HUMAN ACTION RECOGNITION USING SLOW FEATURE ANALYSIS

BARDIA YOUSEFI

RESEARCH REPORT SUBMITTED IN PARTIAL FULFILLMENT FOR THE DEGREE OF DOCTOR OF PHILOSOPHY (ARTIFICIAL INTELLIGENCE)

FACULTY OF COMPUTER SCIENCE & INFORMATION TECHNOLOGY
UNIVERSITY OF MALAYA
KUALA LUMPUR

2016

University of Malaya

UNIVERSITY OF MALAY ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: **BARDIA YOUSEFI**
Registration/Matric. No: **WHA11069**

Name of Degree: **DOCTOR OF PHILOSOPHY** Title of Project Paper/Research Report/Dissertation/Thesis ("this Work"):

HUMAN ACTION RECOGNITION USING SLOW FEATURE ANALYSIS

Field of study: **ARTIFICIAL INTELLIGENCE**

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;

(2) This work is original;

(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the work and its authorship have been acknowledged in this work;

(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;

(5) I hereby assign all and every rights in the copyright to this work to the University of Malaya ("UM"),who henceforth shall be owner of the copyright in this work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;

(6) I am fully aware that if in the course of making this work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date

Subscribed and solemnly declared before,

Witness's Signature Date

Name:

Designation:

University of Malaya

Abstract

Studies on computational neuroscience through functional magnetic resonance imaging and following human visual systems state that the mammalian brain pursues two distinct pathways in the model. These pathways are designed to analyze not only motion information (optical flow) but also the ventral processing stream in the brain that proceeds with form features, in which Gabor wavelet is widely used. The original model of the mammalian visual system represents two independent pathways, which become a subject of interest among researchers. Model development is performed via systematic organization, where the active basis model is added into the ventral processing stream. The Gabor wavelet-based and supervised method is efficient in terms of Gabor beam utilization and object recognition-directed task through form pathway. In addition, the motion information that is generated via optical flow in motion pathway is stabilized through applying the fuzzy membership scoring, which delays the changes in optical flow outcomes and provides further robustness to the system. The interaction between these processing pathways is another substantial matter implied in the model. The cross-connection of the two pathways is implied throughout the present research via direct consideration, such as shared sketch algorithm and optical flow information, fuzzy max-product involvement, and scoring among each other. In addition, the model is considered a form information through active basis model based on incremental slow feature analysis (denoted as slow features). In this study, the motion perception in human visual system comprises fast and slow feature interactions, which render biological movement understandable. Primarily, a form feature is defined. This feature biologically follows the visual system through applying active basis model and incremental slow feature analysis for extraction of the slowest form features of human object for ventral stream. The interaction is considered within the time that provides valuable features to recognize biological movements. Incremental slow feature analysis provides a chance for fast action prototypes and bag-of-word techniques, and opens a new perspective to recognize the original biological movement model. Episodic observation is required to extract the slowest features. However, fast

University of Malaya

features of dorsal processing pathway through episodic ventral analysis update the processing of motion information. Experimental results in the development of the biological movement model indicate promising accuracies for proposed improvements and favorable performance on different datasets (KTH and Weizmann). The results also provide promising direction on this area.

University of Malaya

Abstrak

Kajian ke atas pengiraan neurosains melalui fungsian Pengimejan Resonans Magnetik (fMRI) dan mengikuti sistem visual manusia menyatakan bahawa otak mamalia mengejar dua laluan yang berbeza dalam model. Ia khusus untuk analisis maklumat gerakan (aliran optik) yang terdapat pada satu lagi laluan mengenai aliran pemprosesan berkenaan dengan perut di dalam otak yang meneruskan dengan ciri-ciri bentuk, dimana ombak-kecil Gabor telah digunakan secara meluas. Model asal sistem visual mamalia mewakili dua laluan yang bebas dan ia telah dikaji oleh ramai penyelidik lain, dengan beberapa perkembangan telah dilakukan. Perkembangan model yang dilakukan dengan cara yang teratur dan sis- tematik dengan menambah model asas aktif dalam aliran pemprosesan berkaitan dengan perut. Ia mempunyai dua sebab utama disebabkan ombak-kecil Gabor dan kaedah yang diselia yang boleh menjadi lebih cekap dari segi pengunaan dan pengiktirafan objek ditugaskan melalui bentuk laluan. Selain itu, maklumat gerakan yang telah dihasilkan oleh aliran optik dalam gerakan laluan telah lebih stabil dengan menggunakan keadah pemarkahan keahlian kabur yang ditangguhkan perubahan hasil aliran optik, disamping memberikan lebih kemantapan kepada sistem. Interaksi antara laluan pemprosesan juga merupakan perkara besar yang akan tersirat di dalam model. Sambungan antara dua laluan telah tersirat dalam kajian ini dengan pertimbangan langsung, seperti perkongsian lakarkan algoritma dan maklumat optik-aliran, penglibatan kabur max-produk, dan pemarkahan dari satu kepada yang lain. Di samping itu, model yang dianggap mempunyai ciri-ciri yang cepat dan membentuk maklumat melalui model asas aktif berdasarkan peningkatan analisis ciri perlahan. Di sini, persepsi gerakan dalam sistem visual manusia terdiri daripada cepat dan lambat interaksi ciri yang membuat pergerakan pemahaman biologi diambil kira. Secara prinsipnya, satu bentuk ciri-ciri biologi mengikut sistem visual memohon model asas aktif ke dalam analisis ciri tambahan perlahan untuk pengekstrakan ciri-ciri bentuk paling perlahan objek manusia untuk aliran berkenaan dengan perut. Interaksinya juga menganggap dalam siri masa yang memberikan ciri-ciri yang berharga bagi pengiktirafan pergerakan biologi. Tambahan analisis ciri lambat menyediakan peluang

University of Malaya

untuk jalan pintas melalui prototaip tindakan, teknik beg perkataan, dan membuka perspektif baru untuk melihat pengiktirafan model asal pergerakan model. Walau bagaimanapun, untuk mendapatkan ciri-ciri yang paling perlahan pemerhatian episod diperlukan, tetapi ciri-ciri cepat dorsal pemprosesan laluan sepanjang analisis episodically tentang maklumat terbaru berkenaan dengan pemprosesan maklumat gerakan perut. Keputusan eksperimen dalam hierarki pembangunan model pergerakan biologi menunjukkan ketepatan yang menjanjikan untuk penambahbaikan yang dicadangkan dalam model dan prestasi yang baik di set data yang berbeza (KTH dan Weizmann).

University of Malaya

Acknowledgments

I would like to express my sincere gratitude to Prof. Dr. Loo Chu Kiong and many thanks for his guidance, supports, his patience, and tolerance. He has driven me forward on my research and was always there for me. I am also thankful to Dr. Manjeevan Singh Seera for his constructive comments.

I appreciate all the efforts and considerations in Faculty of Computer Science and Information Technology, University of Malaya and their respected professors, management office and staffs for their kind support throughout the steps of this project. I cherish all the moments of my research study in University of Malaya and consider it as a very pleasant memory of my life.

Last but not least, I would like to thank my family and friends especially my lab-mates for putting up with me and constant support, patience and continuous encouragements.

This work was funded by Flagship grant and High Impact Research (HIR) Centre in University of Malaya.

University of Malaya

TABLE OF CONTENTS

University of Malaya

University of Malaya

University of Malaya

University of Malaya

LIST OF FIGURES

University of Malaya

University of Malaya

University of Malaya

University of Malaya

University of Malaya

University of Malaya

University of Malaya

University of Malaya

University of Malaya

University of Malaya

University of Malaya

University of Malaya

University of Malaya

LIST OF TABLES

University of Malaya

List of Abbreviations

1D:	Dimension.
2D:	Two Dimension.
2AFC:	Two-alternative forced choice.
3D:	Three Dimension.
ABM:	Active Basis Model.
AC:	auditory cortex.
AI:	Artificial Intelligence.
ANN:	Artificial Neural Network.
ASCs:	autism spectrum conditions.
AUs:	action units.
BSFA:	Batch SFA.
BIM-STIP:	bio-inspired model based spatiotemporal interesting points.
BLOD:	Blood-oxygen-level-dependent.
CVt:	Centroidal Voronoi tessellation.
CCIPCA:	Candid Covariance-free Incremental Principal Components Analysis.
CIMCA:	covariance-free incremental minor components analysis.
CLLC:	context and locality constrained linear coding.
CNN:	Convolutional neural networks.
ConvNet:	Convolutional neural networks.
D-SFA:	Discriminative-Slow Feature Analysis.
DR:	Who developed visual agnosia (*i.e.*, damage to the ventrolateral occipital).
DSN:	directional selective neurons.
EBA:	Extrastriate body area.
ELM:	Extreme Learning Machine.
ELM-Sig:	Sigmoid kernelled Extreme Learning Machine.
ELM-RBF:	Extreme Learning Machine(Radial Basis Function).

University of Malaya

ELM-Wav:	Extreme Learning Machine- Wavelet kerneled.
fMRI:	functional Magnetic Resonance Imaging.
FA:	Area oF STS and IT.
FBA:	Fusiform Body Area.
FFA:	fusiform face area.
FS:	Area oF STS and IT.
GSRC:	group-wise sparse representation based classification.
GTS:	Gilles de la Tourette syndrome area.
HAMMER:	Hierarchical Attentive Multiple Models for Execution and Recognition.
HMM:	Hidden Markov Model.
HOR:	Histogram-of-Oriented-Rectangles.
IDM:	integrated dynamic motion model.
IEKF:	iteration expanded of Kalman filter.
IncSFA:	Incremental Slow Feature Analysis.
IPSmot:	Intra-Parietal Sulcus motion.
IT:	Inferotemporal Cortex.
JDG:	Ju-Du-Gunzburger.
KNN:	K-Nearest Neighborhood .
KO:	Kinetic Occipital area.
KTH:	Human Action Dataset including four different scenarios.
LGMD:	Lobula giant movement detectors.
LM:	lemniscus medialis.
LOC:	Lateral Occipital Complex.
MAX.SUM:	Maximum Summation.
MCA:	Minor Component Analysis.
MEG/EEG:	magneto- and electroencephalography.
MLD:	Moving Light Display.
MPOD:	Most Probable Optimum Design.
MST:	Medial Superior Temporal.

University of Malaya

MSTd: Medial superior temporal (dorsal).

MSTl: Medial superior temporal (lateral).

MT: Medial Temporal.

OF: Optic Flow.

OFA: occipital activation.

OPA: Operator Perceived Activity.

PCA: Principle Component Analysis.

pLSA: probabilistic Latent Semantic Analysis.

PGa: area PGA of temporal cortex.

PPA: parahippocampal place area.

PSO: Particle Swarm Optimization.

pRF: population receptive field.

QPSO: Quantum Particle Swarm Optimization.

RBF: Radial Basis Function.

RF: Receptive Field.

RSC: retrosplenial cortex.

RV: Who developed optic ataxia (*i.e.*, damage to the occipitoparietal cortex).

S-SFA: Supervised Slow Feature Analysis.

SD-SFA: Spatial Discriminative Slow Feature Analysis.

SFA: Slow Feature Analysis .

SII: Secondary somatosensory cortex.

SLFN: Single Hidden Layer FeedForward Neural Network.

SNN: Synergetic Neural Network.

SSA: Shared Sketch Algorithm.

STS: Superior Temporal Sulcus.

TE: Tracheoesophageal(cortical areas).

TEO: Temporo-occipital(cortical areas).

TOS: transverse occipital sulcus.

TPO: Temporo-Parieto-Occipitalis.

University of Malaya

V1:	Visual Primary Cortex.
V2:	Second area of visual extrastriate areas.
V3:	Third area of visual extrastriate areas.
V4:	Fourth portion of visual extrastriate areas.
VNMF:	translation-invariant nonnegative sparse coding.
VQ:	vector quantization.
VWFA:	visual word form area.
U-SFA:	Unsupervised Slow Feature Analysis.

University of Malaya

CHAPTER 1:

Introduction

1.1 Overview

Human action recognition in monocular video is an important subject regarding video applications, such as in human computer interaction, video search, and others. This biological recognition is researched in different fields, such as neurophysiological, psychophysical, and experimentations on imaging, and some cortical areas have been acknowledged. In general, human action recognition in video streams using video processing and such methods in the proposed area can be categorized into two techniques. One technique uses global feature extraction from video streams to allocate a particular label to the whole video. This technique clearly needs an unchanged observer within the video, and the environments where actions are occurring should be considered (Santofimia et al., 2013). The second technique considers local features in each frame and label for distinct action. Afterward, sequences can be attained through simple voting for global labeling. Temporal analysis for obtaining the features in each frame and classification is based on the observation in temporal window. Both approaches should have attained significant outcomes in such area (Schindler & Van Gool, 2008). An important factor of complex action recognition and discrimination among different human motion styles and individuals is learning (Hogg et al., 1995). Learning is also fundamental in recognizing 3D stationary human motion (Efros et al., 2003). Human action recognition using video frames can be categorized as an object recognition problem. Such recognition is supposed to handle object variations (e.g., style and size), and the human brain can excellently categorize human objects in different classes of action; recent methods have been inspired by the biological outcomes of computational neuroscience (Daugman, 1980; Olshausen, 1996). In the primary visual cortex (V1), the image procedure is particularly sensitive on bar-like structures. V1 responses are combined by extrastriate visual areas and passed to the inferotemporal cortex (IT) for recognition (Riesenhuber & Poggio, 2002).

University of Malaya

1.2 Research Problem and Problem Statement

Computer vision researchers are categorized into human action recognition in terms of computer vision-based problems, which are different from the biological model of the visual system. The obtained results of such approaches have a high accuracy. However, some of these methods are not considered any contribution in biologically inspired models. The original biologically inspired model [1] has proposed two independent pathways [2], which model the dorsal and ventral processing streams in the mammalian visual system. The form pathways, which represent the ventral streams, should utilize the Gabor filter function to obtain the shape and form information and as a good representation of simple and complex cells. Moreover, the motion pathway has been used for optical flow in the extraction of motion information. The slowness principle is not being used within the pathways and the original model. Nevertheless, slowness features are important motivations in biologically inspired area and object recognition. Proposing a supervised or unsupervised learning tool in the ventral stream can increase the robustness and recognition of human object within the streams. Here, the summary of the research problem are mentioned below for better representation:

Problem 1: Using Gabor which has no learning mechanism in the current model, it is not followed the actual brain model(even bio-inspired model).

Problem 2: The presented model in literature is still not efficient for interaction between two processing pathways.

Problem 3: Slowness principle is not used as a biological inspired model (two pathways model).

1.3 Objectives

This research aims to develop the mechanism of biological movement. It modifies a supervised learning method in the form processing stream to increase the performance

[1] We will use the term original model in this thesis which represents the model reviewed by Giese & Poggio (2003).

[2] Pathways are representing the streams in the brain visual system

with respect to the original model of recognition of biological movements. In addition, physiological models and evidence reveal some feedback toward these independent and separated pathways, which should be considered in the original model, having the interaction between the form and motion processing streams. Moreover, the utilization of slow feature analysis in the recognition of biological movement is a good parameter to represent this model in terms of slowness and fast features and their combinations within the pathways. In general, the objectives of this research summarized corresponding to the problems:

Objective 1: Improve the abilities of ventral streams in terms of detection of human body shape with ABM.

Objective 2: Improving in the interaction between the dorsal and ventral processing pathways.

Objective 3: Improving the motion information processing in dorsal processing stream with fuzzy optical flow division method which guides ventral stream.

Objective 4: Modifying the mechanism for recognition of biological movement applying slowness principle.

1.4 Hypothesis

This research focuses on the improvement of recognition through the development of the original model for biological movement recognition (the general configuration has been shown in figure 1.1). This research follows three major hypotheses, which can be summarized as follows:

• Active basis model (ABM), as a supervised object recognition method based on a model, can be utilized in the ventral processing stream in the original model;

• The use of motion information in the middle of processing for both processing pathways can be a good representation of the interaction between the ventral and dorsal streams in visual processing paths.

University of Malaya

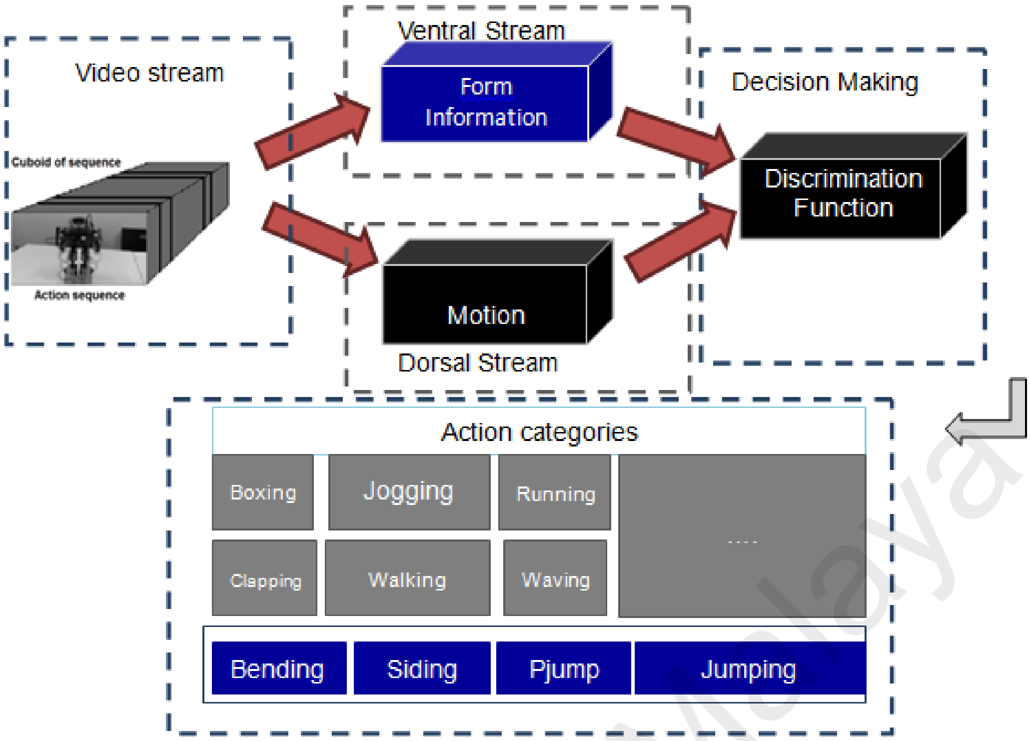

Figure 1.1: It is a general scheme diagram followed the principle of original model of biological movement mechanism.

• The fast and slow feature interaction makes the biological movement more understandable.

