pada video atau jujukan imej, hanya sedikit kerja sahaja yang dilakukan pada imej pegun. Proses menganggarkan konfigurasi pose dalam imej pegun dipanggil sebagai penganggaran postur manusia (HPE). Salah satu masalah yang berurusan dengan imej pegun untuk mendapatkan pengecaman aksi manusia adalah bahawa wujud banyak articulated titik manusia yang sukar ditangkap dalam satu imej. Selain itu, lebih sering daripada tidak, keupayaan untuk mendapat postur yang bertindak sebagai petunjuk tambahan maklumat kontekstual pengecaman aksi manusia. Selain itu, elemen latar belakang yang berlebihan tidak diperlukan dan sering menyumbang kepada kesalahan pengesanan algoritma pose anggaran. Objektif projek ini adalah mereka model yang berkesan dalam penganggaran postur manusia atau struktur dalam imej-imej pegun dengan menunjukkan garis rangka saiz yang berbeza yang menggambarkan bahagian-bahagian tubuh manusia yang berlainan. Untuk menganalisis postur dalam imej-imej pegun, video resolusi rendah dipisahkan kepada beberapa bingkai dan setiap bingkai dipertingkatkan dengan mengurangkan latar belakang untuk pengesanan yang tepat. Kemudian, bingkai tersebut dihuraikan dalam algoritma pose anggaran untuk menangkap struktur manusia. Dari hasil penilaian prestasi, penolakan latar belakang berjaya meningkatkan kadar positif sebenar tetapi kekurangan ketepatan. Di sisi lain, pengenalan rantau kepentingan (ROI) berjaya meningkatkan ketepatan pengesanan HPE sebanyak 2.16% pada kadar positif dan 16.46% dalam kadar negatif untuk penilaian yang dicadangkan apabila ambang bersamaan dengan 25. Walau bagaimanapun, TPR bagi ROI (88.84%) menunjukkan sedikit lebih rendah daripada algoritma asal (93.39%) disebabkan oleh bingkai tertentu yang tidak dapat dikesan. Sebagai kesimpulan, kaedah yang dicadangkan akan melakukan sekurang-kurangnya sama dengan kaedah terkini dalam menganggarkan posture manusia dan kemudiannya mengklasifikasikan tindakan manusia.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

BBE	-	Bounding Box Estimation
BS	-	Background Subtraction
CCTV	-	Closed-Circuit Television
FN	-	False Negative
FP	-	False Positive
HAR	-	Human Action Recognition
HOG	-	Histogram of Oriented Gradients
HPE	-	Human Pose Estimation
MISR	-	Multiple Image Super-Resolution
PAL	-	Phase Altering Line
PCC	-	Probability of Correct Coloring
PCK	-	Probability of Correct Keypoints
PR	-	Precision
ROI	-	Region of Interest
SISR	-	Single Image Super-Resolution
TN	-	True Negative
TP	-	True Positive
TPR	-	True Positive Rate
Tr	-	Threshold

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Problem Background

Due to the fast advancement of technology, Human Action Recognition (HAR) and Human Pose Estimation (HPE) had become hot research topics in image processing and computer vision communities. Various applications can benefit from these technologies, especially in detecting certain activities or as important factor in decision making. HAR focuses on classifying the action of human while HPE is used to estimate the configuration of human-parts in a still image. Most of the existing human detection is done based on video and image sequence [1]. Figure 1.1 shows the process of action classification solely depends on HAR. This type of HAR largely depends on Motion History Image, which is able to record down the motion into a single image [2].

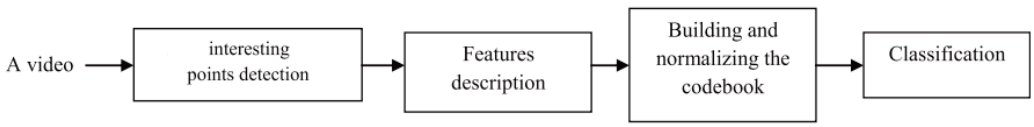

Figure 1.1: HAR without using HPE

Although HPE and HAR have different goals, some researchers prefer to use HPE as an intermediate stage for HAR [3] because Motion History Image is unable to detect very small action like waving hands repeatedly. Figure 1.2 shows the block diagram of action recognition that used HPE as an input. Furthermore, there are only a few research works being done on HPE compared to HAR. Besides, HPE is much more simple than HAR model [4].

