# FEATURE EXTRACTION FOR HUMAN ACTION RECOGNITION BASED ON SALIENCY MAP

TAN YI PING

UNIVERSITI TEKNOLOGI MALAYSIA

# FEATURE EXTRACTION FOR HUMAN ACTION RECOGNITION BASED ON SALIENCY MAP

TAN YI PING

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

Faculty of Electrical Engineering
Universiti Teknologi Malaysia

JUNE 2018

*Specially dedicated*
*to my supervisor, friends and family who encouraged*
*me throughout my journey of*
*education.*

# ACKNOWLEDGEMENT

First of all, I would like to express my deepest gratitude to my supervisor, Prof. Dr. Syed Abdul Rahman bin Syed Abu Bakar for his passion, patient guidance and assistance through my research studies on postgraduate project. His broad knowledge in computer vision and image processing areas always motivated and supportive to me in completing the research work. Besides, I also gained lots knowledge from him on how to be a good researcher where he spends time on providing basic knowledge on the related studies for the research topic. His advice have always been a great and meaningful value to me. I would not able to complete this report if without his contribution on encouragement, recommendation and coordination in this thesis writing.

Next, I would like to greatest gratitude to my family especially my parents. They have constantly supporting me all the time throughout the postgraduate master engineering course and always be my side whenever I was facing bottom necks. Thank you very much for your caring and love to me. Last but not least, I also owed many thanks to my fellow friends and seniors where some discussions and conversation on their experiences on researches did inspired me on my innovation on the research.

# ABSTRACT

Human Action Recognition (HAR) plays an important role in computer vision for the interaction between human and environments which has been widely used in many applications. The focus of the research in recent years is the reliability of the feature extraction to achieve high performance with the usage of saliency map. However, this task is challenging where problems are faced during human action detection when most of  videos are taken with cluttered background scenery and increasing the difficulties to detect or recognize the human action accurately due to merging effects and different level of interest. In this project, the main objective is to design a model that utilizes feature extraction with optical flow method and edge detector. Besides, the accuracy of the saliency map generation is needed to improve with the feature extracted to recognize various human actions. For feature extraction, motion and edge features are proposed as two spatial-temporal cues that using edge detector and Motion Boundary Histogram (MBH) descriptor respectively. Both of them are able to describe the pixels with gradients and other vector components. In addition, the features extracted are implemented into saliency computation using Spectral Residual (SR) method to represent the Fourier transform of vectors to log spectrum and eliminating excessive noises with filtering and data compressing. Computation of the saliency map after obtaining the remaining salient regions are combined to form a final saliency map. Simulation result and data analysis is done with benchmark  datasets of human actions using Matlab implementation. The expectation for proposed methodology is to achieve the state-of-art result in recognizing the human actions.

# ABSTRAK

Pengenalian aksi individu memainkan peranan yang sangat penting dalam visi komputer semasa berinteraksi antara manusia dengan persekitaran dan merupakan salah satu fungsi yang boleh digunakan dalam pelbagai aplikasi dengan lingkungan yang luas. Sejak kebelakangan ini, tumpuan bagi kajian adalah kredibiliti bagi pengekstrakan ciri-ciri untuk mencapai prestasi yang cemerlang dengan penggunaan peta yang mempunyai informasi yang istimewa dan bererti. Walau bagaimanapun, tugas ini mencabar di mana masalah dihadapi semasa pengesanan tindakan manusia apabila kebanyakan video diambil dengan pemandangan latar belakang yang berantakan dan meningkatkan kesukaran untuk mengesan atau mengenali tindakan manusia secara tepat disebabkan kesan penggabungan dan tahap kepentingan yang berbeza. Dalam projek ini, objektif utama adalah untuk merekabentuk model yang menggunakan pengekstrakan ciri dengan kaedah aliran optik dan pengesan pinggir. Selain itu, ketepatan penanda peta diperlukan untuk memperbaiki ciri-ciri yang diekstrak untuk mengenali pelbagai tindakan manusia. Untuk pengekstrakan ciri, ciri gerakan dan pinggir dicadangkan sebagai dua syarat untuk ruang dan masa yang menggunakan pengesan tepi dan deskriptor bagi Histogram Sempadan Pengerakan (MBH) masing-masing. Kedua-dua cara ini dapat menerangkan piksel dengan gradien dan komponen vektor lain. Selain itu, ciri-ciri yang diekstrak dilaksanakan dalam pengiraan peta yang boleh menonjol dengan menggunakan kaedah Spektral Residual (SR) untuk mewakili transformasi vektor Fourier bagi log spektrum dan menyingkirkan kebisingan dengan penapisan dan pemampatan data. Pengiraan peta kedalaman selepas memperoleh baki daripada bahagian yang berlebihan dan digabungkan untuk membentuk peta muktamad terakhir. Hasil simulasi dan analisis data dilakukan dengan kumpulan data dengan benchmark tindakan manusia menggunakan implementasi Matlab. Jangkaan projek ini adalah memperolehi hasil yang dapat mencapai tahap yang sama atau menandingi kaedah yang sedia ada pada hari ini.