1.5 Scope of Study

The scope of this study is defined by the improvement in the original model for the recognition of biological movement through improving functionality using a supervised object recognition approach (i.e. Active Basis Model) for the ventral processing stream in the human visual system, and increasing robustness and relative comparable accuracy as compare with the state-of-the-art methods. In addition, modeling the ventral stream by using the slowness principle for extraction of human object form and motion information shows optical flow application. The interaction between these two pathways occurs in the categorization part, which obtains significant results in the decision and recognition of movements. However, each of these pathways can separately help to recognize biological movements by using the considerable disparity rate. Two human action recognition

University of Malaya

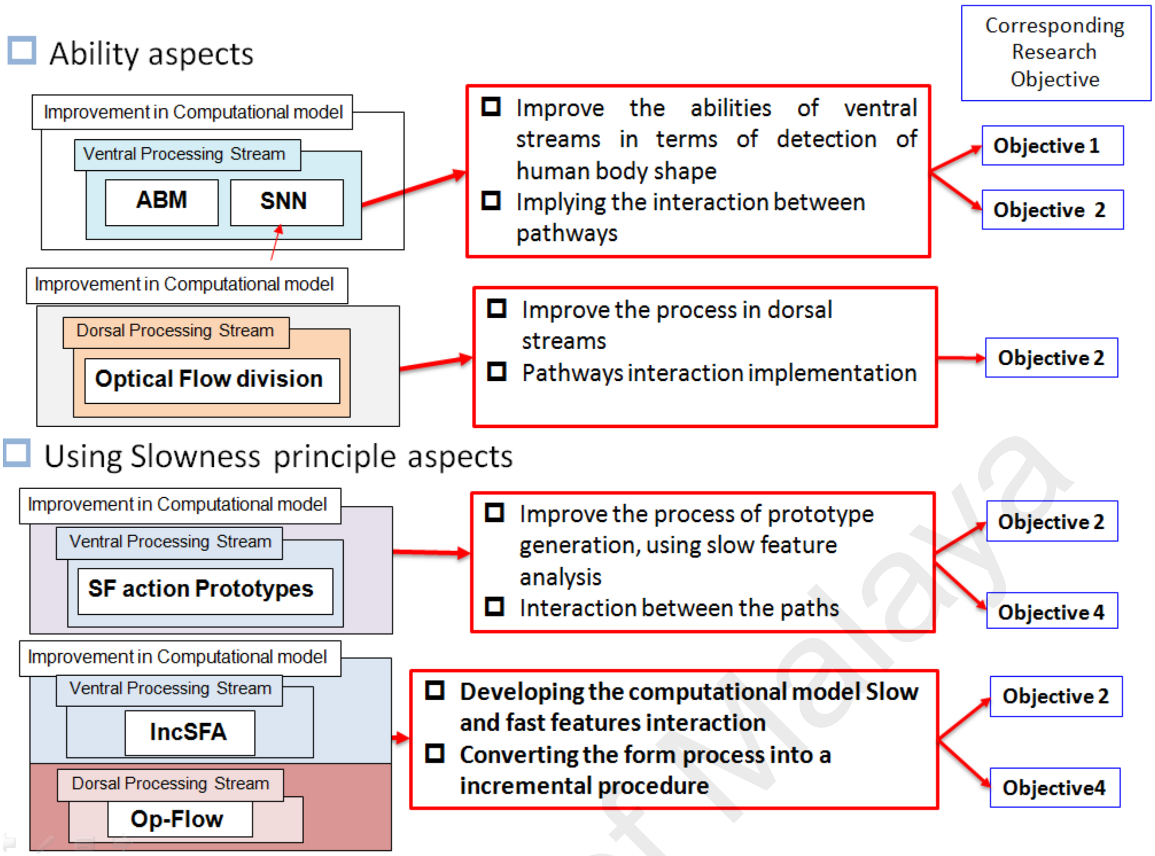

Figure 1.2: Figure represents an overview on modifications of biological movement mechanism along with it contribution and their corresponding objective.

datasets have been utilized to benchmark the system performance.

1.6 Significance of The Study

This research study contributes in computational intelligence following the evidences in the fields such as philological, neuroscience, neurophysiology, holonomic brain theory, computational and theoretical neuroscience concerning the information of the models defined for mammalian (human) visual system. Video processing applications are mentioned in this research (e.g., human action recognition, human activity recognition). Moreover, applying slowness principle [3] into the mechanism provides different perspective for two independent pathways.

[3]Slowness principle is introduced by Wiskott & Sejnowski (2002), it is reviewed in the next chapters

University of Malaya

1.7 Research contributions and summary

This chapter provides a brief introduction to the research problems, objectives, hypothesis, and research significance. Here, the contributions are pointed out. There are three different problems existed in current biologically inspired human action recognition approaches (object shape learning in ventral streams, interaction between two pathways, and slowness principle in the mechanism). The objectives corresponding to these problems are listed below along with their contributions:

Adding ABM in ventral stream

* It improves the abilities of ventral streams in terms of detection of human body shape (object recognition task improvement)(*it corresponds to the first objective*).

* It implies the interaction between pathways (*related to the second objective*), guidance of optical flow into the Share Sketch Algorithm (SSA).

Optical Flow division

* It improves the process in dorsal streams using the divisions of optical flow which helps more robust outcome (*Corresponding to objective 3*).

* The optical flow division updates the interaction between two pathways which is done in previous method(*Corresponding to objective 2*).

Slow features(SF) action Prototypes

* It improves the process of prototype generation, using slow feature analysis (*fulfills the* 4^{th} *objective*).

* It updates interaction between the pathways, through the update the SSA into the SFs instead of melting algorithm (*it responds to objective 2*).

A dual Slow and Fast features

* It develops the computational model slow and fast features interaction (*Objective 2*).

* It uses the slowness principle into the mechanism and converts the form process into a incremental procedure (Objective 4^{th}).

University of Malaya

1.8 Outline of the thesis

Chapter 1 presents the research introduction. Background and literature of research and biologically inspired models are reviewed in Chapter 2. Chapters 3 and 4 discuss the methodology and corresponding results. Concluding remarks and discussions of the developments and their contributions in the mechanism of biological movement are given in Section 5.

University of Malaya

CHAPTER 2:

Literature Review

2.1 Introduction

Analysis of biological motion recognition is categorized in different research fields, such as neurophysiology, neuropsychological, computer vision, and artificial intelligence (AI). The present article mainly investigates the AI and computer vision perceptive of this task, which considers the subdivision of a computer science field dealing with machine intelligent behavior and learning. Since the 1950s, the primary experimental methods have been classified into two main divisions: 1) symbolic methods, which follow classic approaches, are similar to expert systems and termed as connectionism approaches; 2) scruffy methods concentrate on the intelligence evolution or following artificial neural networks. Both directions undergo rigorous restrictions. The untimely objectives, such as human behavior reproduction and simulation, are entirely overlooked. Machine and man have opposite abilities: man can estimate, infer, and recognize in parallel, whereas machine can sequentially perform quick computations. Recently, biologically inspired models, which are similar to behavior-based AI, have attracted attention. These models are more focused on the performance than on the internal processing of the machine. Many projects have intensive collection of the mentioned facts, but none can create a machine that takes direct advantage of this information. Nevertheless, in the field of vision combined with AI , some outcomes are obtained, but the remaining results might not be sufficient. Modern machines can learn on the basis of statistics and follow few determined objectives. One objective is to deal with information extraction from large datasets, unsupervised learning, pattern recognition, and calculation based on statistical analysis statements following hypothesis evidence. Such a machine is practically utilized for actual problems, such as speech, image, and object recognition. Other topics, such as actual machine intelligence and principles, are also practiced to be philosophical and theoretical devices. Biological movement and its recognition comprise a multifield research that follows many biological principles and engineering approaches, and are based

University of Malaya

on the human (or mammalian) visual system. In the present study, this subject is further reviewed in terms of mechanisms and models proposed for this task.

2.2 Human action recognition and biological movement

Human action recognition is an imperative and notable part of computer vision study. Currently, such recognition is also one of the most promising applications in this area and has attracted considerable attention for various applications, such as human computer interaction and video search. The task of human action recognition can be summarized by automatically determining the type of human object action in video sequences or image frames. Recognizing complex motion is also important for distance recognition, communication, and imitation learning in complex motor actions. Motion recognition has been studied in neurophysiological, psychophysical, and imaging experiments, as well as in numerous fields. E. J. Marey and E. Muybridge carried out the initial studies on human movement in the 1850s; they photographed moving subjects and presented on its locomotion (Turaga et al., 2008). One of the earliest research on visual perception and introduction of actual movement was conducted by Rubin (1927) and Duncker (1929) (Beintema & Lappe, 2002; Hu et al., 2016). Johansson et al. (1973) presented the initial part of a study on moving object characteristic and movement perception (Leek et al. 2012). They considered some motion patterns for living being, such as humans and animals, as biological motion and few important points in the main joints of moving body, which represent motion patterns. Such elements in proximal stimulus can represent the kinetic–geometric model regarding the analysis of visual vector and basically expand in mechanical form when biological motion perception and its patterns are combined. The classic moving light display (MLD) provides an excellent impetus regarding the perception of human motion in neuroscience analysis and study (Leek et al., 2012; Johansson, 1975). Recognition of human walker's gender is performed without knowledge cues by using point-light sources established on human body's important joints, which is not similar to statistical experiments that have sufficient accuracy in this task. Changes in the speed of walking and degree of arm swing, particularly in high speeds, are associated with females, and upper body joints are excellent in determining the accuracy for gender recognition in light analysis (Kozlowski & Cutting, 1977). Marr et al. (1978) showed the problem of computational process in human visual system and information obtained via retinal images; 3D shapes are considered for problem presentation by introducing some notes as follows: three criteria are introduced for shape recognition in judging, three aspects of design representation are considered (e.g., coordinate, primary shape unit information, and information organization), shape description (e.g., coordination of the object center, size variations, modular organization, view transferring mechanisms, and identification of natural axes), and constraints regarding conservation recognition applying further information from the image. Perrett et al. (1985) reported on the temporal cortex of macaque monkey; they found that most of the cells in the brain region are sensitive to the type of movement and respond to specific body movements (Perrett et al., 1985). The two cell types introduced are sensitive to the rotation and view of body movements. Furthermore, the response of majority of the cells in the areas of temporo-parieto-occipital and PGA of temporal cortex has been considered in providing descriptions of view-centered and view-independent responses among the mentioned cells (Pribram, 1991). Goddard (1989) used connectionist techniques, along with spatial and temporal feature incorporation through diffused MLD data, and represented the walking recognition in 400 ms MLDs (Goddard, 1989). This integration occurs in the low-level features of the shape and motion by this target to make high-level features. Low-level features include sequential trajectories points, and they are grouped in line segmentation with one another to obtain proper lower and upper body limb forms. Remarkably, shape and form pathways are hierarchically joined to detect the three levels of complexity, i.e., component, segment, and assembly levels that signify temporal series on procedures (Goddard, 1989). Goddard followed the biologically inspired human action recognition in determining the complex structured motion by using MLDs. He analyzed major computational problems, such as time-varying representation, visual stimuli integration, gestalt formation, contextual formation, and particular spatial location focusing on process and its representation. Moreover, he showed the process of "what" and "where" in the visual system tightly coupled in a synergistic manner (Goddard, 1992). To better follow the different aspects

University of Malaya

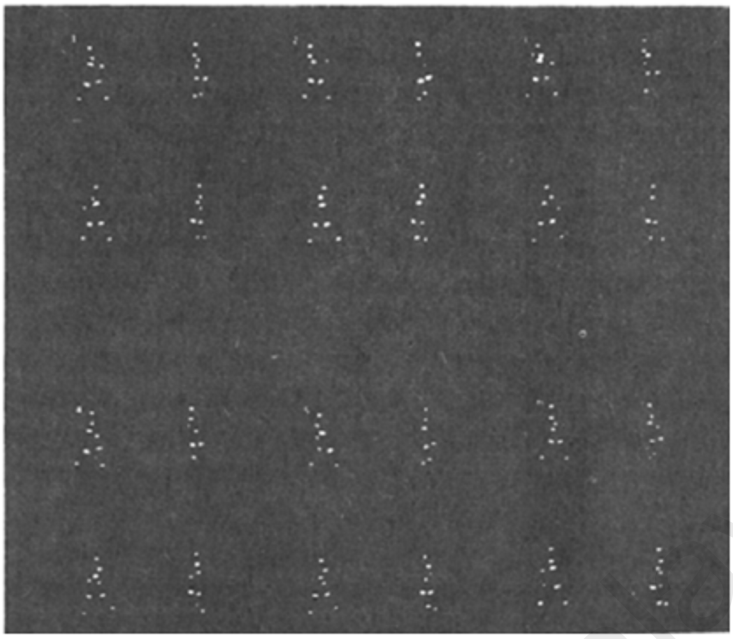

Figure 2.1: Static approximations of the dynamic stimuli for the six walkers. Adapted from (Aggarwal & Cai 1997).

of biological human action, the subject is divided into several different subsections. To summarize this review, the following sections will discuss motion perception, modeling, psychological and neuroscientific approaches, and other various aspects on the analysis gathered in the previous section.

2.3 Perception of the motion

Goodale and Milner (1992) researched on the separation of the perception and actions for recognition and identification on the ventral processing stream concerning the object recognition task. This separation allows the observer to move the hand for picking up of object and considers the projection perceptual information for object identification from striate and IT; furthermore, the posterior parietal region of the striate cortex has dorsal stream projection and needs visual sensorimotor transformations (Goodale & Milner, 1992).

Cédras and Shah (1995) reviewed motion development from the recognition aspect. Two main stages were presented for motion-based recognition by organizing it into motion models and matching unknown input with the constructed model. Several recognitions,

University of Malaya

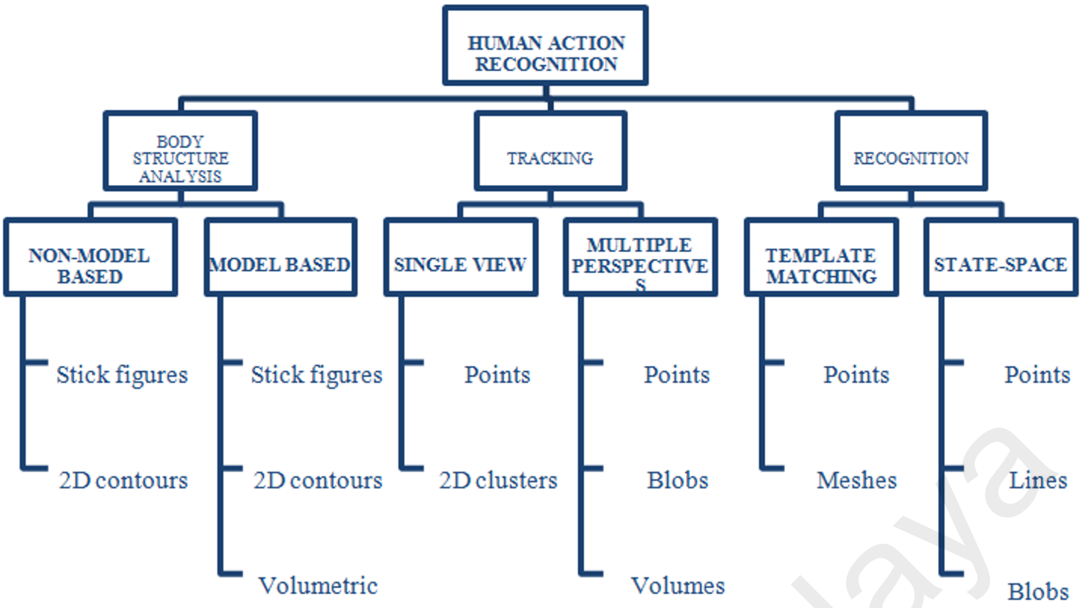

Figure 2.2: The categorization of human action recognition and biological movement has been represented in form of flowchart. Picture is adapted from (Aggarwal & Cai 1997).

such as cyclic motion detection and recognition, hand gesture interpretation, tacking, and human motion recognition, were reported. Perkins (1995) presented real-time animation, along with rhythmic and stochastic noise, for conveying only the texture of motion; this research also avoided the computational dynamic and constraint solvers (Perkins, 1995). He showed that each action has an internal rhythm and transition among movements, and realized a real-time animation. The detection of cyclic motion in frequency domain techniques of the magnitude information of Fourier transform and autocorrelation is represented as a curvature (1D signal) in function of time at 2D trajectories. Such detection is tested by synthetic and actual data of a walking person (Tsai et al., 1993). The spatiotemporal method and the hidden Markov model (HMM) technique were presented for MLD identification and classification, respectively. It provides decision based on the spatiotemporal sequence of the observed object features, and relatively little spatial information is caused by the segmentation of MLD image sequences, along with object identification; such information is highly temporal and is accessed by the HMM system, a major high classification rate (Fielding et al., 1995).

Gallese et al. (1996) analyzed the electrical activity in the brain of macaque monkeys from 532 rostral parts of six inferior neuron areas. Visual activation is required for collaboration between the agent (i.e., hand and mouth) of object and its relevant action. Finally, the mirror neurons form a system aimed at matching observation, motor action execution, and its contribution in action recognition (Gallese et al., 1996). Aggarwal et al. (1999) reviewed the human motion analysis approach that has received attention through computer vision research, which increasingly encourages a broad spectrum of applications (Aggarwal & Cai, 1997). Some tasks in the three major areas of analysis are related to interpreting human motion as human body parts, tracking body using single/multiple cameras, and recognizing human activities through frame sequences. This low-level segmentation of human body parts involves joints and projection of the 3D structure of the human body in 2D representation (Aggarwal & Cai, 1997). The problem in most conventional empty spaces and buildings lies in the difficulty of creating computer-generated characters that display real-time, engaging interaction, and realistic motion. The process of action and its perception have a common representational structure. The understanding of action observation (in the observers) in another individual has a similar neural code used to produce the same actions. Evidence of this hypothesis includes brain image studies and examination of the functional segregation light of the perceptual mechanisms subtending visual recognition and the same mechanism used for action (Decety & Grèzes, 1999).

Rangarajan et al. (1992) researched on matching the extended trajectories of biological motion (implying the points in the moving objects) with the information from object recognition. They used scale-space representation, which considers the direction and speed of the trajectories and uses spatial information to match the trajectories. The performance of these two algorithms has been tested in synthetic and actual cases (Rangarajan et al., 1992). Neri et al. (1998) presented the ability of the visual system to integrate the motion information of the standing still of walkers and such actions over time and space, as well as their capacity comparison by viewing simple translational motion. An analysis regarding the temporal differences within the recognition of biological movement involves the study of point-light display and human perception of movements.