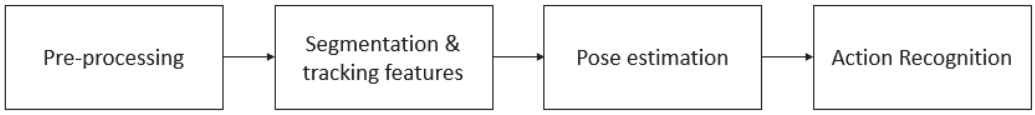

Figure 1.2: Pose estimation based HAR

1.2 Problem Statement

One of the main problems in Human Pose Estimation (HPE) is human has many articulated human joints which are difficult to be captured just within one single still image [3]. Furthermore, pose may differ for different people even though for the same action like running and sitting.

Most of the existing algorithms use high resolution still images or high occupancy images as the inputs to the HPE algorithm. However, in some situations the quality of images stored such as Closed-Circuit Television (CCTV) footage is of low resolution due to limited storage space. According to Marcin Eicher, poor quality image is one of the main factors that increases the failure rate of the HPE algorithm [5]. Furthermore, too much of background elements can cause either incorrect or poor pose estimation.

Lastly, an alternative evaluation method is required to distinguish the performance by adding the image modification as preprocessing stage. This is due to the current evaluation method proposed by existing HPE paper is using MATLAB-based tool based which makes it harder to be implemented in using C++ program.

1.3 Objectives

The proposed HPE project has the following objectives:

1. To introduce background subtraction method in existing human pose estimation method in image sequences.
2. To implement region of interest in improving the performance of human pose estimation in image sequences.
3. To propose an evaluation method which is able to produce comparable result.

1.4 Scope of Project

In this project, C++ will be used for implementing the language of HPE algorithm. All analysis is done under offline condition for in still images focusing on walking and standing. Besides, the dataset of low resolution image comes from a single frame of the half-resolution Phase Altering Line (PAL) standard video. A half-resolution PAL standard video contains characteristic of 384 x 288 pixels and 25 frames per second. Moreover, the provided video must face the subject in upright position with stationary background. Lastly, The HPE algorithm only target one people at a scene.

1.5 Organization

This chapter describes an overview of the project, problem statement, objectives and scope of this project.

Chapter 2 is the literature review related to this project. The background studies include HPE, the pixel resolution of still image and others. The comparison of related works also will be discussed at the end of this chapter.

Chapter 3 presents the methodology of this project, which includes the system framework, the system block diagram, the formula of performance evaluation and the Gantt Charts of this project.

Chapter 4 analyzes and discusses the results from findings. This chapter also compared the performance of algorithms.

Chapter 5 concludes this project and provides some recommendation for future works.

REFERENCES

1. Moussa, M. M., Hamayed, E., Fayek, M. B. and El Nemr, H. A. An enhanced method for human action recognition. *Journal of advanced research*, 2015. 6(2): 163–169.

2. Davis, J. W. Hierarchical motion history images for recognizing human motion. *Proceedings IEEE Workshop on Detection and Recognition of Events in Video*. 2001. 39–46. doi:10.1109/EVENT.2001.938864.

3. Ahad, M. A. R. *Motion history images for action recognition and understanding*. Springer Science & Business Media. 2012.

4. Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M., Moore, R., Kipman, A. and Blake, A. Real-time human pose recognition in parts from single depth images. *CVPR 2011*. 2011. ISSN 1063-6919. 1297–1304. doi: 10.1109/CVPR.2011.5995316.

5. Eichner, M. and Ferrari, V. Human Pose Co-Estimation and Applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012. 34(11): 2282–2288. ISSN 0162-8828. doi:10.1109/TPAMI.2012.85.