# TABLE OF CONTENTS

## LIST OF TABLES

# LIST OF FIGURES

# LIST OF ABBREVIATIONS

| | | |
|---|---|---|
| HAR | - | Human Action Recognition28 |
| MBH | - | Motion Boundary Histogram28 |
| SR | - | Spectral Residual28 |
| KLT | - | Kanade Lucas Tom29-30asi |
| SIFT | - | Scale Invariant Feature T30ransform |
| SURF | - | Speeded Up Robust feature32s |
| GLOH | - | Gradient Location and Orient33ation Histogram |
| HOG | - | Histogram of Oriented Gradients |
| HOF | - | Histogram of Optical Flow |
| FT | - | Frequency tuned |
| CA | - | Context-aware |
| DoG | - | Difference of Gaussians |
| RPCA | - | Robust Principal Component Analysis |
| CRF | - | Conditional Random Fields |
| FFT | - | Fast Fourier Transform |

# LIST OF SYMBOLS

| | | |
|---|---|---|
| 3D | - | 3-Dimensional |
| N | - | Number |
| x | - | x-axis |
| y | - | y-axis |
| t | - | Time |
| u | - | Horizontal component |
| v | - | Vertical component |
| $\varphi$ | - | Phase angle |
| $r$ | - | Magnitude |
| $\Im$ | - | Real part |
| $\Re$ | - | Imaginary part |
| $\mathcal{F}$ | - | Fourier Transform |
| $\mathcal{R}$ | - | Spectral residual |
| $g$ | - | Gaussian |
| f | - | Frame |
| $\mathcal{S}$ | - | Saliency |

# CHAPTER 1

# INTRODUCTION

## 1.1 Introduction

Human Action Recognition (HAR) is a process that recognizes the action given in images or from videos with the involvement of the local interest points [1] or regions across the time and space.  Both images and videos contain useful information that can be applied in the process to recognize the action that been captured. HAR plays significant role in computer vision and image processing societies which focusing on the interaction between human and environment. This is due to the wide spectrum of the applications such as security and surveillence, video retrieval [1], health care for the elderly and handicaps, and man-machine interface with highly commercialization potential. Based on the human action recognition system, there are some of the important characteristics that need to be clarified as follows:

i) High performance – the successful of the human action system is determined by the performance of the action recognition

ii) Region of interest – Important part of the image or video sequences that can be extracted or selected for action recognition

iii) Computation complexity – the time taken to react to the system or algorithm to recognize an action

Feature extraction is the transformation of the arbitrary input data such as image and text into sets of features which are pattern properties that contributing in categorization application [2]. The variety of the feature extraction at both low and high level [3] helps in recognizing the action by using different cues where fusion or combination are allowed to achieve the outcome and produce a qualitative result. Apart of that, saliency map is an image representation that shows the important of a pixel to its surrounding neighbours[2]. The design of the saliency map itself is meant to converting the image representation into a state that is easier to be handled and analyzed. Each pixel in the image contains information where some of the pixels do share similar characteristics that are able to be grouped together and computed for its value. It can be translated into another way of explanation where the more the pixel is important then the higher will be its value.

## 1.2 Problem Statement

In recent years, there have been many methods proposed by the researchers to recognize the human action with salient object detection based on the feature extraction and most were successful in classifying the action. However, there are some inevitable problems faced during human action detection when the video sequences are taken or captured with cluttered background. This in turns increasing the difficulties to recognize the human action accurately. Therefore, human action system in video sequences requires reliable features or cues extraction that contains useful information for action recognition. Besides, the saliency map generated based on the features extracted needed to be accurate and attractive to human eyes to detect and recognize the human action.

## 1.3 Objectives of Project

The main objective of this study is to overcome this issue by developing an efficient saliency map that able to use the feature extracted for human action recognition. In achieving this, two specific goals are considered in this study:

i) To utilize feature extraction with Optical flow method and Sobel edge detector in generating saliency map from human action recognition videos.

ii) To improve and analyze the accuracy of the saliency map generation with Spectral Residual (SR) method.

## 1.4 Scope of the Study

The scope of study are defined in order to complete the work on time with satisfying performance. In this project, the design flow of saliency map will be displayed as follow:

i) Focusing on KTH and Weizmann dataset on offline human actions recognitions videos recorded using normal camera.

ii) The design of the saliency map generation with feature extraction for human action recognition is implemented in MATLAB framework.