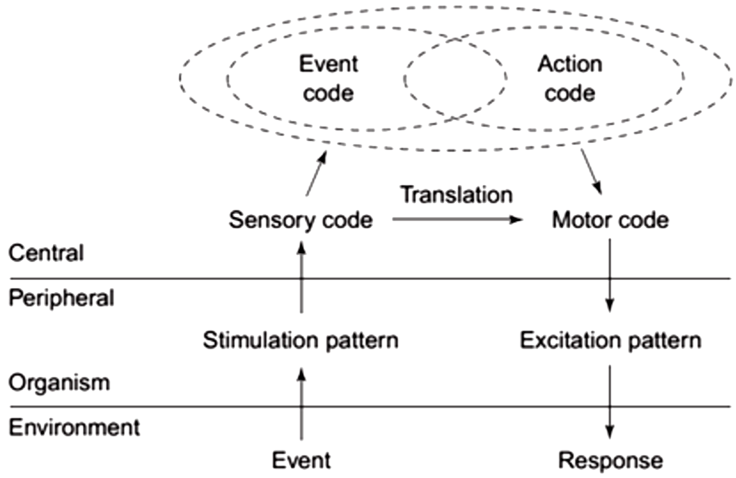

Figure 2.3: The control of action and perception showed in term of functionality in the figure. The sensory codes generated in brain (central) and simulation's patterns in sense organs (peripheral) revealed by left-hand side and upward arrows. In downward arrows in the right-hand side are excitation patterns from motor codes in muscles within the actions. (Adapted from Ref. a, by permission of Psychology Press Limited, Hove, UK, Image and caption adopted from (Decety & Grèzes 1999))

During this study, the temporal properties exposed to exaggeration techniques and spatial information are constant, which affects the duration of motion sequence. The recognition of exaggerated motions is better than the original, and the absolute duration is not a significant cue for the recognition of biological movements. Therefore, the exaggeration in temporal information returns the overall principle of diagnostic information, which is encoded for recognition in various domains (Hill & Pollick, 2000). Giese and Poggio (2000) presented morphable models by linear combination of prototypical views to recognize biological movements and image synthesis for stationary 3D objects and its involvement in complex motion patterns. The linear combination of prototypical image sequences is used to recognize action patterns (even complex movements). The mentioned new approach can be used to analyze and synthesize the biological movement, which involves the actual and simulated video data and various patterns (which has local properties of the linear vector space) of locomotion (Giese & Poggio, 2000). Moeslund and Granum (2001) conducted a comprehensive survey on motion capture involving computer vision. This survey targeted an overview on the taxonomy of system functionalities and summarized it into four processes (initialization, tracking, estimation of agent pose, and recognition); each of these processes is divided into its subprocesses and various categories (Moeslund & Granum, 2001). Moreover, Grèzes et al. (2001) analyzed human perception in biological motion. Their study considers key role for action interpretation, identification, and predication. The main hypothesis of their approach lies under neural network specifications and its verifications through fMRI for 10 healthy volunteers. Seven types of visual motion displays are used: random dot cube, drifting random dots, random dot cube with masking elements, upright point-light walker display with masking elements, inverted point-light walker display with masking elements, upright point-light walker, and inverted point-light walker. In this approach, the hemodynamic responses of both rigid and non-rigid biological motions are connected (rigid motion responses are localized posteriorly to the rigid responses). The left intraparietal cortex is involved in non-rigid biological movement perception and associated with the posterior superior temporal sulcus (STS) and left anterior portion of IPS responses. Regions, such as LOS/KO, MT/V5, and the posterior STS, are included in these activations (Grezes et al., 2001). An examination on the visual perception effects used point-light display for the arm movements of two actors for knocking and drinking movements. These actions were performed in 10 various effects. The point-light animation influenced by the phase-scrambled and upside-down versions of actions was shown to the actors for classification. The experimental results indicated that perception affects the corresponding action kinematics and movement of the phase related to the different limb segmentations (Pollick et al., 2001). A computational approach for biological movement perception presented the computer–human interface algorithm for initialization. The approach detects human body movement and automatically labels it. It also estimates the maximization of the joint probability density function of the velocity and position of the body parts (Song et al. 2001). Wiley and Hahn (1997) proposed a virtual reality approach regarding the computer-generated characters and their engaging interactions. The kinetics of human body and biological motion for the interpolation synthesis of articulated figure motion were previously studied (Wiley & Hahn, 1997). In addition, Grossman and Blake (2002) conducted visual system analysis on neural mechanisms, anatomical and functionality into two forms, and motion distinct pathways. The analysis of point-light animation involves the mutual perception of form and motion within the act of biological movement. Their study referred to a previous work regarding the activation of posterior superior temporal sulcus (STSp) and presented a new finding for the activation of fusiform (FFA) and occipital within the biological movement and generation of neural signals, which can differentiate a biological motion from a non-biological one. LOC and EBA involved in human form perception were also presented. The neural in the form and motion pathways causes the biological motion perception (Grossman & Blake, 2002). Jastorff at el. (2002) proposed an approach to investigate the recognition process during neural mechanism and whether the brain can learn a completely new complex pattern of action. They generated a new artificial–biological motion using the linear combination of time–space prototypical trajectories recorded through motion capturing. This method provides a significant improvement in discrimination for all the stimuli. The human brain can learn entire novel action patterns (Giese et al., 2002). An investigation on the spatiotemporal generalization of biological movement perception revealed the response of motion stimuli and interpolated this generalization among natural biological motion patterns. A linear combination of spatiotemporal patterns is estimated using natural movement patterns. The weight of prototypes in the morphs and the continuous and smooth variations in category probabilities are observed in this approach. A generalization exists within the motion patterns classes in the visual system (Giese & Lappe, 2002). Beintema and Lappe (2002) analyzed the perception of form pattern of human action through moving light points. The biological motion stimulus follows limited time perception of human motion. The figure of human motion is spontaneously recognized without the local image motion, and the direction of the walker and walking figure coherence are recognizable without the existence of image motion. The approach shows that image motions are not the foundation for biological motion perception, and the biological motion is derived from form information, which is dynamic (Beintem & Lappe, 2002). Kilner et al. (2003) elucidated the connection between the actions and their perception by representing the action in motor programs. The mirror system involves the execution

University of Malaya

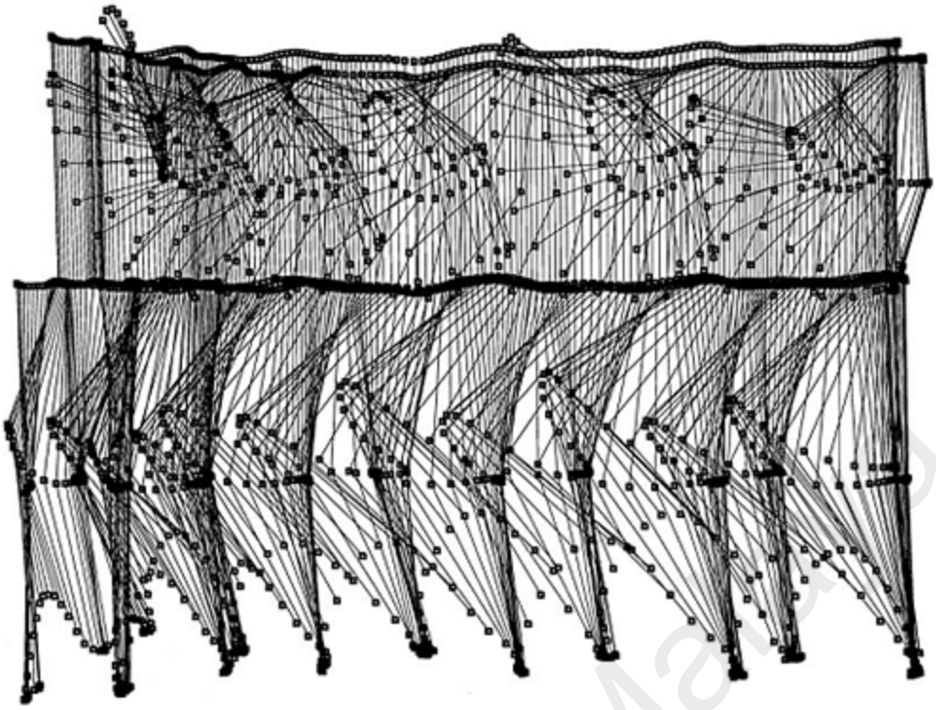

Figure 2.4: The figure represents the motion for the human skeleton during the simple movement and the trajectories of the person whom perform the actions also revealed. (picture adopted from Wachter, S., & Nagel, H. H. (1997, June) .)

of actions after considering the interference between the observation and performance. The approach hypothesizes that the overlap between observation and execution causes inconsistent performance. Observation on incongruent movement performance also elicits considerable interference effect on action execution and suggests the considerable cost to motor control (Kilner et al. 2003). An approach on perceiving the complex shape orientation and local geometric attribute perceptual integration was also presented for global representations through two-part shape adjustment and 2AFC task. Such research is not related to action recognition, but the shape analysis for visual estimation suggests that robust statistic influences segmentation (Cohen & Singh, 2006).

The ability to recognize moving human figure using moving point light is considered a biological motion perception. These point lights provide information on body shape and local motion signals to such a vivid perception. Lange et al. (2006) investigated the global form analysis by applying simple-template matching techniques through statistical pos-

University of Malaya

tures of walking human subjects with near-absent local motion signals. The simulation results were compared with psychological data and extraction of sparse form data in point-light walkers (Lange et al., 2006). An approach on the perception of object recognition analyzes the observation features of humans. The transfer of information from particular trained object components has been implied for other objects sharing the components, and this transfer depends on their geometric relationship to the objects. The shared components of objects cause a high level of recognition in the objects, but the component transfer between the objects is limited to not more than one (Gölcü & Gilbert, 2009). A research on motion blind patient (LM patient) considers the suspected human homolog of V5/MT concerning the moving stimuli. The patient reports biological motion (human action) as Johansson display (moving dots, which have particular direction and velocity). Thus, the motion in the visual cortex is interpreted traditionally and associated with motion processing. However, the patient did not report the spatial disposition of the actor and ability for figural segregation on the movement basis cue and interpretation of the movements (moving parts) independently (McLeod, 1996). Daems and Verfaillie (2010) presented the analysis of body postures in different viewpoints and human identification using four experiments. Some parts of their findings conclude that people who can identify the actions are basic-level objects and that an abstraction occurs in the visual system (Daems & Verfaillie, 1999). Peripheral vision and pattern recognition concerning the theory of form perception were summarized by Strasburger et al. (2011). Their research includes the extension of Schwartz's cortical mapping function, discussion about limitations, demonstration of Bouma's law, and an extensive range of psychophysical tasks. It also considers the low presentation complexity and speed of pattern categorization and cognitive processing for peripheral vision for low-level functions (Strasburger et al., 2011). Servos et al. (2002) researched on the relationship between biological motion and control unpredicted stimuli by examining shape perception, motion neural subtractions, and motor imagery. During this research, a biological motion called BOLD signal is found in the lingual gyrus at the cuneus border (Servos et al., 2002). A study about single cells, neuroimaging data function, and field potential records shows the visual mechanism in STS

University of Malaya

in primates and humans; it also simplifies biological motion display by using point-light markers on the limbs of walkers (Puce & Perrett 2003). Data retrieval regarding the direction from scrambled point-light displays in humans and animals leads to a hypothesis that biological motion mechanism serves for general detection system (Troje & Westhoff, 2006). A review on perception considers the human action and its advancement in motoric and visual effects, as well as the elucidation of perceptional neural concomitants (Blake & Shiffrar, 2007). A comparison of point-light view and human process in biological motion shows that non-human actions do not convey the actions in the same manner as point lights. The neural responses in pITG are also not engaged, and STSp plus FFA/FBA and ITS are decreased merely in the point-light versions. STSp is involved only in human actions (Pyles et al., 2007). Another research analyzed the STSp region and found that its functionality underlies the BLOD response. Such research used fMRI to analyze the actions with point-light animations. Viewpoint invariance of human action in STS is an abstraction in centering its representation of visual analysis (Grossman, 2010).

Giese (2014) presented an approach to describe a body in the aspect of psychology. This approach shows that body motion perception needs an integration of multiple visual processes involving Gastalt-like pattern and aggregation of the bottom-up and up-down processes with recognition based on learning (Giese 2014). A research on visual motion perception uses the integrated dynamic motion model to handle diverse moving stimuli. This research involves the random dot kinematograms and considers the motion integration and motion detection and perception in decision. Analysis is performed by generating the parameters in dynamic simulation (Tlapale et al., 2015). Jung and Gu (2015) showed another approach that combines perception and modeling in the visual motion. This approach follows the visual perception knowledge to diagnose the main details of video and efficiently remove the noise (Jung et al., 2015). Another approach characterizes the patterns and perception duration and categorizes them into three groups according to their direction cues, namely, cardinal, diagonal, and toward diagonal (Meso & Masson, 2015). The link between imagery and perception was investigated by putting the observers in the dark, which rotates in the left or right. The velocity of chair rotation should be high,

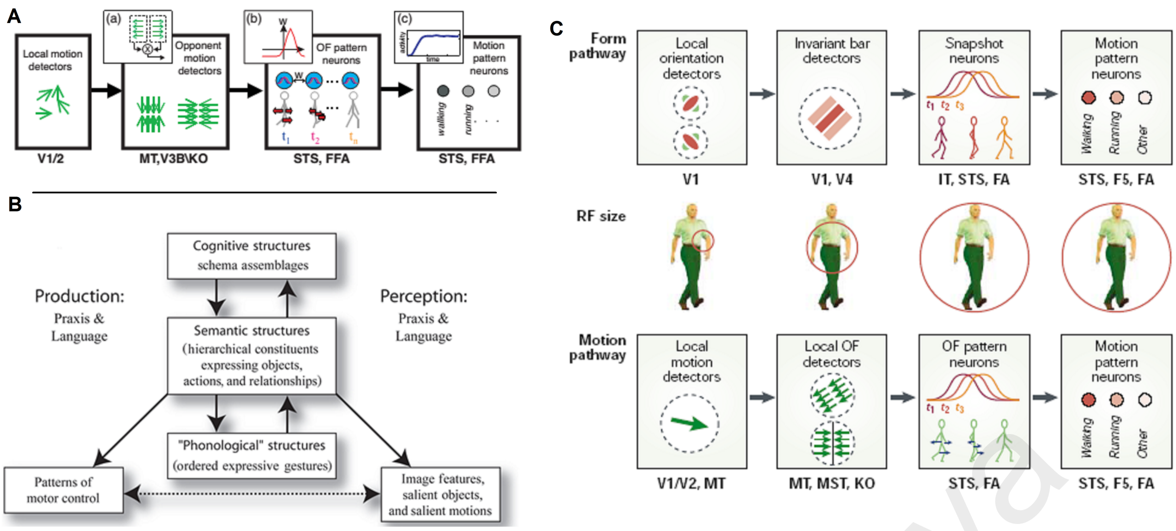

Figure 2.5: **A)**Schema of the model shown and symbols are shown following the brain areas and their functionality: MT: middle temporal area; V1: primary visual cortex; FFA: fusiform face area; STS: superior temporal sulcus; KO: kinetic occipital area. These areas and their functionalities are considered in their timing $t_1, t_2, ..., t_n$ for the input data frames and their encoded information gathered by radial base function and optical flow. Also (a) reveals the opponent motion detector; (b) shows the lateral coupling in complex optical flows; (c) response of the motion pattern detector Casile & Giese (2005).). **B)**The language production and perception performance review is shown in the figure(Arbib 2005)). **C)** illustrate the model presented in Giese, Martin& Poggio (2003) with concerns of receptive field as well.

and the direction of imagined rotation is different from physical rotation (Nigmatullina et al., 2015). In addition, Tadin (2015) used perception information to suppress the surrounding spatially, which involves sensory input. The ability to suppress information is a neural process that involves both perception and cognition (Tadin, 2015). Matsumoto et al. (2015) analyzed schizophrenia patients who have impairments in cognition, perception, and visual attention; they also analyzed the biological motion perception in 17 patients and 18 healthy controls (Matsumoto et al. 2015). Ahveninen et al. (2016) investigated the combination of spatial and non-spatial information in the auditory cortex (AC) of two parallel streams, namely, “what” and “where” that are modulated for visual cortex subsystems, as well as their integration regarding object perception. This approach uses animated video clips of two audiovisual objects, namely, black and gray cats, and records the magneto- and electroencephalography data. The events in sound are initially linked to object perception in posterior AC, with modulation representations in anterior

University of Malaya

Table 2.1: The perception approaches presented with their contribution in the field.

Perception		
Approach	**Topic of the approach**	**Connection to other researches**
E. J. Marey and E. Muybridge (1850s)	moving photographs presenting locomotion	
Rubin (1927)	visual perception of real movement	
Duncker (1929)	visual perception of real movement	
Johansson et al. (1973)	motion patterns for humans& animals as biological motion (MLD)	
Turaga et al., 2008	locomotion analysis	
Johansson, 1975	perception of human motion in neuroscience analysis	
Kozlowski & Cutting, 1977	with females and upper body MLD, gender recognition	
Marr et al. (1978)	computational process in human visual system, 3D shapes	**perception**
Perrett et al. (1985)	temporal cortex of macaque monkey analysis, found two cells in brain sensitive for rotation and view of the body movements	
Perrett et al. (1989)	view centered, view independent responses among the brain cells	
Goddard (1989)	spatial and temporal feature incorporation through diffuses MLD data	**perception-computer**
Goddard (1992)	synergistic manner of the process of “what” and “where” in visual system	**neuroscience**
Goodale & Milner (1992)	projection perceptual information from striate and inferotemporal cortex	**neuroscience-object identification**
Cédras and Shah (1995)	motion based recognition into motion models	**modelling**
Perkins (1995)	animated real-time, texture of motion, avoiding computational	**modelling**
Tsai et al. (1993)	detection of cyclic motion, applying Fourier transform	**highly related to computational modelling**
Fielding & Ruck (1995)	Hidden Markov Model (HMM) technique for classification	**highly related to computational modelling**
Gallese et al. (1996)	analysis the electrical activity in macaque monkey's brain	**neuroscience**
Aggarwal et al. (1999)	human motion analysis review and computer vision approaches	**computer vision**
Aggarwal & Cai (1997)	interpreting human motion, tracking, framely recognizing human activities	**perception**
Decety & Grèzes (1999)	Process of action and its perception, functional segregation MLD	
Rangarajan ey al. (1992)	matching the biological motion trajectories (object recognition)	**computer vision**
Neri et al. (1998)	visual system ability to integrate the motion information of walkers	
Hill & Pollick (2000)	temporal differences in MLD, recognition of the exaggerated motions	
Giese & Poggio (2000)	linear combination of prototypical views,3D stationary object recognition	**computer vision**
Moeslund & Granum (2001)	comprehensive survey on the motion capture	**computer vision**
Grèzes et al. (2001)	neural network specifications and its verifications through fMRI	**computer vision-neuroscience**
Pollick et al. (2001)	visual perception effects used point-light display(MLD)	**computer vision**
Song et al. (2001)	Computer-human interface. using joint probability density function (PDF)	**computer vision**
Wiley & Hahn (1997)	virtual reality approach regarding the computer generated characters	**computer**
Grossman & Blake (2002)	neural mechanisms, anatomical and functionality into two pathways	**neuroscience**
Jastorff at el. (2002)	investigating of recognition process in the neural mechanism	**neuroscience**
Giese & Lappe (2002)	spatio-temporal generalization of the biological movement perception	**computer vision**

AC (Ahveninen et al., 2016).

2.4 Knowledge based modeling approaches

Modeling of biological movements into systematical and mathematical models follows the neuro-physiological, physiological, and neuro-science evidence. This modeling is increasingly developed and considered one branch of this research field; many computer vision approaches also underlie this model. An engineering approach uses HMM and features based bottom-up approaches in time sequence images. The computer vision aspects of action recognition are more considered than the recognition of biological movements (Yamato et al., 1992). The linear combination of motion sequence prototypical views is considered an effective method to recognize and analyze the 3D stationary object images. The problem on corresponding space–time computations is solved using

University of Malaya

the algorithm presented by Giese and Poggio (1999). Biological movement simulates images and applies the prototypical superposition in motion sequences, thereby creating new video sequences, which are used to analyze action recognition. The topology over space for action patterns is considered in the video sequences. A proposed method follows the structural risk minimization principle, which provides knowledge regarding pattern space topology for recognition (Giese & Poggio, 1999). Gavrila (1999) presented a survey article in visual analysis regarding human movement. This survey represented several applications in this domain (particularly for the hand and whole-body motion) and included 2D approaches with or without shape models and 3D methods (Gavrila, 1999). A quantitative description on the geometry of human objects with movements is considered through fitting a 3D model projection to construct image frame sequences. The person model kinematic using homogenous transformation tree plus models the right elliptical cones in the body parts of the human subject. This model concerns the degree of freedom, which normally varies and is determined by solving the iteration expanded of Kalman filter; the velocity of the person model is constant, and only simple motions are included (Wachter & Nagel, 1997).