6. Ke, S. R., Zhu, L., Hwang, J. N., Pai, H. I., Lan, K. M. and Liao, C. P. Real-Time 3D Human Pose Estimation from Monocular View with Applications to Event Detection and Video Gaming. *2010 7th IEEE International Conference on Advanced Video and Signal Based Surveillance*. 2010. 489–496. doi: 10.1109/AVSS.2010.80.

7. Nie, B. X., Xiong, C. and Zhu, S. C. Joint action recognition and pose estimation from video. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2015. ISSN 1063-6919. 1293–1301. doi: 10.1109/CVPR.2015.7298734.

8. Zatsiorsky, V. and Prilutsky, B. *Biomechanics of skeletal muscles*. Human Kinetics. 2012.

9. Nagase, K., Katsura, S., Kasahara, Y. and Ohnishi, K. Advanced motion copying system of multi degree-of-freedom human motion. *2009 International Conference on Electrical Machines and Systems*. 2009. 1–6. doi:10.1109/

ICEMS.2009.5382785.

10. Maita, D. and Venture, G. Influence of the model's degree of freedom on human body dynamics identification. *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 2013. ISSN 1094-687X. 4609–4612. doi:10.1109/EMBC.2013.6610574.

11. Senan, M. F. E. M., Abdullah, S. N. H. S., Kharudin, W. M. and Saupi, N. A. M. CCTV quality assessment for forensics facial recognition analysis. *2017 7th International Conference on Cloud Computing, Data Science Engineering - Confluence*. 2017. 649–655. doi:10.1109/CONFLUENCE.2017.7943232.

12. Straka, M., Hauswiesner, S., Rüther, M. and Bischof, H. Skeletal Graph Based Human Pose Estimation in Real-Time. *BMVC*. 2011. 1–12.

13. Gong, W., Zhang, X., Gonzàlez, J., Sobral, A., Bouwmans, T., Tu, C. and Zahzah, E.-h. Human Pose Estimation from Monocular Images: A Comprehensive Survey. *Sensors*, 2016. 16(12): 1966.

14. Zhao, L., Gao, X., Tao, D. and Li, X. A deep structure for human pose estimation. *Signal Processing*, 2015. 108(Supplement C): 36 – 45. ISSN 0165-1684. doi:https://doi.org/10.1016/j.sigpro.2014.07.031. URL `http://www.sciencedirect.com/science/article/pii/S016516841400406X`.

15. Belagiannis, V., Amin, S., Andriluka, M., Schiele, B., Navab, N. and Ilic, S. 3D Pictorial Structures for Multiple Human Pose Estimation. *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014: 1669–1676. ISSN 1063-6919. doi:10.1109/CVPR.2014.216.

16. Yang, Y. and Ramanan, D. Articulated Human Detection with Flexible Mixtures of Parts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013. 35(12): 2878–2890. ISSN 0162-8828. doi:10.1109/TPAMI.2012.261.

17. Felzenszwalb, P. F., Girshick, R. B., McAllester, D. and Ramanan, D. Object Detection with Discriminatively Trained Part-Based Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010. 32(9): 1627–1645. ISSN 0162-8828. doi:10.1109/TPAMI.2009.167.

18. Park, S., Chang, J. Y., Jeong, H., Lee, J. H. and Park, J. Y. Accurate and Efficient 3D Human Pose Estimation Algorithm Using Single Depth Images for Pose Analysis in Golf. *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 2017. 105–113. doi:10.1109/

CVPRW.2017.19.

19. Liang, G., Lan, X., Wang, J. and Zheng, N. Human pose estimation based on human limbs. *2016 23rd International Conference on Pattern Recognition (ICPR)*. 2016. 913–918. doi:10.1109/ICPR.2016.7899752.

20. Vincent, V. What is the difference between a low and high resolution image, 2015. URL `http://valavincentphotography.com/blog/page/5/`.