Evaluation of data is carried out based on visual saliency and amount of salient points detected to determine the performance of the saliency map generated.

## 1.5 Project Report Outline

The rest of the progress is organized accordingly throughout the research work. As a kick start, Chapter 2 describes the literature review of the saliency detection for the human action recognition. In this section, different features approaches, and computation techniques applied in related works are discussed. Comparison are made here for advantages and disadvantages of each methodology. Chapter 3 describes on the proposed methodology of the project. Section for results and discussion will be explained in Chapter 4. Last but not least, Conclusion is made based on the objectives defined in the project with several recommendation and future work proposed in Chapter 5.

# REFERENCES

[1] J. Stöttinger, A. Hanbury and N. Sebe, "Sparse Color Interest Points for Image retrieval and Object Categorization", *Image Processing*, vol. 21, no. 5, pp. 2681-2692, 2012.

[2] Kumar and Bhatia, "A Detailed Review of Feature Extraction in Image Processing Systems", 4th International Conference on Advanced Computing & Communication Technologies, 2014.

[3] Y. Liu, T. Gevers and X. Li, "Color constancy by combining low-mid-high level image cues", *Computer Vision and Image Understanding*, vol. 140, pp. 1-8, 2015.

[4] A. Borji, "Boosting Bottom-up and Top-down Visual Features for Saliency Estimation," in *IEEE Conference on Computer vision and Pattern Recognition (CVPR),* Providence, RI, 2012.

[5] S. Stepanyuk, "The detailed consideration of saliency-based visual attention model," in *MEMSTECH Proceedings of VIIth International COnference*, Polyana, Ukraine, 2011.

[6] Z. S. Chen, Y. Tu and L. Wang, "An Improved Saliency Detection Algorithm Based On Itti's Model" *Technical Gazette*, vol. 21, no. 6, pp.1337-1344, 2014

[7] Sebastian Brannstrom, "Extraction, Evaluation and Selection of Motion Features for Human Activity Recognition Purposes",

[8] W. Wei, B. Liu, Z. K. Pan, Z. Wang, "A Simplified HS algorithm in optical flow estimation", in *3rd International Conference on Information Science and Control Engineering,* Qingdao, China, 2016.

[9] X. T. Zhen, "Feature Extraction and Representation for Human Action Recognition", *Emerging and Selected Topics in Circuits and Systems*, vol.3, no. 2, pp. 145-154, 2013.

[10] C. Yang, L. Zhang, H. Lu, X. Ruan and M. H. Yang, "Saliency Detection via Graph-Based manifold ranking," in *IEEE Conference on Computer Vision and Pattern Recognition*, Portland, OR, 2013.

[11] S. Li, C. Zeng, S.P. Liu, and Y.Fu, "Merging fixation for saliency detection in a multilayer graph," *Neurocomputing,* vol. 230, pp. -, 22 march 2017.

[12] T. Xi, W. Zhao, H. Wang, "Salient Object Detection with Spatiotemporal Background Priors for Video," *Image Processing,* vol. 26, no. 7, pp. 3425-3436, July 2017.

[13] H. Li, Y.Xie, B, Luo, L. Tang, B. Zeng, K. N. Ngan, F. Meng, "Using Mid-High Level Cues To Detect Salient Object" in *IEEE Conference on Multimedia and Expo (ICME),* pp. 1-6, 2014.

[14] R.M. Kumar, K. Sreekumar, "A Survey on Image Feature Descriptors", *Computer science and Information technologies*, vol. 5, no. 6, pp. 7668-7673, 2014.

[15] P. Wang,Z.Zhou, W.Liu and H. Qiao, "Salient region detection based on local and global saliency," in *IEEE International Conference on Robotics and Automation (ICRA)* , Hong Kong, 2014.

[16] J.J. Luo, "Feature Extraction and Recognition for Human Action Recognition," in, 2014.

[17] J. Uijlings, I.C. Duta, E. Sangineto and Nicu Sebe, "Video Classification with Densely Extracted HOG/HOF/MBH Features: An Evaluation of the Accuracy / Computational Efficiency Trade-off," *IJMIR,* vol. 4, no. 1, pp. 33-44, 2014.

[18] H. Wang, A. Klaser, C. Schmid, and C. L. Liu, "Dense trajectories and motion boundary descriptors for action recognition," *International Journal of Computer Vision,* vol. 103, no. 1, pp. 60-79, 2013.

[19] R. Achanta, S. Hemami, F. Estrada and S. Susstrunk, "Frequency-tuned salient region detection," in *IEEE Conference on Computer Vision and Pattern Recognition*, Miami, FL, June 2009.