Gises and Poggio (2003) reviewed a complex research regarding the dual processing pathways in the visual system and their functions. This review discussed the mechanism for recognizing biological movement in the mammalian brain and analyzed the motor control by using quantitative models and neurophysiologically plausible tools for model establishment (Giese & Poggio, 2003). Human perception of movement is followed by stimulus approaches, such as point-light display. An evidence is presented on multilevel generalization by using simple mid-level optic flow features on coarse spatial arrangement. Some findings are as follows: point-light and normal walkers statistically share similar dominant local optic flow features, inconsistency of the human body skeleton with point-light stimulus, and critical features for considerable recognition (degraded stimuli). The dominant form and motion feature extraction in the mid-level of the moving subject is conducted by using principal component analysis (PCA). This analysis is commonly employed to extract the informative directions of high-dimensional information spaces.

Mid-level optic flow features from spatial localization are considerably effective to recognize biological movement (Casile & Giese, 2005).
Neural and its functionality grounding for language skills are analyzed by observing the premotor area F5 in monkeys and Broca's area in humans, which contain a mirror system to observe and execute manual actions. The major concern of this research is related to imitation (Arbib, 2005). Valstar and Pantic (2006) compared logically and biologically inspired methods for facial expression in human machine interaction. Face recognition in six basic emotional states and atomic facial muscles called action units (AUs) was analyzed. Classic psychological studies considered a finite number of rules to classify basic emotions and followed the recent studies using ANNs . A comparison was conducted between the detection of emotions from features versus the determination of AUs into an ANN for recognition. The results suggested the suitability of the biologically inspired approach on logical methods for this application (Valstar & Pantic, 2006). Demiris and Khadhouri (2006) analyzed the action perception in motor systems. They provided a review on computational architecture and hierarchical attentive multiple models for execution and recognition (HAMMER) for motor control systems on a robot by selecting an action competitively and performing and perceiving it during demonstration. Computational experiments showed the differences for controlling HAMMER. Biological evidence compatibility of action plans in recognition, which uses action and perception features, is also shown (Demiris & Khadhouri, 2006). Minler and Goodale (2007) also analyzed two cortical systems regarding the vision in action and perception inspired by Larry Weiskrantz. They summarized some essential concepts regarding the model, refinements, and particular clarifications (Milner & Goodale, 2008). A mid-level learning for motion features is presented by Fathi and Mori (2008). This approach is not directly biologically inspired, but it involves the motion patterns and optical flow (Fathi & Mori, 2008). Schindler and Gool (2008) presented a successful implementation of a biologically inspired model of human action recognition, which involves the motion and form information interaction. They showed the recognition of simple actions instantaneously by using short sequences (snippets) of 1–10 frames. A reasonable performance was obtained

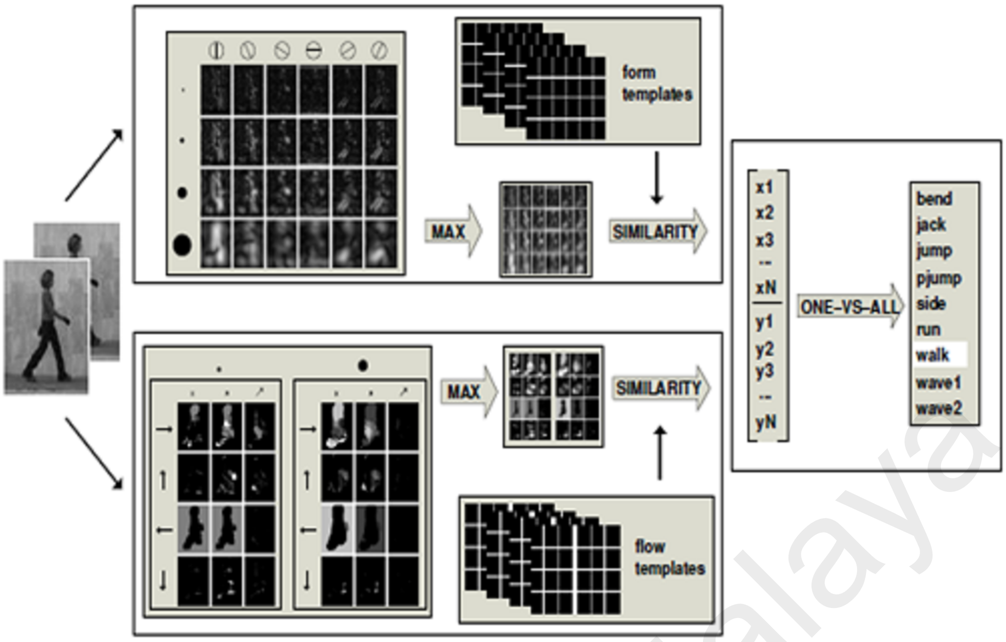

Figure 2.6: A biological inspired recognition of human action overview diagram. The two parallel processing streams are considered that log-Gabor filtering in the multiple directions and scales generate the form information and in the other hand optical flow extracts the information of motion considering the directions, scales and speeds. The max-pooling have been used in the both streams. This information concatenates to create input for linear classifiers Schindler & Van Gool (2008b).)

for the proposed approach, but its main objective is not about the biologically inspired model (Schindler & Van Gool, 2008a). Schindler and Gool (2008a) developed a method to recognize the action consisting of form (local shape) and motion (local flow) features in the video sequence. This method considers two features, which are inspired by separate independent pathways from the human visual system. Local pooling of feature maps, down sampling, comparison with an established template, and generation of similarity score vector for each channel were conducted. These scores are merged and given to a discriminative classifier (Schindler & Van Gool, 2008b).

Furthermore, Webb et al. (2008) presented the mechanism in intermediate levels of visual processing and investigation to detect circular and radial forms. This mechanism analyzes the detection of the global structure in spiral form using the array consisting of 100 Gabor that is randomly positioned within the window. The Gabor filter randomly ro-

University of Malaya

tates, and the structure can be detected when the mask and test have the same spiral pitch. The Gabor filter is extensively used in the form pathway, and the approach is significant for elucidating the mechanism of visual processing streams (Webb et al., 2008). Yau et al. (2009) investigated the recognition and interaction between vision and touch by studying the single neurons in macaque monkey intermediate visual (V4 area) and somatosensory (SII area) cortex to match the shape stimuli. The curvature direction for the mentioned regions was tuned, and PCA was utilized to identify the underlying patterns and shape feature selectivity for large response variances (Yau et al., 2009). A bio-inspired feed-forward of spiking network model was performed for the influences of the motion system (V1 and MT) on human action recognition. Two characteristics of neural code, namely, neuron synchronization and their firing rate, were considered. Spiking networks can be a potential alternative in actual visual applications (Escobar et al., 2009).

The dynamic representation of action recognition was analyzed through a pose descriptor called histogram of oriented-rectangles to represent the human action recognition in the video streams via rotating the human silhouette. The approach is more computational rather than bio-inspired, but it provides a considerable method (Ikizler & Duygulu, 2009). Another machine vision approach that can be considered a partially biological method uses the bag-of-word (BOW) representation of visual features as visual word and semilatent topic models (Wang & Mori, 2009). Ryoo and Aggarwal (2009) presented the spatiotemporal relation for recognizing human activity. The approach considers the known background, and spatiotemporal local features are used for short video and simple periodic actions. Notably, local spatiotemporal features are considerable because of their similarity to the receptive field utilization in the visual system (in the ventral and dorsal processing streams). The proposed method is an engineering approach that allows the localization and detection of the actual complex activities (Ryoo & Aggarwal, 2009). Shabani et al. (2010) proposed an engineering method on multiscale salient features from motion energy, which is encoded by local events and biologically plausible on perception model. The opponent-based motion energy from oriented motion filters is constructed from bio-inspired time causal filters. This approach also uses the BOW idea (Shabani

University of Malaya

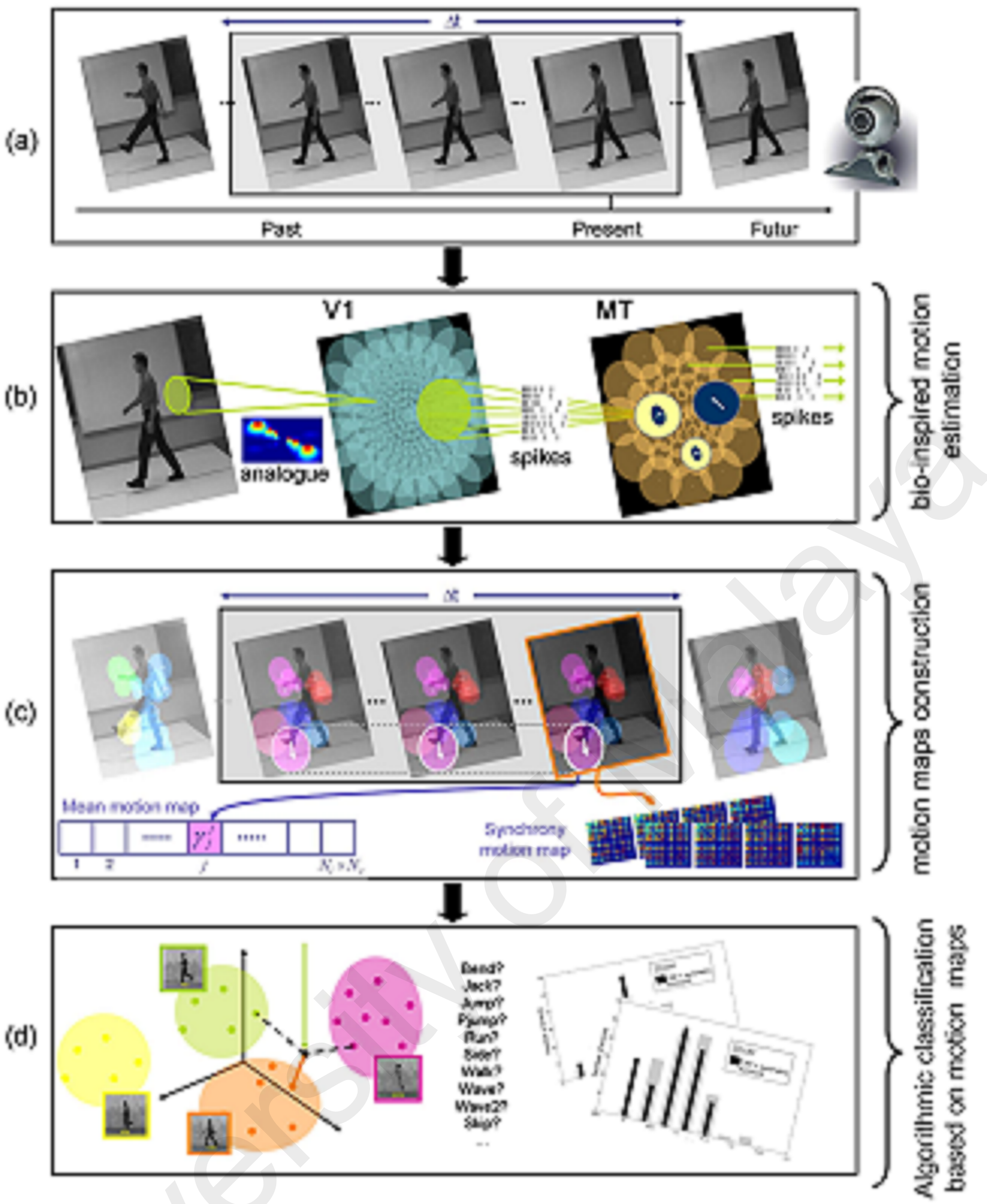

Figure 2.7: The different step of the presented approach in the block diagram is shown for input image sequence. a) depicted the video sequence of input video of human action, b) directional selection filtering applied to input images frames using log-polar as V1 output and these spike train and feed the spiking MT (gathering the spatio-temporal information), c) the motion maps through the estimation of the mean firing rates of MT or synchrony map of the spikes for MT cells. d) Show the classification step considering the information gathered from motion maps and training set Escobar et al. (2009).)

University of Malaya

et al., 2010). Poppe (2010) summarized the visual-based human action recognition and addressed some robust solutions on visual surveillance, human–computer interaction, and video retrieval. Ward et al. (2010) investigated the reference frames applied in terms of visual information by using fMRI. The analysis considers the receptive field scene processing areas, such as transverse occipital sulcus (TOS), retrosplenial complex (RSC), and parahippocampal place area (PPA). PPA and TOS show the position response curves on the fixation points to the screen (or the pattern), whereas RSC area does not (Ward et al., 2010). A review survey paper summarized the human action/activity recognition approaches that are categorized on the basis of the representation of spatial and temporal structures of actions (Weinland et al., 2011). Bio-inspired features in action recognition are presented by involving the motion in the models of cortical areas V1 and MT (shape and characteristics of their receptive field). A model with different surrounding geometries for MT cell receptive field is presented, which leads to bio-inspired features regarding the average activity of MT cells and how these features are used as a standard in the classification of activity recognition (Escobar & Kornprobst, 2012).

In addition, a fully automatic system for human action recognition is presented using convolutional neural networks (CNNs) in the uncontrolled environment. CNN is a deep learning approach, which is bio-inspired and develops a 3D CNN for the task. This approach extracts the features from spatial and temporal dimensions via a convolutional network, captures the motion information encoded from adjacent frames, and generates and combines multiple channel information. The presented approach has successfully implied a bio-inspired method through CNN and motion information combination for actual environments (Ji et al., 2013). Lehky et al. (2011) investigated the characteristic of sparseness selection in the anterior inferatemporal cortex on a large dataset. This research involves the information on 674 monkey inferotemporal cells for 806 object photographs and the two-way analysis of the responses of the entire neurons in single image (population sparseness) and column-wise (response of single neurons to all images). This research is related to the statistical analysis of stimuli in the primates and shows a large number of various critical features to tune different neurons. The approach also represents inconsistent structural-based object recognition tasks, and the objects are decomposed into small standard features (Lehky et al., 2011). Collisions that mostly apply for future robot and human interactions in complex visual environments are detected by analyzing

University of Malaya

two types of neurons, namely, Lobula giant movement detectors (LGMD) and directional selective neurons, in visual pathways of locusts. This research involves a model that tunes these two networks for collision tasks, compares them separately, and analyzes them co-operationally. The results show that LGMD can detect collision faster and more robust than other configurations. This research does not focus on biological movement recognition but considerably analyzes visual motion (Yue & Rind, 2013). Cai et al. (2014) presented bio-inspired model-based spatiotemporal interesting points for human action recognition to allow interest point detection and descriptor construction compared with the STIP framework. This model has been used for a long time and follows the dual pathway in the visual system model (Cai et al., 2014).

Guthier et al. (2014) studied the interaction and combination of the pathways in the visual system. The investigation focused on the recognition of complex biological articulated movements, such as gestures and expressions. They introduced a model that utilizes gradient and optical flow. The patterns are used by an unsupervised learning algorithm, translation-invariant nonnegative sparse coding called VNMF, and shaped prototypical optical flow patterns. In the learning processes, a lateral reserve term that eliminates competing pattern activations provide small sparse activations (Guthier et al., 2014).

Another study investigated a bio-inspired model of human action recognition that focuses on the influence of spiking neural networks in the visual cortex (Shu et al., 2014). The study provided a hierarchical architecture and considered two visual cortical areas, namely, middle temporal area (MT) and primary visual cortex (V1) for motion processing. The result was obtained by analyzing the horizontal connection of spiking neurons in every cortical area. A connection between V1 and MT areas called cross-talk, as well as its lateral connection, was analyzed considering the linear combination of normalized V1 direction-tuned signals. A 3D Gabor filter was tailored for V1 cells. A mean motion map for mean firing rates in MT area neurons called action code was used to analyze the biological movements. The action code comparatively improves recognition accuracy and computational efficiency. The method also increases the recognition performance

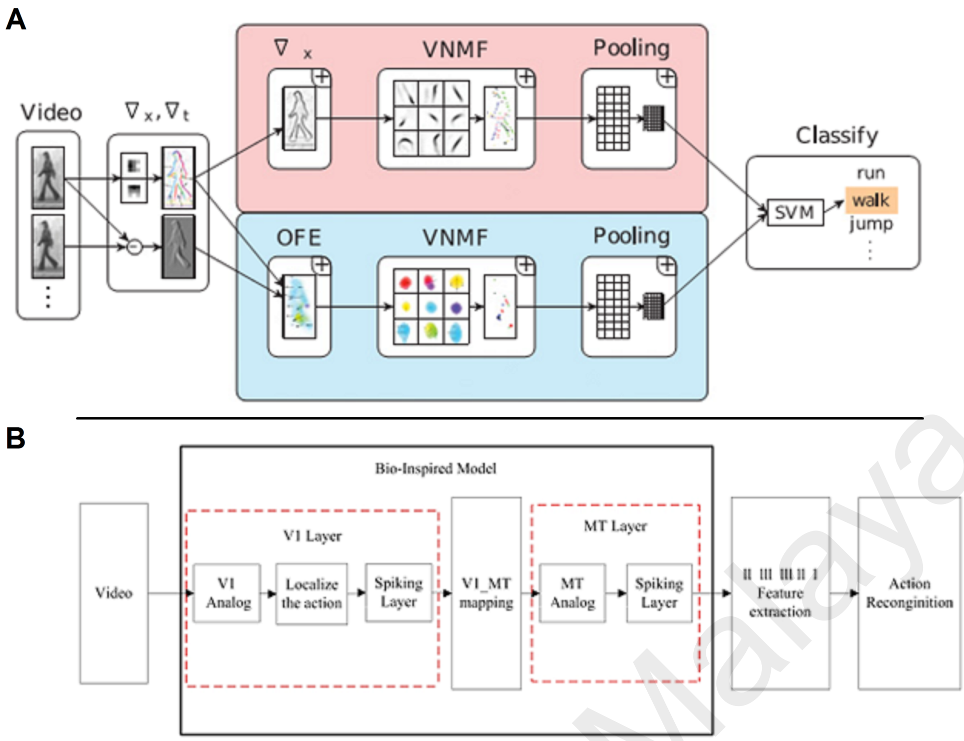

Figure 2.8: **A)** Two parallel processing streams in the model of biological motion recognition has been shown using the spatio-temporal gradients in the static path (red), it works by pre-learning the gradient patterns and sub-sequential pooling. On the other hand, projection of pre-learned optical flow field (OPE) in dynamic path (blue color) collected and sub-sequential pooling is performed in it. At the end, combined pooled activations are classified by SVM. VNMF algorithm has been used for midlevel patterns concerning nonnegative components usage Guthier et al. (2014)). **B)** A serial configuration of proposed approach consists of four different parts for recognition of human actions has been presented. The core part of the model is involved by V1 layer and V1 and MT and MT layer. In the V1 layer non-linear combination of perception information from simple cells using 3D Gabor spatiotemporal filter and in the MT layer, MT cells perform pooling the information from V1 cells following mapping between the MT and V1. The extraction from spiking neurons and used for classification by SVM Shu et al. (2014).

considering the serial model for the motion and form pathways against the generalized bio-inspired model (Shu et al., 2014). A bio-inspired approach for robust recognition of faces using C2 features in HMAX follows the dorsal and ventral stream visual cortex neural behavior (Esmaili et al., 2014); C2 features are extracted from visual attention points. The application of the BOW method in human action recognition was investigated on the basis of vector quantization (VQ), which presents an efficient method called context and locality constrained linear coding; a group-wise sparse representation-based classi-

University of Malaya

fication method that involves the sparse representation of the human action recognition was also investigated (Tian et al., 2015). BOW has been used at maxima of the sparse interest point operators, but the inconsistency of visual processing in biological systems with sparse approaches has not been addressed. The proposed approach has three contributions for bridging the gap. The approach conducts valuable analysis in human action recognition; however, a computer vision-based method is used instead of a bio-inspired modeling analysis (Mathe Sminchisescu, 2015). The approach was considered by Marc Jeanerod as the basic method of action (semantics and pragmatics) and movement. The ordinary representational resources of pragmatics and semantic types of actions following the evidence of simulation and language understanding were investigated. Three theoretical frameworks were mentioned by Prinz (2014): 1. Semantics is based on pragmatics; 2. Pragmatics is anchored on semantics; and 3. Pragmatics is a part and parcel of semantics. Sparse coding was employed for the BOW application in the action recognition approaches for basic shape extraction at the temporal structure of the action group. The representation of BOW for every video representation in the form of histogram of group sparse coefficients and its geometry was analyzed (Moayedi et al., 2015).