21. Nejad, Y. K., Masnadi-Shirazi, M., Yazdi, M. and Shahvar, M. Z. Quality enhancement of low-resolution face images. *2015 9th Iranian Conference on Machine Vision and Image Processing (MVIP)*. 2015. 228–231. doi:10.1109/IranianMVIP.2015.7397542.

22. Chia, W. C., Yeong, L. S., Ch'ng, S. I. and Kam, Y. L. The effect of using super-resolution to improve feature extraction and registration of low resolution images in sensor networks. *2015 7th International Conference of Soft Computing and Pattern Recognition (SoCPaR)*. 2015. 340–345. doi:10.1109/SOCPAR.2015.7492770.

23. Sajjadi, M. S. M., SchÃ¶lkopf, B. and Hirsch, M. EnhanceNet: Single Image Super-Resolution Through Automated Texture Synthesis. *2017 IEEE International Conference on Computer Vision (ICCV)*. 2017. 4501–4510. doi:10.1109/ICCV.2017.481.

24. Ding, C., Bao, T., Karmoshi, S. and Zhu, M. Low-resolution face recognition via convolutional neural network. *2017 IEEE 9th International Conference on Communication Software and Networks (ICCSN)*. 2017. 1157–1161. doi:10.1109/ICCSN.2017.8230292.

25. Long, Y., Xiao, X., Shu, X. and Chen, S. Vehicle Tracking Method Using Background Subtraction and MeanShift Algorithm. *2010 International Conference on E-Product E-Service and E-Entertainment*. 2010. 1–4. doi:10.1109/ICEEE.2010.5661108.

26. Kumar, A. N. and Sureshkumar, C. Background subtraction based on threshold detection using modified K-means algorithm. *2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering*. 2013. 378–382. doi:10.1109/ICPRIME.2013.6496505.

27. Choudhury, S. K., Sa, P. K., Bakshi, S. and Majhi, B. An Evaluation of Background Subtraction for Object Detection Vis-a-Vis Mitigating Challenging Scenarios. *IEEE Access*, 2016. 4: 6133–6150. doi:10.1109/ACCESS.2016.2608847.

28. Guo, J. M., Hsia, C. H., Shih, M. H., Liu, Y. F. and Wu, J. Y. High speed multi-layer background subtraction. *2012 International Symposium on Intelligent Signal Processing and Communications Systems*. 2012. 74–79. doi:10.1109/ISPACS.2012.6473456.

29. Kim, S. and Kwon, S. Improvement of traffic sign recognition by accurate ROI refinement. *2015 15th International Conference on Control, Automation and Systems (ICCAS)*. 2015. ISSN 2093-7121. 926–928. doi:10.1109/ICCAS.2015.7364755.

30. Li, Y., Li, W. and Ma, Y. Accurate Iris Location Based on Region of Interest. *2012 International Conference on Biomedical Engineering and Biotechnology*. 2012. 704–707. doi:10.1109/iCBEB.2012.47.

31. Rujikietgumjorn, S. and Watcharapinchai, N. Real-Time HOG-based pedestrian detection in thermal images for an embedded system. *2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*. 2017. 1–6. doi:10.1109/AVSS.2017.8078561.

32. Lin, B. Z. and Lin, C. C. Pedestrian detection by fusing 3D points and color images. *2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS)*. 2016. 1–5. doi:10.1109/ICIS.2016.7550787.

33. Marciniak, T., Chmielewska, A., Weychan, R., Parzych, M. and Dabrowski, A. Influence of low resolution of images on reliability of face detection and recognition. *Multimedia Tools and Applications*, 2015. 74(12): 4329–4349.

34. Andriluka, M., Pishchulin, L., Gehler, P. and Schiele, B. 2D Human Pose Estimation: New Benchmark and State of the Art Analysis. *2014 IEEE Conference on Computer Vision and Pattern Recognition*. 2014. ISSN 1063-6919. 3686–3693. doi:10.1109/CVPR.2014.471.

35. Fisher, R., Santos-Victor, J. and Crowley, J. CAVIAR: Context aware vision using image-based active recognition, 2005.