[20] X. Sun, Z. Shu, X. Liu, Y. Shang and Q. Yu, "Frequency-spatial domain based salient region detection," *Optics,* vol. 126, no. 9-10, pp. 942-949, 2015.

[21] A. H. Shabani, D. Clausi and J. S. Zelek, "Improved Spatio-temporal Salient Feature Detection for Action Recognition," in *Proceedings of the British*

*Machine Vision Conference*, 2011.

[22] X. Yan and X. Liu, "The improved two-dimensional Gabor filter based interest objects detection," in *8th International Congress on Image and Signal Processing (CISP),* Shenyang, China, 14-16 Oct 2015.

[23] Y. Xue, X. Guo, and X. Cao, "MOTION SALIENCY DETECTION USING LOW-RANK AND SPARSE DECOMPOSITION," in *IEEE Conference on Acoustics, Speech and Signal Processing (ICASSP),* Kyoto, Japan, 2012.

[24] M. M. Cheng, N. J. Mitra, X. Huang, P. H. S. Torr, and S. M. Hu, "Global Contrast Based Salient Region Detection," *IEEE Transactions on Pattern Analysis and machine Intelligence,* vol. 37, no. 3, pp. 569-582, March 2015.

[25] P. Wang, Z. Zhou, W. Liu, and H. Qiao, "Salient region detection based on local and global saliency," in *IEEE International Conference on Robotics and Automation (ICRA)* , Hong Kong, 2014.

[26] J. Yang, Y. Wang, G. Wang, and M. Li, "Salient object detection based on global multi-scale superpixel contrast," *IET Computer Vision,* vol. 11, no. 8, pp. 710-716, 2017.

[27] L. Xu, L. Zeng, and H. Duan, "An effective vector model for global-contrast-based saliency detection," *J.Vis.Commu.Image R.,* vol. 30, pp. 64-74, 2015.

[28] S. Goferman and L. Z. Manor, "Context-Aware Saliency detection," *IEEE Transactions on Pattern recognition and Machine Inteliigence* , vol. 34, no. 10, pp. 1915-1926, 2012.

[29] Sourya Roy and Pabitra Mitra, "Visual saliency detection: a Kalman filter based approach," *Computer Science - Computer Vision and Pattern Recognition,* pp. 1-12, 2016.

[30] Sikha O K, Sachin Kumar S, K.P. Soman, "Salient region detection and Segmentation in Images using Dynamic Mode Decomposition," *Computer Vision and Pattern Recognition,* pp. 1-8, 2016.

[31] W. Qiu, X. Gao, and B. Han, "A superpixel-based CRF saliency detection approach," *Neurocomputing,* vol. 244, pp. 19-32, 2017.

[32] J. Zhao, Y. Zhong, H. Shu, and L. Zhang, "High-Resolution Image Classification Integrating," *IEEE Transactions on Image Processing,* vol. 25, no. 9, pp. 4033-4045, 2016.

[33] Z. Liu, X. Zhang, S. Luo, O. L. Meur, "Superpixel-based saliency detection," in *14th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS),* Paris, 2013.

[34] Z. Liu, X. Zhang, S. Luo, and O. l. Meur, "Superpixel-Based Spatiotemporal Saliency Detection," *IEE Transactions on Circuits and Systems for Video Technology*, vol. 24, no. 9, pp. 1522-1540, 2014.

[35] X. Hou and L. Zhang, "Saliency Detection: A Spectral Residual Approach", in *IEEE Conference on Computer Vision and Pattern Recognition*, no.1033, pp. 1-8, 2007.

[36] C. C. Loy, T Xiang, and S. Gong, "Salient Motion Detection in Crowded Scenes", in *Proceeding of the 5th Symposium on Communications, Control and Signal Processing*, Rome, Italy, 2012.

[37] A. S. Aguado and M. S. Nixon, "Chapter 4: low-level feature extraction (including edge detection)," in *Feature Extraction & Image Processing for Computer Vision*, Elsevier Ltd, 2013, pp. 137-216.

[38] S. S. Sengar and S. Mukhopadhyay, "Motion detection block based bi-directional optical flow method", *Visual Communication and Image Representation*, vol. 49, pp. 89-103, 2017.

[39] J. Zhang and S. Sclaroff, "Exploiting Surroundedness for Saliency Detection: A Boolean Map Approach," *IEEE Transactions on Pattern Recognition and Machine Intelligence,* vol. 38, no. 5, pp. 889-902, 2015.

[40] J. Zhang and S. Sclaroff, "Saliency Detection: A Boolean Map Approach" in *IEEE International Conference on Computer vision (ICCV),* Sydney, 2013.