An engineering approach led to the creation of a video database for human action recognition presented by Schldt et al. (2004). This approach analyzes adaptive local space time features that are captured in the local events located in the video. A computational model follows the neural plausibility assumptions for the interaction of the form and motion signals in biological motion perception from figural form cues; the receptive fields in the images of a static human body were also analyzed (Lange Lappe, 2006). Willert et al. (2007) presented an approach that estimates motion using optical flow through a dynamic Bayesian network. The method involves spatiotemporal features that interact with motion information from two-filter inference in online and offline parameter optimization (Willert et al., 2007). This technique is similar to those used in other studies that estimated motion to analyze small numbers of temporal consecutive frames of action; at the same time, this method presents a certain class of transition probability functions that approximate

University of Malaya

inferences based on Belief propagation (Willert Eggert, 2009). Sun et al. (2010) found that the median filtering of the intermediate flow fields is the optimization gains of the performance and provided superior energy solution. An adaptive representation of visual motion processing was developed by Willert and Eggert (2011), resulting in an online adaptation of velocity-tuning curves inspired by physiological experiments on macaque MT. Another motion analysis approach considers movements in all directions, namely, circular, radial, spiral, and translational; this property makes it an important factor in the analysis of the primate dorsal visual system in visual perception. This approach investigates the egomotion-compatible visual stimulation in human systems and analyzes six sensitive motion areas, namely, V3A, V6, MST+, CSv, MT, and an Intra-Parietal Sulcus motion (IPSmot) region for different types of optical flow stimuli (Pitzalis et al., 2013).

Jhuang et al. (2007) presented a hierarchical feed forward architecture on the object recognition task in biological models of the visual cortex. The model contains spatiotemporal features with a relevant hierarchy that processes them and provides position-invariant feature detection (Jhuang et al., 2007). Another approach involves learning sparse spatiotemporal codes from the basis vectors, considering that the scale, direction, and velocity depend on the spatiotemporal features learned in unsupervised fashion. Sparse coding was used to provide an initial basis and was expanded to create new basis vectors recursively with a large temporal extent; this sparse coding propositionally conserved the previous weights (Dean et al., 2009). A model was developed to demonstrate an intermediate-level visual presentation comprising two stages; the first stage provides a representation of the early features in layer, whereas the second one is related to invariance. This model provides a rich representation of dynamic natural images in the visual cortex (Cadieu & Olshausen, 2012). A human action in the sparse representation shown in a set of overcompleted basis (dictionary) was used for human action recognition in the video sequence. These overcompleted bases were obtained using spatiotemporal feature descriptors and provided some linear combinations of dictionary elements; a compacted way of representation was eventually achieved, which includes VQ and clustering (Guha

University of Malaya

& Ward, 2012). This result was obtained within the condition of a non-negative sparse coding in biological motion presented by Guthier et al. (2012). The visual motion involved uses the optical flow and shape information obtained from the moving objects. This approach aims to capture the characteristics of local motion patterns following constant brightness and using the feature combinations from non-negative sparse coding; this method results in the learning of the basic patterns of motion (Guthier et al. 2013). Nayak & Roy-Chowdhury (2014) presented an approach using spatiotemporal features and their unsupervised relationship to dictionary learning in the model of activity recognition. This approach provides an unsupervised sparse decomposition framework for the relationship between the spatiotemporal features and the local information from descriptors, which create classifiers through multiple kernel learning (Nayak & Roy-Chowdhury, 2014). An extension to the proto-object model based on the saliency map involved depth information using the 3D eye tracking datasets presented by Hu et al. (2016). Another unsupervised approach performs synthetic biological movement recognition (Babaeian et al., 2015a, Babaeian et al., 2015b) and shows great potential for use in the mechanism of biological movements. Haghigh et al. (2016) studied human-like movements processed in the human brain and motor control. The study involved the concept of artificial intelligence and robotics, as well as learning the latent simple motions for imitation in more complex movements. It proposed MOSAIC structure in motor control modeling (Haghighi et al., 2016). Yousefi et al. (2013) presented models to recognize biological movement involving a supervised Gabor-based object recognition approach called ABM (Wu et al. 2010) in the ventral processing stream (Yousefi et al., 2013; Yousefi et al., 2014). A fuzzy-based optical flow proposed for dorsal streams was used to improve the model (Yousefi & Loo, 2014a; Yousefi & Loo, 2014b). Furthermore, an approach to involve slow features was presented for ventral processing stream (Yousefi & Loo, 2014).

2.5 Psychological and neuroscience point of view

The cellular population located in the temporal lobe of macaque monkeys' inner superior temporal sulcus (partially called STPa or STSa) was analyzed. The responses of

University of Malaya

Table 2.2: The Knowledge based modeling approaches presented with their contribution in the field.

Knowledge based modeling		
Approach	**Topic of the approach**	**Connection to other researches**
Yamato (1992, June)	HMM and feature based bottom up approaches in time sequence images	modelling
Giese & Poggio (1999)	Linear combination of motion sequence prototypical views, 3D object recognition	
Gavrila (1999)	survey article in visual analysis regarding the human movement	
Wachter & Nagel (1997, June)	quantitative description on the geometry of human object	
Gises & Poggio (2003)	dual processing pathways in the visual system	
Casile & Giese (2005)	multilevel generalization using simple mid-Level optic flow features	perception
Arbib (2005)	analysis of neural and its functionality grounding for the Language skills	perception, neuroscience
Valstar & Pantic (2006)	comparison of Logical and biological inspired methods for facial expression	computer vision
Demiris & Khadhouri (2006)	computational architecture and HAMMER for motor control systems	perception
Minler and Goodale (2007)	analysis of two cortical systems regarding the vision in action	perception
Fathi & Mori (2008)	mid-Level Learning for the motion features	modelling
Schindler & Gool (2008, June)	recognition of simple actions instantaneously by short sequences (snippets) 1-10 frames	computer vision
Schindler & Gool (2008)	recognition of form (Local shape) and motion (Local flow) features	computer vision
Webb et al. (2008)	intermediate Levels of visual processing, detection circular and radial form	
Yau et al. (2009)	interaction of vision and touch, PCA for patterns shape features identification	
Escobar et al. (2009)	bio-inspired feed-forward of spiking network model	neuroscience
Ikizler & Duygulu (2009)	analyzing the dynamic representation of action recognition using HOR	
Wang & Mori (2009)	visual features as visual word and semi-Latent topic models	modelling
Ryoo & Aggarwal (2009)	Spatiotemporal relation for recognition of human activity	
Shabani et al. (2010)	multiscale salient features from motion energy	modelling
Poppe (2010)	Visual based human action recognition	computer vision
Ward et al. (2010)	references frames applied for visual information using fMRI	
WeinLand et al. (2011)	review paper for human action/activity recognition	
Escobar & Kornprobst (2012)	analysis motion in the models of cortical areas V1 and MT	neuroscience-perception
Ji et al. (2013)	fully automatic system for human action recognition by CNN	modelling
Lehky et al. (2011)	characteristic of selection of sparseness in anterior inferotemporal cortex	neuroscience
Yue & Rind (2013)	detection of collisions, analysis of two types neurons: LGMD and DSNs	neuroscience
Cai et al. (2014)	spatiotemporal feature in the bio-inspired model, BIM-STIP	
Guthier et al. (2014)	survey, modelling using nonnegative sparse coding, VNMF	
Shu et al. (2014)	bio-inspired modeling human action recognition, spiking neural network	
Esmaili et al. (2014)	robust recognition of face using C2 features in HMAX	computer vision
Tian (2015)	BOW method, VQ ,CLLC, GSRC in the human action recognition	
Mathe & Sminchisescu (2012)	BOW in maxima of sparse interest operators	
Prinz (2014)	analysis of action semantics and pragmatics	perception
Moayedi (2015)	basic shape extraction of action group sparse coding employed BOW	
SchuLdt et al. (2004)	adaptive local space time features	Computer vision
Lange & Lappe (2006)	Neural plausibility assumptions for interaction of the form and motion signals	
Willert et al. (2007)	estimating the motion using optical flow by dynamic Bayesian network	
Willert & Eggert (2009)	estimation of motion to analyze the small number of temporal consecutive frames	
Sun et al. (2010)	median filtering of the intermediate flow fields	
Willert & Eggert (2011)	representation of visual motion processing	
PitzaLis et al. (2013)	motion analysis approach considers the movements in all directions	perception
Jhuang et al. (2007)	hierarchical feed forward architecture on the object recognition	
Dean et al. (2009, December)	Learning sparse spatiotemporal codes from the basis vectors	
Cadieu & Olshausen (2012)	intermediate-level visual presentation	
Guha & Ward (2012)	human action in the sparse representation in overcompleted basis (dictionary) set	
Guthier et al. (2012)	non-negative sparse coding on biological motion	
Yousefi et al. (2013)	Introducing Active Basis Model for ventral stream	Computer vision
Yousefi & Loo(2014)	fuzzy optical flow division in Dorsal stream	Computer vision
Yousefi & Loo(2014)	Interaction between dorsal and ventral streams	Computer vision
Yousefi & Loo(2015)	Slowness principal into modelling	Computer vision
Yousefi & Loo(2015)	Hybrid Max-Product Neuro-Fuzzy Classifier and Quantum-Behaved PSO in the model	Computer vision
Yousefi & Loo(2015)	Slowness prototypes in ventral stream	Computer vision
Nayak & Roy-Chowdhury (2014)	spatiotemporal features, unsupervised way into a dictionary Learning	
Hu et al. (2016)	proto-object based on the saliency map	computer vision
Haghigh et al. (2016)	human-Like movements	

Table 2.3: The psychological and neuroscience approaches presented with their contribution in the field.

Psychological and neuroscience		
Approach	**Topic of the approach**	**Connection to other researches**
Jellema et al. (2000)	analysis of the cellular population located in the temporal lobe of macaque monkey	**Psychology**
Billard et al. (2000, June)	action imitation considered the actions high-level abstractions	
Vaina et al. (2001)	investigation regarding the neural network, fMRI in MLD	
Goodale & Westwood (2004)	evaluating the labour division at visual pathways	
Banquet et al. (2005)	associative learning for object location level, in CA3-CA1 region	
Cook et al. (2009)	ASCs for comparing detection of non-biological and biological motion	
Wyk et al. (2012)	action representation at STS	**Psychology**
Milner & Goodale (2006)	involvement of dorsal stream in movement to target following ventral stream	
Hesse & Schenk (2014)	visuomotor performance of a D.F. patient tested for letter-posting task	
Schenk (2012a)	DF patient analyzes the ability to get the object	
Schenk (2012b)	Using fMRI the functionality of DF patient	
Whitwell et al. (2014)	test by different width of the objects and DF, distinguish the shape perceptually	**perception**
Whitwell et al. (2015)	ability grip scaling is may rely on online visual or haptic feedback (for DF patient)	
Krigolson et al. (2015)	review of the behavior using EEG	**Psychology**
Cavina-Pratesi et al. (2015)	brain circulation using fMRI regarding the word recognition ability	
Ganos et al. (2015)	voluntary movement considering GTS area	
Fleischer et al. (2013)	visual recognition from motion	
Theusner et al. (2014)	motion energy based on the luminance of objects motion detectors	
Yamamoto & Miura (2016)	visual object motions on time perception	
Schindler & Bartels (2016)	on 3 dimensional visual cue involving the motion parallax analyzing	
Venezia et al. (2016)	the sensorimotor integration of visual speech through the perception	**perception**
Harvey & Dumoulin (2016)	visual motion effects on neural receptive field and fMRI response	

these cells were associated with the agent's action performance as it reached the targeted position. The properties of these cells in the STSa and their relationship to these actions were described (Jellema et al., 2000). A selective response of subset cells considers eye gazes, body posture, and faces. The second subset cells provide response to the limb movements in particular directions and are modulated by the direction of attention. The combination of the direction of attention and movements of the body significantly supports action detection (Jellema et al., 2000). Another biologically inspired approach that focuses on action imitation considers the actions high-level abstractions related to spinal cord, pre-motor and primary cortex, temporal cortex, and cerebellum regions at the connectionist level. The movement of the spinal cord is predefined by rhythmic arm and leg movements related to open-loop walking (and such actions). Such learning has been done through DRAMA neural architectures in spatiotemporal invariance and time series. The approach analyzes three types of action learning, namely, oscillatory movements, repetitive form of the arms and legs, and exact movements of reaching and grasping (Billard & Matarić, 2000). Vaina et al. (2001) investigated the neural network engaged in the recognition of biological motion through fMRI using point-light for the major joints of

University of Malaya

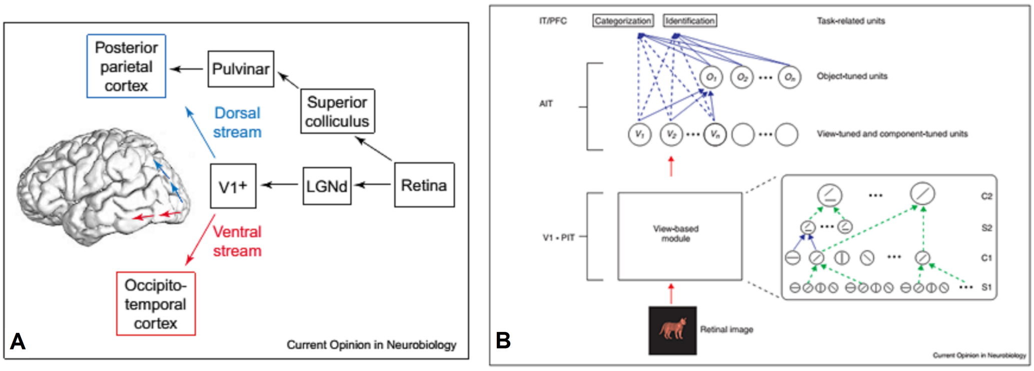

Figure 2.9: **A)**The invariant selection in the level of recognition and modeling of cortical object recognition has been simply shown (Riesenhuber & Poggio 2002). **B)** The scheme representation of two streams of visual processing located in human cerebral cortex and in the information gathered from retina for projections the data in the dorsal path for lateral geniculate nucleus in the thalamus (LGNd), and the projection to primary visual cortex (V1). The ventral stream (red) information of early visual areas (V1ρ) and its projection in the occipito-temporal cortex and dorsal stream (blue) projection of the information to the posterior parietal cortex are shown. The indicated routes represented by the arrows and involvement of complex interconnections .

the agent. The study also involved the rigid and non-rigid motion responses of biological motion with gender concerns. In particular, brain activity responses for the recognition of biological motion involve the areas of lateral cerebellum and lateral occipital cortex in response to area KO. The involvement of both ventral and dorsal processing stream, as well as the activation of lingual and fusiform gyrus and Brodmann areas 22 and 38, 19 and 37 STS, and 39 and 7, is mentioned during the biological motion. Their study examined stroke patients with unilateral brain lesions (Vaina et al., 2001). Goodale and Westwood (2004) presented another approach that evaluates the labor division at visual pathways and completed their hypothesis on the primate cerebral cortex between ventral streams dedicated to visual perception and dorsal stream for visual control in action. The study analyzed the psychological evidence and the response to visual motor control; in particular, the neurobiological challenge in mapping these behavioral findings onto the brain was analyzed and compared with known information about ventral and dorsal streams (in primate neurophysiology and human neurophysiology) (Goodale & Westwood, 2004).

An associative learning type analysis was presented for object location level, spatio-

University of Malaya

temporal level in the CA3–CA1 region, and movement-related information in the entorhinal cortex. This letter also analyzed the behavioral implementation and multimodal integration, which suggested the functional interpretation in hippocampo-cortical systems (Banquet et al., 2005). A research on autism spectrum conditions (ASCs) compared the detection of non-biological and biological motion in human adults through psychological evidence obtained from participants who watched biological (hand movements) and non-biological (falling tennis balls) stimuli. The ASC group did not show proper responses to perturbation from biological motions based on velocity profile (Cook et al. 2009). Wyk et al. (2012) presented an action representation at the superior temporal sulcus (STS) and determined its significance in biological motion perception; they evaluated the actions (congruence and incongruence) of 37 children (mean age: 11) and 17 adults (mean age: 25.4) by using fMRI. In congruence actions, both groups bilaterally showed the activation at the posterior STS; in response to incongruence, children showed response changes in the STS regions. An incongruency effect was observed in the older children and adolescents in the experiment (Vander Wyk et al., 2012). The dorsal stream was suggested to be involved in movement to the target following ventral stream visual representation processing delay (Milner & Goodale, 2008). The visuomotor performance of a DF patient was tested through a letter-posting task. The absence of environmental cues was observed in the DF patient, causing them to be unaffected by delay (aforementioned). The findings suggest that ventral stream damage does not consistently influence delayed movements but affects the visual feedback and environmental landmarks (Hesse & Schenk, 2014). Another investigation on DF patients analyzed their ability to acquire an object by distinguishing its geometry. Using fMRI, the functionality of a DF patient uses the intact visuomotor system housed within the posterior parietal lobe in the dorsal stream. Moreover, Schenk (2012a, 2012b) described the non-functioning of visuomotor networks in the dorsal stream, which was caused by a haptic feedback of the targeted object's edges. A test was conducted using different object widths, and the DF patients could grasp them within the healthy range (unlike the hypothesis that they should not); moreover, haptic feedback did not improve the ability of the DF patient to distinguish

University of Malaya

the shape perceptually (Whitwell et al., 2014). Another research mentioned that ability grip scaling may rely on online visual or haptic feedback. The grip scaling of the DF patient did not activate while her vision was suppressed in a grasp movement, showing that the haptic feedback after perception impairs the DF patient's performance. The research showed that DF patient's spared grasping task relies on dorsal stream functioning at the normal mode (Whitwell et al., 2015). Krigolson et al. (2015) presented a review of the behavior on three areas, namely, feedback processing, feed-forward control, and target perturbation; electroencephalography (EEG) was utilized to determine the temporal nature application in the goal-directed action. The cognitive potential and neural processing timing to motor control were further analyzed (Krigolson et al., 2015). A research used fMRI to investigate brain circulation during word recognition in the left fusiform and left inferior frontal areas of the gyrus, as well as the left middle temporal cortex of the DF patient. The left fusiform activations called visual word form area appeared from the FFA and hypothesized that this area lies outside the LO (Cavina-Pratesi et al. 2014). A research on voluntary movement in the Gilles de la Tourette syndrome comprised 25 patients. The results suggested that the brain learns voluntary control by perceptually discriminating signals from noise (Libet et al., 1983).

A research on visual recognition of motion involved several cortical regions, namely, premotor, partial cortex, and STS (Fleischer et al., 2013). It research performs for hand actions more till biological movement of human but it can be partially considered the human action. The model provides unifying quantitative reliability using the electrophysiological results from action-selective visual neurons (Fleischer et al., 2013). Theusner et al. (2014) established a model for action recognition in the brain. The model follows the motion energy based on the luminance of object motion detectors for the cortical representation of body posture similar to spatiotemporal receptive fields in posture-time space. This property was observed in the STS of macaque monkeys, and the 3D views for static and moving bodies were analyzed. Perception appearance, motion phenomenon, and static images obtained from motion neural activation were explained (Giese 2014).

University of Malaya

Yamamoto and Miura (2016) analyzed visual object motions on time perception. They investigated the line segments in front or behind the occluders in different speeds and followed the association of time perception with global motion processing (Yamamoto & Miura, 2015). A study on 3D visual cue involving motion parallax analyzed the link between the visual motion and scene processing by using fMRI. Parallax-selective responses were found in parietal regions IPS4 and IPS3, and in the region of occipital place area. Some regions such as the RSC and PPA do not respond to the parallax (Schindler & Bartels, 2016). Venezia et al. (2016) analyzed the sensorimotor integration of visual speech through perception. The study used fMRI on healthy individuals to identify new visiomotor circuit speech production (Venezia et al., 2016). A research on how visual motion affects neural receptive fields and fMRI response amplitudes was carried out to examine visual motion neural position preferences in the hierarchy of visual field maps using high-field fMRI and population receptive field. The results showed that visual motion induces the transformation of visuo-spatial representations through the visual hierarchy (Harvey & Dumoulin, 2016).

2.6 Some approaches relevant to the subjects

Hubel and Wiesel proposed that object recognition occurs from simple to complex cells, and this was investigated by Riesenhuber and Poggio (1999). Quantitative modeling was conducted on biological feasibility for high-level recognition of the object (the model is based on MAX-like operation) (Riesenhuber & Poggio, 1999). Tarr & Blthoff (1998) proposed an observation analysis regarding 3D object recognition in man, monkey, and machine by analyzing their biological plausibility and computational strengths and weakness. The method was expanded to a biologically inspired model approach for human arm movements and human action high-level abstraction by using the hierarchy of artificial neural networks. This model demonstrated that abstraction occurs in the visuo-motor control area of the brain and detected 37 degrees of freedom and biomechanical simulation with humanoids (Billard & Matarić, 2001). Another approach on action imitation explains the natural action through visual analysis of actions and motor representation of

University of Malaya

the nervous system. The evidence of the existing system is mentioned by mirror system for mapping in primates and humans (Rizzolatti et al., 2001). The human imitation of machines has been investigated for the purpose of flexibility, usefulness, and development of user-friendly machines. The approach concentrates on understanding how robots determine what to imitate, as well as the process of mapping perception onto the action it is imitating (Breazeal & Scassellati, 2002). Another method concentrates on the classification of gender in the biologically inspired approach of Troje (2002), which discusses the sensitivity of biological movement and its social information extraction. The level of encoded information and its retrieval in such a motion were investigated by transforming this information into pattern recognition and statistical data. The classification was analyzed by comparing it with psychological data for human observers strive for gender classification. The results indicate that the dynamic part of the mechanism contains more information on gender than on motion-mediated structural cues; however, its application is not limited to gender and can be used to synthesize new motion patterns (walking) (Troje, 2002). A study on human and monkey behavior using fMRI was conducted to observe the mechanism of cortical object recognition (Riesenhuber & Poggio, 2002). The study covered the invariant selection in the level of recognition and modeling of the cortical object recognition.

The imitation of movement through observation suffers from pose estimation, tracking, movement recognition, and coordination transform, all of which were investigated through the perceptional understanding of biological movements. The mathematical and statistical approaches that tackle parts of the imitation problem and the motor side of the imitation were investigated. The results argued that the perceptual system for movement identification and the spatial information correspond to these actions (Schaal et al., 2003). The cognitive development agent in the imitation and its architecture in the recognition of action was presented and implemented in the robots. The understanding and generation of actions, as well as the ability to learn new composite actions during the mentioned architecture, were also investigated (Demiris & Johnson, 2003). Johnson and Demiris (2005) presented an abstraction model on planning robot actions, as well as its

University of Malaya

recognition and imitations. The study analyzed the forward and inverse models of computational substrate of actions for the mentioned purposes by arranging the hierarchy of inverse and forward models. The top-down and bottom-up processes were checked by human demonstration in the object manipulation tasks performance for multi-level motor abstraction capacity analysis. These abstractions enable recognition in the different visual systems depending on the speed of performance (Johnson & Demiris, 2005); for instance, if the performances are slow, the noise overcomes the movement signal and causes failure in the low level. However, the high-level recognition might be successful, as the total failure actions result from excessively rapid performance in the mentioned architecture (Johnson & Demiris, 2005). Analysis of object recognition and model of motion processing in visual cortex comprises a hierarchy of spatiotemporal feature detectors. Motion direction (with the experiment) and position invariant spatiotemporal feature detectors were analyzed (Jhuang et al., 2007). Blais et al. (2009) proposed another orientation in the investigation of shape representation detection. The study analyzed the rotation modulates and linear non-separable effects on 2D and 3D shape targets for visual search (Blais et al., 2009). A dynamic 3D object recognition decomposes the spatiotemporal signature in long-term observers to examine its coding. The mental representation of the new objects was found to be decomposed under the law of perceptual organization; moreover, the human brain compulsively deals with such signatures, showing a non-influential effect on the observers for the extraction of features from unsmooth sequences after a particular scrambling of temporal scales (Wang & Zhang, 2010). Absi and Abdullah (2010) presented an approach on the human visual system recognition of objects; this approach was inspired by the feedback and feed-forward mechanisms in the human visual system. Livne et al. (2012) used a video-based 3D pose tracking involving biological motion to study human attributes such as gender, weight, and mental state (e.g., sadness and happiness). A high-level shape information analysis was carried out to examine 3D object recognition through eye movement patterns (Leek et al., 2012). In this experiment, eye movements were recorded on the objects that were analyzed. Data fixation was performed for shape analysis based on convex surface curve, internal concave regions, and external bounding contour. This method provided new evidence that supports the influence of eye movements on shape processing in the human visual system (Leek et al., 2012).

2.7 Slow Feature Analysis

2.7.1 Definition

In general, several relatively regular notations are summed up in the subsequent part. The typical Kronecker pointer meaning that it is zero for $i \neq j$ and equivalents one for $i = j$ is specified by δ_{ij}. Moreover, the probability of a signal which is continuum in time domain $x_t \in R_N$ more than the episode $[0,T]$ is indicated by $< x_t >:= \frac{1}{T}\int_0^T x_t dt$. Two length signals T covariance with its balanced empirical evaluator for the time-discrete container is described as:

$$c_{x,y} = cov(x_t, y_t) := \langle x_t y_t^T - < x_t >< y_t >^T \rangle \tag{2.1}$$

$$\hat{C}_{x,y} = c\hat{o}v(x_t, y_t) := \frac{1}{t-1} XY^T - \frac{1}{T^2}(X\mathscr{I})(Y\mathscr{I})^T \tag{2.2}$$

A signal x_t covariance beside a version of time-shifted of itself is worded autocovariance.

$$C_{x,\Delta t} := cov(x_t, x_{t+\Delta t}) \tag{2.3}$$

$$\hat{C}_{x,\Delta t} := c\hat{o}v(x_t, x_{t+\Delta t}) \tag{2.4}$$

Lacking of shifting in the time, one attains the symmetric covariance matrix of the signal x_t.

$$C_x = cov(x_t) := cov(x_t, x_t) \tag{2.5}$$

$$\hat{C}_x = c\hat{o}v(x_t) := c\hat{o}v(x_t, x_t) \tag{2.6}$$

University of Malaya

At this time, $\mathscr{I}$ represents a vector including T times the component $1, I$ is the matrix of unit and $X \in R_{N \times T}$ has all points of data organized as vector column and is described the matrix of data. A data matrix version of including the similar vectors as X but changed in time, Δt, is indicated $X_{\Delta t}$.

2.7.2 General problem statement

The SFA task of learning as Wiskott and Sejnowski (2002) initially formulated as follow: a time series input known as $x_t \in R^N, t \in [0, T]$ K real-valued instantaneously find as a set of functions $g_1(x), ..., g_k(x) \in \mathscr{F}$ which generates the time series as the output $y_t = g(x_t)$ such that

$$slowness_j = s(y_{j,t}) :=< \dot{y}_{j,t}^2 =! = min > \tag{2.7}$$

The objective of the slowness principle under the below mentioned constraints

$$\begin{aligned}
&\forall j \quad < y_{j,t} >= 0 && zero\, mean, \\
&\forall j \quad < y_{j,t}^2 >= 1 && unit\, variance, \\
&i < j \quad < y_{i,t}, y_{j,t} >= 0 && decorrelation\, and\, order.
\end{aligned}$$

Slow feature analysis basically and temporally varies as of filter of simple-low pass. As g_j functions include merely instant scope, which is they plan input x_t at a specific time t at the same time to an output $y_{j,t}$. The main purpose is to attain smoothness temporally although in the immediate restriction strong processing. The further constraints for problem of optimization certify to insignificant solutions are eliminated. The unit variance avoids steady signals to appear, enforces the decorrelation constraint and the zero-mean constraint for convenience. The solution uniqueness is assured while the functions are extracted one following a different such that the function of slowest feature in F is g_1 and the function of slowest feature $y_{j,t}$ that is decorrelated to all signals $y_{i,t}$ for $1 =< i < j$. Variation is temporally computed with the first derivative squared that can be estimated through a finite difference (Berkes & Wiskott (2005); Bohmer et al. (2011); Bray &

University of Malaya

Martinez (2003);).

$$\hat{y}_{j,t} = \lim_{\Delta t \to 0} \frac{y_{j,t+\Delta t} - y_{j,t}}{\Delta t} \overset{\Delta t}{\approx} y_{j,t+1} - y_{j,t} \tag{2.8}$$

In the linear function cases $g_j(x_t) = w_j^T(x_t - < x_t >)$ by input signals having zero mean property, like one of the constraints, the calculation can be prepared as follows:

$$\begin{aligned} s(y_{j,t}) =< \hat{y}_{j,t}^2 >=< (y_{j,t+1} - y_{j,t})^2 >= \langle (w_j^T \dot{x}_t)^2 \rangle \\ = w_j^T \; cov(\dot{x}_t) w_j =: w_j^T A w_j \end{aligned} \tag{2.9}$$

$$\begin{aligned} \langle y_{j,t} . y_{j,t} \rangle = \langle w_i^T (x_t - < x_t >) . w_j^T (x_t - < x_t >) \rangle \\ = w_i^T cov(x_t) w_j =: w_i^T B w_j \end{aligned} \tag{2.10}$$

The solution for problem of the generalized eigenvalues can be obtained by:

$$AW = BW\Lambda \tag{2.11}$$

The second order statistics matrices, A and B are positive-semidefinite and symmetric and in this way all eigenvalues are real values and greater/equivalent to zero. Also Λ denotes for eigenvalue matrix. The optimal weights $w_{1..K}$ straightly yields the eigenvectors related to $\lambda_{1..K}$ as the smallest eigenvalues. Normalization to $w_i^T B w_j = \delta_{ij}$ completes two other constraints in the initial equations. If we consider A and B having the properties of symmetric and thus all eigenvalues are real and greater than or equal to zero and second order statistics they are positive semi-definite. The each function slowness for every g_j is specified through $s(y_{j,t}) = \lambda_j$ and the K slowness components is $s(y) = \sum_{j=1}^{k} \lambda_j$.

2.7.3 Incremental Slow Feature Analysis

An incremental form of slow feature analysis (IncSFA) presented here, united covariance-free incremental minor components analysis (CIMCA) and candid covariance-free incremental principal components analysis (CCIPCA) (Kompella et al. 2012). Features of IncSFA complicatedly revised is linear considering the dimensionality of input, though

University of Malaya

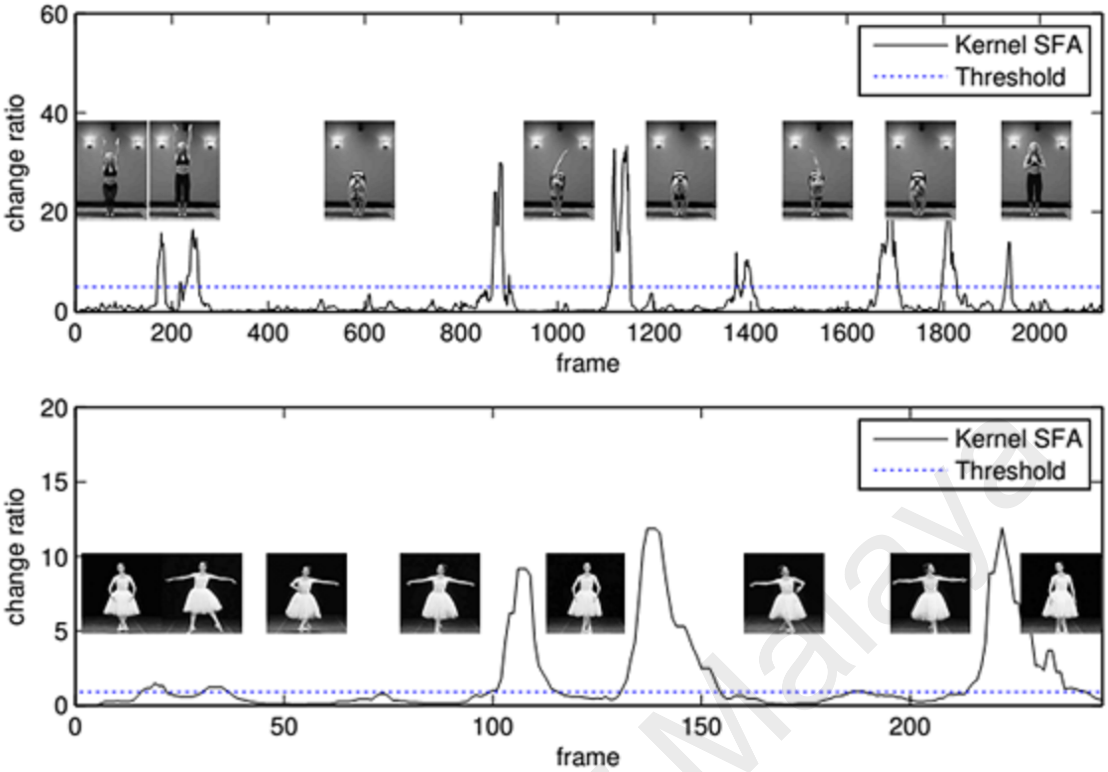

Figure 2.10: The implementation of incremental slow feature analysis in Krein space has been revealed in the figure and it depicts the variations in different actions and it has been examined for two video streams which were showing different actions (photo from Liwicki et al., 2013).

updating of batch SFA's (BSFA) difficulty is cubic. IncSFA does not require accumulating, or even calculating, any covariance matrices. IncSFA disadvantage is in the efficiency of data: it does not apply all data point as efficiently as BSFA. However IncSFA permits SFA to be tractably used, by a few parameters, straight input stream having high dimensional (e.g., an autonomous agent in visual input), though BSFA has to resort to hierarchical architectures of receptive field while the dimension of input is too high. Additionally, updates of IncSFA have Hebbian and anti-Hebbian simple forms, enlarging of SFA in biological plausibility. Experimental results illustrate IncSFA finds out the features in same set as BSFA and can have only some cases that BSFA fails. Slow Feature Analysis (SFA) is a learning inspired method from human visual system subspace, though, it is rarely observed in computer vision. The motivation of application for unsupervised activity analysis; the method increases first implementation of SFA online temporal in segmentation of video for detection of changes in motion within episodes. The method

University of Malaya

operates a particular domain indefinite kernel that obtains the representation of data into account for robustness introduction. Since the kernel is indefinite (i.e. describes in the place of a Hilbert, a Krein space), the approach originates Krein space-SFA. It suggests a framework for incremental kernel SFA that uses the special characteristics of the kernel.

2.8 Chapter summary

Studies in the literature have shown many significant development in the field. Also there are some points can be taken to the account such as:

The research on ventral processing stream can be significantly improved by applying shape learning instead of current approaches (Gabor filter alone).

Despite, there are many researches conducted for interaction between the pathways in the field of neurophysiology and physiology. However, there are less works done in the implementation of interaction performed in the computational mechanism.

Furthermore, SFA introduced as powerful tool and can be introduced into the mechanism. Aforementioned points consider as research gapes and will be reviewed in the methodology chapter to find the solutions.

University of Malaya

CHAPTER 3:

Method

This chapter describes the representative methods proposed during this project. The main reference of this paper is the original model for the recognition of biological movement based on the mammalian visual system (Giese & Poggio, 2003). The highly relevant research subjects that considerably followed this model are from Giese and Poggio (2003), Schindler and Van Gool (2008), and Danafar et al., (2010). This chapter presents ideas that can improve biological movement recognition models by following several steps. First, the approach involves model development in the ventral processing streams using the supervised Gabor wavelet-based technique of object recognition called Active Basis Model (ABM) (Section 3.1) (Wu et al., 2010; Kompella et al., 2012; Yousefi et al., 2013; Yousefi & Loo, 2014b). An improvement in the dorsal processing stream which generates motion information by the fuzzy optical flow division method (in Section 3.1). Finally, an incremental slow feature analysis was utilized into the mechanism which generates the slow features and its combination with the fast feature from the dorsal stream for recognition (Section 3.3).

University of Malaya

3.1 Active Basis Model in the ventral processing stream and balanced synergetic neural network

A supervised Gabor wavelet based technique for object recognition task used in ventral processing stream (it relates to the form of object) (Yousefi et al., 2013). The proposed system addresses a biologically inspired system as in Giese & Poggio (2003), Schindler and Van Gool (2008), and Guthier et al. (2014) where there a two independent paths and the inputs of the mechanism are the video sequences.

3.1.1 Active basis Model

Active basis model (Wu et al., 2010) applies Gabor wavelets (for elements dictionary) offers deformable biological template. Shared Sketch Algorithm (SSA) is followed through AdaBoost. In each iteration, SSA follows matching pursuit chooses an element of wavelet. It checks the objects number in different orientation, location, and scale. Selecting the small number of elements from the dictionary for every images (sparse coding), therefore can be represented of image using linear combination of mentioned elements, considering U as a minor residual.

$$I = \sum_{i=1}^{n} c_i \beta_i + \varepsilon \tag{3.1}$$

where $\beta = (\beta_i, i = 1, ..., n)$is set of Gabor Wavelet elements and components of sin and cosine, $c_i = \langle I, \beta_i \rangle$ and ε is unsolved image coefficient. By using wavelet sparse coding, large number of pixels reduces to small number of wavelet element. Sparse coding can train natural patches of image to a Gabor like wavelet elements dictionary that carries the simple cells in V1 properties (Olshausen & Fiel, 1996). The extraction of local shapes will be separately done for every frame and like responses of filter orientation and density of each pixels computes. Also, the active basis model uses the Gabor filter bank but in different form. A Gabor wavelets dictionary, comprising n directions and m scales is in the form of, $GW_j(\theta, \omega), j = 1, \ldots, m \times n$. Where, $\theta \in \{\frac{k\pi}{n}, k = 0, \ldots, n-1\}$ and $\omega = \{\frac{\sqrt{2}}{i}, i = 1, \ldots, m\}$. Gabor wavelet features signifies the object form as small variance in size, location, and posture. Though overall shape structure, it considers to be maintained

University of Malaya

throughout the process of recognition. Response (convolution) to each element offers form information with θ and ω.

$$B = \langle GW, I \rangle = \sum \sum GW(x_0 - x, y_0 - y : \omega_0, \theta_0) I(x, y). \tag{3.2}$$

Let GW_j be a $[x_g, y_g]$, I is a$[x_i, y_i]$ matrices, response of I to GW is a $[x_i + x_g, y_i + y_g]$.Therefore, previous convolution both matrices must be padded through sufficient zeros. Consequence of convolution can be eliminated via cropping the result. Additional technique would be to shift back the center of the frequencies (zero frequency) to center of the image though it might reason for loosing data.Obtained training image set $\{I^m, m = 1, ..., M\}$, joint sketch algorithm consecutively chooses B_i. The fundamental opinion is to find B_i so that its edge segments obtain from I_m become maximum. Afterward, it is necessary to compute $[I^m.\beta] = \psi \mid \langle I^m.\beta \rangle \mid^2$ for different i where $\beta \in Dictionary$ and ψ represents sigmoid, whitening, and thresholding transformations and then for maximizing $[I^m.\beta]$ for all possible β will be computed. Let $\beta = (\beta_i, i = 1, ..., n)$ is the template, for every training image I^m scoring will be based on:

$$M(I^m, \theta) = \sum_{i=1}^{n} \delta_i \mid I^m, \beta \mid - \log \Phi(\lambda \delta_i). \tag{3.3}$$

M is the match scoring function and δ_i obtained from $\sum_{n=1}^{M} [I^m, \beta]$ regarding steps selection and Φ is nonlinear function. The logarithmic likelihood relation of exponential model attains from the score of template matching. Vectors of the weight calculate by maximum likelihood technique and are revealed by $\Delta = (\delta_i, i = 1, ..., n)$ (Wu et al., 2010).

$$Max(x, y) = max_{(x,y) \in D} M(I_m, \beta). \tag{3.4}$$

$MAX(x, y)$ calculates the maximum matching score obtained previously. D represents the lattice of I. Here, there is no summation because of updating the size based on training system on frame $(t - 1)$. Moreover, the method tracks the object applying motion feature for getting displacement of moving object.

University of Malaya

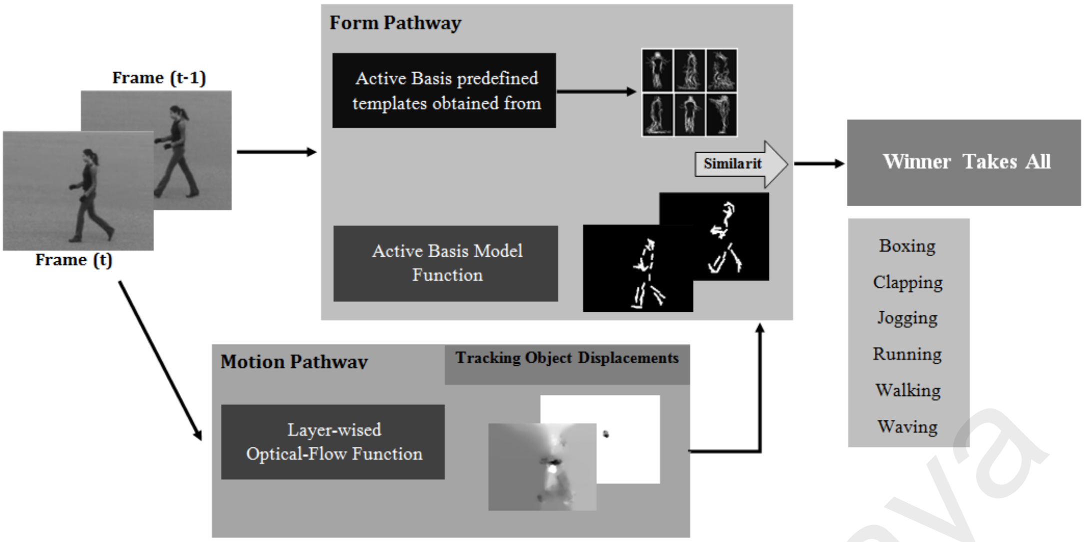

Figure 3.1: The figure reveals the flowchart of our algorithm regarding human action recognition.

3.1.2 Motion Information

For having the features regarding the motion of subject, the layer-wised optical flow estimation has been done. A mask that reveals the each layer visibility is the main difference between estimation of traditional and layer-wised optical flow. The mask shape is able to perform fractal and arbitrary, and only matching applies for the pixels, which fall inside the mask (Liu, 2009). The layer-wised optical flow method by Liu (2009)which has baseline optical flow algorithm of Alvarez et al. (2002), Brox et al. (2004), and Bruhn et al. (2005) is used. M_1 and M_2 are visible masks for two frames $I_1(t)$ and $I_2(t-1)$, the field of flow from I_1 to I_2 and I_2 to I_1 are represent by (u_1, v_1), (u_2, v_2). The following terms will be considered for layer-wise optical flow estimation. Objective function consists of summing three parts, visible layer masks match to these two images using Gaussian filter which called data term matching $E_\gamma^{(i)}$, symmetric $E_\delta^{(i)}$,and smoothness $E_\mu^{(i)}$.

$$E(u_1, v_1, u_2, v_2) = \sum_{i=1}^{2} E_\gamma^{(i)} + \rho E_\delta^{(i)} + \xi E_\mu^{(i)}. \tag{3.5}$$

After optimization of objective function and using outer and inner fixed-point iterations, image warping and coarse to fine search, flow is attained for both bidirectional. Compressed optic flow for all the frames are calculated by straight matching of template to

University of Malaya

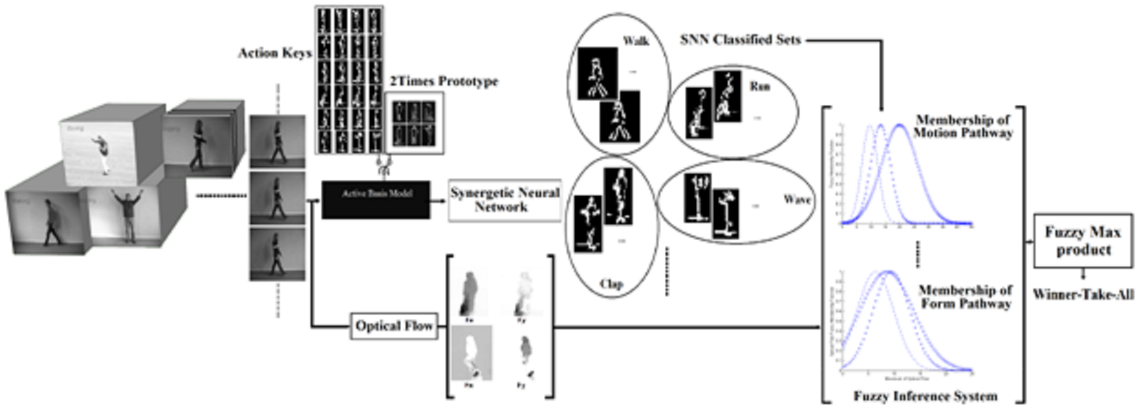

Figure 3.2: The figure reveals the flowchart of modification in the mechanism of biological movement.

the earlier frame by applying the summation of absolute difference (L1-norm). Although optic flow is particularly noisy, no smoothing techniques has been done on it as the field of flow will be blurred in gaps and specially the places that information of motion is significant (Jhuang et al., 2007). To obtain the proper response of the optical flow regarding its application in the proposed model, optical flow will be applied for adjusting the active basis model and making it more efficient. To achieve a representation reliable through the form pathway, the optic flow estimates the velocity and flow direction. The response of the filter based on local matching of velocity and direction will be maximal as these two parameters are continuously changing.

3.1.3 Balanced Synergetic neural network classifier

Synergetic neural network(SNN) is presented by Haken as one the pattern recognition process, which performs in brains of human. A joint method to association of trained samples is valued in feature averaging. Though it is not enough flexible for direction changing, therefore the boundaries of these templates are not clear. Applying learning object in the same view is a technique for dealing with inflexibility, which will limit the task of classification. "Algorithm of melting" is introduced by Hogg et al. (1995) for objects combination in diverse pose. Assume a trained object sample $\acute{I}_i$ contains of n pixel

University of Malaya

values. By reshaping $\acute{I}_i$ to v_i that is a column vector matrix and normalization will have:

$$\sum_{j=1}^{n} v_i j = 0, \sum_{j=1}^{n} v_{ij}^2 = 1. \tag{3.6}$$

Connected prototype matrix V^+ calculates: $V^+ = (V^+V)V(1)$. Let V is the all learned samples set $v_i = 1, ..., m.$ and every column satisfies condition of orthonormality: $v_k^+ v_j = \delta_{ij}$, for all j and k. Where δ_{ij} is delta of Kronecker. For a sample examination q, parameters of order signify test-sampling matching. Class parameter of order for k derives as,$\varepsilon_k = v_k^+, k = 1, ..., m.$ Due to pseudo inverse over-fitting sometime melting fails to generalize the learning. A penalty function presents as Most Probable Optimum Design (MPOD) to improve the generalization and classify face object pose application (Lee & Loo, 2010). Following this modification, the melting combination of similar object patterns into a template is useful for classification. So synergic template is:

$$v_p^+ = E(V^T V + P_1 O + P_2 I)^{-1} V^T. \tag{3.7}$$

I, O, P_1, and P_2 are identity matrix, unitary matrix, and coefficients of penalty. E is an enhanced identity matrix; every element of E is a row vector of size j as the following:

$$E = \begin{bmatrix} e_1^{n(1)} & e_0^{n(2)} & \cdots & e_0^{n(M)} \\ e_0^{n(1)} & e_1^{n(2)} & \cdots & e_0^{n(M)} \\ \cdots & \vdots & \ddots & e_0^{n(M)} \\ e_0^{n(1)} & e_0^{n(2)} & \ldots & e_1^{n(M)} \end{bmatrix} e_0^i = (0, \cdots, 0), e_1^i = (1, \cdots, 1). \tag{3.8}$$

The proposed model uses two times synergetic neural network, once for making the templates in each pathways and second time in the final classification.

3.2 Unbalanced-SNN and Quantum-Behaved PSO and Fuzzy Max-Product for interaction between two pathways

Analyzing the human brain cognitive processes (Hanken, 1991; Haken, 1995; Gao et al., 2001; Haken, 2004), particularly the visual analysis, it shows that the brain per-

University of Malaya

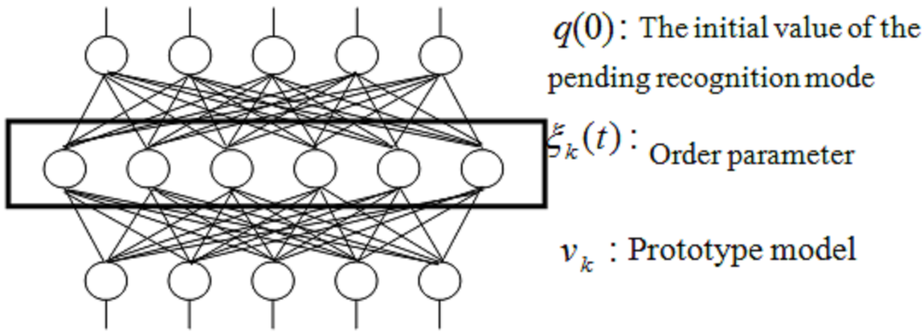

Figure 3.3: The synergistic pattern recognition.

sistently involved in a big amount of the perception re-processing, subconscious mind, filtering, decomposition and synthesis. The brain of the human is a cooperative system, in some cases; cognitive processes can depend on the self-organizing pattern formation. A joint method to association of trained samples is values of feature averaging (Gao et al., 2001). He revealed a collaborative pattern recognition of a top-down thinking: pattern recognition process can be comprehended like a specific order parameter competition process for recognition mode q can construct a dynamic process, so q after middle state $q(t)$ into a prototype pattern v_k. There will be a similarity in the evaluation of some parameters like balanced mode. The kinetic equation based on using q is as follow:

$$\dot{q} = \sum_{k=1}^{M} \lambda_k v_k (v_k^+ q) - B \sum_{\acute{k} \neq k} (v_k^+ q) v_k - C(q^+ q) q + F(t) \tag{3.9}$$

The corresponding kinetic equation for the order parameter

$$\dot{\varepsilon}_k = \lambda_k \varepsilon_k - B \sum_{\acute{k} \neq k} \varepsilon_{\acute{k}}^2 \varepsilon_k - C \sum_{\acute{k}=1} \varepsilon_{\acute{k}}^2 \varepsilon_k \tag{3.10}$$

Based on the competition, the order parameter, which is the strongest, will have a victory, that is, to accomplish the pattern recognition purpose. This idea can be realized through a layer-wised network that is depicted in Figure **??**. Haken (1991) suggested the approach with logarithmic mapping-based, F_T, and followed coordinates transform technique. The supposed algorithms of learning that assign adjoin vector process of prototype vector. Here, two ways are presented regarding assigning prototypes which is utilized synergetic neural networks twice and another one uses key frames of actions for predict-

University of Malaya

ing of actions. Attention parameter is also will be determined using quantum particle swarm optimization technique that will be presented afterward.

3.2.1 Quantum-Behaved Particle Swarm Optimization for kinetic equation of order parameter

Quantum-behaved particle swarm optimization (QPSO) (Sun et al., 2012) is driven by conceptions from quantum mechanics and particle swarm optimization (PSO), an algorithm regarding probabilistic optimization adapted from the bare-bones PSO family (Kennedy, 2010). Like PSO by M individuals, which each of them is considered as a volume-less particle in an N-dimensional space, by the recent position vector and the velocity vector of particle $i, 1 \leq i \leq M$ on the $n_t h$ iteration represented as and correspondingly. The particle moves based on:

$$V^j_{i,n+1} = V^j_{i,n} + c_1 r^i_{i,n}(P^j_{i,n} - X^j_{i,n}) + c_2 R^j_{i,n}(G^j_{i,n} - X^j_{i,n}). \tag{3.11}$$

$$X^j_{i,n+1} = X^j_{i,n} + V^j_{i,n+1}. \tag{3.12}$$

For $j = 1, 2, ..., N$, where c_1 and c_2 are known as the acceleration coefficients. The best earlier position vector of particle i is shown by $P_{i,n} = (P^1_{i,n}, P^2_{i,n}, ..., P^N_{i,n})$ (personal best or pbest), and the position vector of the best particle between whole particles in the population is presented by $G_n = (G^1_n, G^2_n, ..., G^N_n)$ (global best or gbest). Following minimization problem will be considered:

$$\min \quad f(x), \quad s.t. \quad x \in S \subseteq R^N \tag{3.13}$$

Where objective function presents by $f(x)$ which in considers as almost continuous function everywhere in feasible space S. Thus, for updating $P_{i,n}$ we will have

$$P_{i,n} = \begin{cases} X_{i,n} & f(X_{i,n} < f(P_{i,n-1})) \\ P_{i,n-1} & otherwise \end{cases} \tag{3.14}$$

and G_n can be created by $G_n = P_{g,n}$, where $g = argmin_{1 \leq i \leq M} f(P_{i,n})$ PSO algorithm may be converging wherever every particle converges to its local attractor $P_{i,n}$ which are defined as:

$$P^j_{i,n} = \psi^j_{i,n} P^j_{i,n} + (1 - \psi^j_{i,n}) G^j_n, \psi^j_{i,n} \approx U(0,1) \tag{3.15}$$

$\psi^j_{i,n}$) is a sequence of random number between 0, 1, uniformly. Equation shows that stochastic attractor of i^{th} considers in hyper-rectangle and moves by $p_{i,n}$ and G_n. Sun et al. (2012) presented the position of the particle is an updated using equation as follows:

$$X^j_{i,n} = p^j_{i,n} \pm \alpha |X^j_{i,n} - \beta| ln(\frac{1}{u^j_{i,n+1}}) \tag{3.16}$$

where α is the Contraction-Expansion (CE) coefficient which is a positive real number and can be adjusted to balance the global and local search of the algorithm within its process. Random numbers uniformly distributed on $(0,1)$ revealed as sequence is shown by $u^j_{i,n+1}$, varying with n for each i and j. Also the mean best (mbest) position is presented by $C_n = (C^1_n, C^2_n, ..., C^N_n)$ which is the average of the best positions (pbest) of all particles, that is, $c^j_n = \frac{1}{x} \sum^M_{i=1} P^j_{i,n}$.

3.2.2 Centroidal Voronoi Tessellations for Choosing a Starting the attention parameter

In the previous section, it is mentioned that quantum-behaved particle swarm optimization is applied for finding the optimum order parameter. As revealed in the kinetic equation of synergetic neural networks, initialization of the attention parameter (λ_k) is required to calculate the order parameters updates. Voronoi tessellations can be applied as a way to partition a viable space into partitions. The set of generators considers as a group of points in the space, which divided into subsets following the approximation of the generators points. Generators are associated with subsets and points are nearer to its corresponding generators rather than any of other generators considering distance function (*e.g.*, the Lz-norm). Note that the generators are not very evenly distributed throughout the space. By dividing the spaces into the partitions, several generators are set at almost precisely the same point in the space. Although, the centroidal voronoi tessellations sets the generators at centre of the voronoi cells overcomes to the poor and non-uniform distribution of the cells (Richards & Ventura, 2004).

Here, the generators were chosen similar way regarding initialization of initial attention parameters for Particle Swarm Optimization. The proposed approach follows Ju-Du-Gunzburger (JDG) algorithm (Ju et al., 2002) which produces the feasible computational approximation of CVTs and its combining the elements of MacQueen's method (MacQueen, 1967) and Lloyd's method (Du et al., 1999). This algorithm finds the attention parameter initial positions in quantum-behaved particle swarm optimization of order parameter updates more uniformly distributed in the search space.

3.2.3 Combination of two pathways and Max-product fuzzy

The recognition stage schematic regarding classification of human action recognition based on biological inspired model is shown in Figure (**??**) and Figure (1.1) considering the features calculated for both pathways, the main concern is their combination. For that, max product fuzzy method has been utilized for transferring the information of both pathways by Gaussian membership function and maximum of their product into fuzzy domain that represents the class which action does it belongs. Fuzzy logic is a kind of logic having multi-valued that is originated from the theory of fuzzy set found by Zadeh (1965) and it deals with reasoning approximation (Kennedy, 2010). It offers high level framework aimed at reasoning approximation which can suitably provide imprecision and uncertainty together in linguistic semantics, model expert heuristics and handles requisite high level organizing principles (Kumar, 2004). Artificial neural networks refer to computational or mathematical models based on biological neural network and provide self-organizing substrates for presenting information with adaptation capabilities in low level. Fuzzy logic can be a significant complementary method for neural networks because of plausibility and justified for combining the approaches together regarding design classification systems which referred as fuzzy neural network classifier (Kumar, 2004; Lin et al., 2011). Also Bourke and Fisher (1996) presented that the max-product gives better outcomes than the usual max-min operator. Consequently, similar algorithms by having

University of Malaya

effective learning scheme have been mentioned by Bourke and Fisher (1998), Loetamonphong and Fang (1999), and Wong et al. (2010) using the max-product composition later. Here , fuzzy Max-production composition is applied inside the synergetic neural network regarding form and motion pathways aggregation. It means the initial order parameter will be obtained by combination of these two pathways for better decision-making.

3.2.4 Definition of motion pathway classes in different action

All possible action of human object optical flow captured and store in a database considered as references. Each references optical flow data in every action assigns in a specific amount of optical flow regarding specific actions, which will be assigned by interpretation of an operator (human observer) as a training map, generally description of which could be called Operator perceived activity (OPA) (Owens et al., 2002). Considering that mean and standard deviation of every class are different from each other, the operator comments on each of reference data will be different and classification among the classes will be done.

3.2.5 Max-product fuzzy classifier

Fuzzy production among two pathways classification is carried out through general strategy of having result estimated as following composition from both pathways presented as below:

$\mu_{FP\omega}(\dot{\varepsilon}_k, C_i, t)$ and $G_{MP\omega}(f_k, C_i, t)$ are outputs of quaternion correlator in enrollment stage belong to form and motion pathways, respectively. Fuzzification is done through Gaussian membership function as activation functions:

$$G_{MP\omega}(f_k, C_i, t) = exp[-\frac{FP\omega(\dot{\varepsilon}_k, C_i, t) - \mu_{FP\omega}}{\sigma^2}] \tag{3.17}$$

(1) Where $\dot{\varepsilon}_k$ comes from unbalanced order parameter k^{th} subject in frame' index t belongs to C_i estimate from active basis model as form pathway and directly relates to λ_k as its k^{th} attention parameter tuned by quantum-behaved particle swarm optimization in the training stage. Also for motion pathway membership is Gaussian functions deviation

University of Malaya

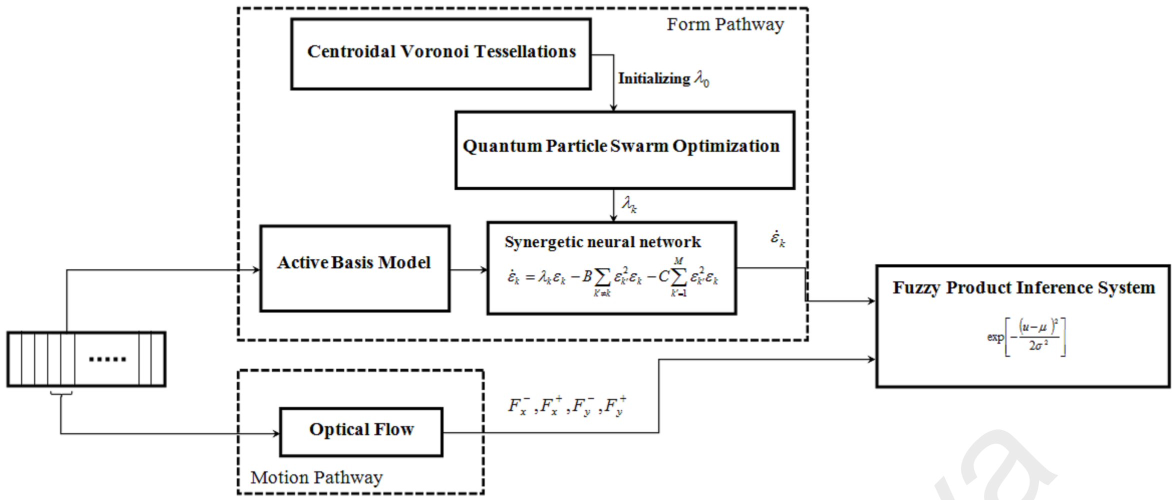

Figure 3.4: Schematic of recognition in proposed model.

as below:

$$G_{MP}(f_k^{\pm}, C_i, t) = exp[-\frac{(MP\omega(f_k^{\pm}, C_i, t) - \mu_{MP\omega})^2}{\sigma^2}]$$

$$G_{MP}(f_k^{\pm}, C_i, t) = G_{MP}(f_k^{-}, C_i, t) \times G_{MP}(f_k^{+}, C_i, t)$$

$$G_{MP}(f_k^{\pm}, C_i, t) = G_{MP}(f_{y,k}^{-}, C_i, t) \times G_{MP}(f_{y,k}^{+}, C_i, t)$$

$$\mu_{MP}(f_{\tau,k}^{\pm}, C_i, t) = G_{MP}(f_{x,k}^{\pm}, C_i, t) \times G_{MP}(f_{y,k}^{\pm}, C_i, t)$$

(3.18)

Where $f_{\tau,k}^{\pm}$ is positive or negative (direction) flow in $\tau = x\, or\, y$ of k^{th} subject in frame's index t as representation of motion pathway amount for every class c_i. $\mu_{FP\omega}(\dot{\varepsilon}_k)$ and $\mu_M P\omega$ are mean value and is standard deviation of both pathway.

(2) Determine the value of product by considering trained attention parameter in

University of Malaya

form pathway and trained parameters of motion pathway in k^{th} subject in frame's index t.

$$\mu_{M\acute{\omega}} = G_{FP\omega}(\dot{\varepsilon}_k, C_i, t) \times \mu_{MP}(f^{\pm}_{\tau,k}, C_i, t) \tag{3.19}$$

(3) Gather the values of product in an array similar for amount of membership in class of every action with both pathways separately:

$$E=\begin{bmatrix} \mu_{P\acute{\omega}_1,C_1} & \mu_{P\acute{\omega}_2,C_2} & \cdots & \mu_{P\acute{\omega}_1,C_i} \\ \mu_{P\acute{\omega}_2,C_1} & \mu_{P\acute{\omega}_2,C_2} & \cdots & \mu_{P\acute{\omega}_2,C_i} \\ & \vdots & \ddots & \vdots \\ \mu_{P\acute{\omega}_k,C_1} & \mu_{P\acute{\omega}_k,C_2} & \cdots & \mu_{P\acute{\omega}_k,C_i} \end{bmatrix} \tag{3.21}$$

(4) Presents output array and a set of produced membership amounts reveals the belonging degrees to every class C_i. The biggest amount represents the degrees of belong to each classes and winner take all.

(5) Determine which element in classification matrix Y_μ has maximum degree of the membership among all i classes.

ψ = number of element position in classification matrix Y_μ which has the maximum value with C_i class. ψ presents the assigned number of reference image in database.

(6) Following one fuzzy IF-THEN rule, perform defuzzification:

R^1_s : IF $\mu_{P,\acute{\omega}_\alpha,C_i}$ from subject α in class has maximum degree in membership function as compare with others, THEN subject classified as class C_i (Yousefi & Loo, 2014b).

University of Malaya

3.3 Fuzzy Optical flow divisions and interaction between two pathways

This method considers stability in the motion pathway regarding the information of motion. Optical flow division helps to have more stable decision of these two pathways, while following the psychological evidence about the influence of the motion information from dorsal processing streams on the form information generated in ventral processing stream. Making this way by fuzzy inference system creates a robustness and resistance of the motion information regarding instantaneous changes within the biological movements. In the experimental results, this part was compared with the proposed method, applying on KTH human action dataset. It provides high level framework targeted at approximation reasoning which can appropriately deliver the imprecision and uncertainty together in linguistic semantics, model expert heuristics, and handles requisite high level organizing principles. Fuzzy logic can be an important balancing method which is plausibility and justified for combining approaches together for design the classification, decision and inference systems (Lin et al., 2011). Different fuzzy inference systems have been proposed and there has been substantial that many of its applications are composition of max–min as functional basis. However, Leotamonphong and Fang (1999) mentioned that composition of max–min is "suitable only when a system allows no compensatory among the elements of a solution vector" (Sun et al., 2012). A time dependent fuzzy system has also been used many times regarding solution of control and classification. Chen et al. (2005) presents a delay-dependent robust fuzzy control for a class of nonlinear delay systems via state feedback (Gorelick et al., 2007).

3.3.1 Problem statement and preliminary

After applying optical flow, the velocity of human object will be considered for both x and y directions. In general, $v_x, v_y \in R^{m\times n}$ which m and n are sizes of image frame from input video stream.

(1) $\mu_{v_x}^{C_{1,2}}(x), \mu_{v_x}^{C_{2,4}}(x), \mu_{v_x}^{C_{1,2}}(y), \mu_{v_x}^{C_{2,4}}(y)$ are triangular membership functions for v_x and it will be the same for v_y velocity in x and y directions and represent outputs of quaternion correlator in the enrollment stage belong to motion pathways, respectively. The fuzzifica-

University of Malaya

tion is done through triangular membership function as activation functions:

$$\mu_{v_x}^{C_{1,2}}(x) = \begin{cases} \frac{x}{C_{1,3}} & 0 < x \leq C_{1,3} \\ \\ \frac{C_{2,4}-x}{C_{1,2}-C_{3,4}} & C_{1,3} < y \leq C_{2,4} \end{cases}$$

,

$$\mu_{v_x}^{C_{2,4}}(x) = \begin{cases} \frac{x-C_{1,3}}{C_{2,4}-C_{1,3}} & C_{1,3} < x \leq C_{2,4} \\ \\ \frac{C_{1,3}-x}{C_{1,3}-C_{2,4}} & C_{2,4} < x \leq n \end{cases}$$

$$\mu_{v_x}^{C_{2,4}}(y) = \begin{cases} \frac{y}{C_{3,4}} & 0 < y \leq C_{3,4} \\ \\ \frac{C_{1,2}-y}{C_{1,2}-C_{3,4}} & C_{3,4} < y \leq C_{1,2} \end{cases} , \mu_{v_x}^{C_{1,2}}(y) = \begin{cases} \frac{y-C_{3,4}}{C_{1,2}-C_{3,4}} & C_{3,4} < x \leq C_{1,2} \\ \\ \frac{m-y}{m-C_{1,2}} & C_{1,2} < y \leq m \end{cases}$$

The position of highest velocity in *x,y* estimated by evaluating the amount of membership functions and then membership function related to every cell will be based on aggregating *x,y* for each velocity. It will evaluate for both cases of velocities. $\mu_{v_x}^{C_i}$ and $\mu_{v_y}^{C_i}$ are shown the membership in each cell where z is number of the cell(i=*1,...,4*).

$$\mu_{v_x}^{C_i} = max\{\mu_{v_x}^{C_i}, \mu_{v_y}^{C_i}\} \qquad (3.24)$$

(2) Determine the value of motion information in motion pathway in frame's index **t**. As information of velocities can be unstable due to shaking the camera or different style in human object, meanwhile it is acting in front of camera, the amount of velocity is dependent of time. The definition of time in this context is based on the frame's index per second. Here, this dependency implements by considering the previous frame membership value. Proposed time dependent fuzzy optical flow division can be utilized for representing a class of optical flow divisions with fuzzy inference rules concerning time regarding every frame of video stream as unit of the time defined here, as follows:

$$\begin{gathered} \widetilde{\mu}_v^{C_i}(t) = \widetilde{\mu}_v^{C_i}(t-\tau) + \eta_v^{C_i}(t)(1-\widetilde{\mu}_v^{C_i}(t-\tau)) \\ t \in [t_0, t_0 + k\tau], k \in (0,1,...,N) \end{gathered} \tag{3.25}$$

Where τ is the frame's index, which is a parameter for camera and k is numbers of frames pasted from the cell changing (it means k will be reset after varying of the cell membership). N is the maximum number of frame distance from present frame, which does not unreasonably increase membership function value. We call $\eta_v^{C_i}(t)$ memory coefficient function and add to the membership function of the winner cell and define as follows:

$$\eta_v^{C_i}(t) = \begin{cases} \frac{1}{k+\beta} & \mu_v^{C_j}(t) \leq \widetilde{\mu}_v^{C_j}(t=\tau) \\ & \\ \frac{-1}{k+\beta} & \mu_v^{C_j}(t) > \widetilde{\mu}_v^{C_j}(t=\tau) \end{cases} , k \geq 0$$

(3.26)

Let β as adjustment parameter can be manually tuned in the system. C_j presents the cell, which is different form C_j and has maximum velocity among all cells in optical flow division. t is the time of frame which one division of the optical flow has the highest membership amount as compare with other divisions and it will be restarted by changing the division.

(3) Gather values produced in previous memberships in every optical flow divisions in each frame by following rules:

1) Flow of upper limb: is attained by association of optical flow fuzzy amounts for C_1 and C_2. Membership value reveals the flow for upper limb of human object. It is mentioned

University of Malaya

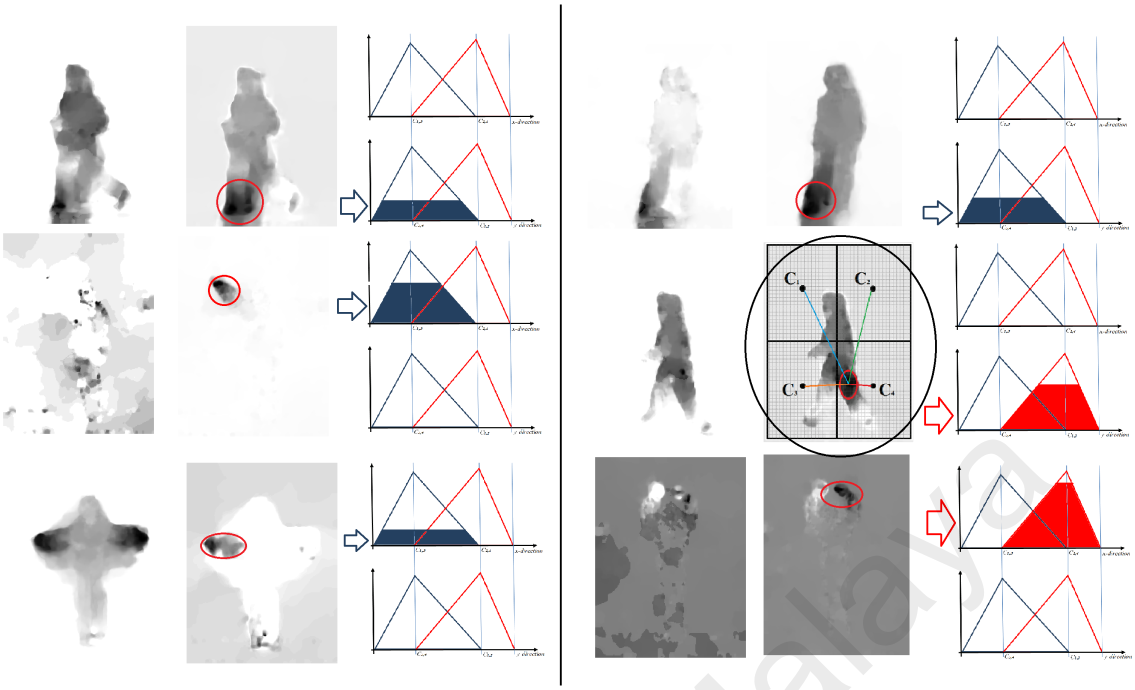

Figure 3.5: The results of dorsal processing stream applying Optical Flow and the Optical Flow division into the fuzzification has been depicted. The resolutions of divisions are designed for categorization of actions group to have additional interfere of dorsal and ventral processing streams. It can be a good representative of the interaction on MT, middle temporal of dorsal stream, and V4, ventral stream(for shape and orientation), or the MST area with inferior temporal(IT). The membership function of the action will be estimated from the position of maximum flow in the flow image. Membership values are aggregated through the proposed technique to increase the robustness. The input images of action mentioned in the figures are obtained from KTH human action recognition dataset.

as follow:

$$\mu_{Upper-Limb}(t) = \widetilde{\mu}_v^{C_1 \cup C_2}(t) = max\{\widetilde{\mu}_v^{C_1}(t), \widetilde{\mu}_v^{C_2}(t)\} \tag{3.27}$$

2) Flow related to lower limb: calculates from union the amounts of optical flow in C_1 and C_2 with each-other in time t:

$$\mu_{Lower-Limb}(t) = \widetilde{\mu}_v^{C_3 \cup C_4}(t) = max\{\widetilde{\mu}_v^{C_3}(t), \widetilde{\mu}_v^{C_4}(t)\} \tag{3.28}$$

Optional: Flow of left and right limb: are calculate by considering the optical flow membership amount among C_1, C_3 and C_2, C_4, respectively.

$$\mu_{Left-Limb}(t) = \widetilde{\mu}_v^{C_1 \cup C_3}(t) = max\{\widetilde{\mu}_v^{C_1}(t), \widetilde{\mu}_v^{C_3}(t)\} \tag{3.29}$$

University of Malaya

$$\mu_{Right-Limb}(t) = \widetilde{\mu}_v^{C_2 \cup C_4}(t) = max\{\widetilde{\mu}_v^{C_2}(t), \widetilde{\mu}_v^{C_4}(t)\} \tag{3.30}$$

This part is optional suggested but we have not used it.

(4) Following one fuzzy IF-THEN rule, perform defuzzification:

R1s: IF every membership function from the subject has maximum degree in membership function as compare with others, THEN the subject limits just some relevant action eligible for form pathway selection and selection possibility of other actions by form pathway eliminates.

(5) Output of aforementioned membership values can be considered as belonging scores among the classes of actions, which shows specific movements in human subject limbs. The biggest amount represents degree of belongs for each classes and winner take all. e.g. running, jogging and walking involve the lower limb activities whereas boxing, clapping and waving make flow in the upper limb of human object (Yousefi & Loo, 2104a; 2014c).

University of Malaya

3.4 Mechanism for the recognition of biological movement using combination of the Fast features along with slowness principle

Slowness features of human object into the ventral stream, form pathway, along with information of the motion pathway, fast features; to modify the original biologically inspired model in visual system (Yousefi & Loo, 2014d; 2015) is considered. In the main experiment, a total of approximately 38,000 frames cuboid of human subject movements were entered into the proposed model. Form and motion, slow and fast, information analyzed though the model. The response of every input was estimate throughout the time-series data applying IncSFA for slowness and its combination with motion, fast, information.

The recognition of biological movements in the original model considered two parallel pathways in the mammalian visual system. The proposed perspective of the current model occurs through slowness and quickness in the ventral and dorsal processing streams. Modeling the ventral stream utilized the slowness principle for the active bases for the extraction of the human object form and motion information that attained the optical flow. The interaction between these two pathways occurs in the categorization, which obtains significant results in the decision and recognition of movements. However, each pathway separately assigns the recognition of biological movements by the considerable disparity rate.

The performances of two patients were analyzed, including DF, who developed visual agnosia (*i.e.*, damage to the ventrolateral occipital), and RV, who developed optic ataxia (*i.e.*, damage to the occipitoparietal cortex) (Goodale et al., 1994). The perspective of this approach considers the original model with a specified mathematical explanation of the system.

3.4.1 Model motivation and Biological Inspired Concept

The model which presented here is for recognition of biological movements and motivated from original biologically inspired models (Giese & Poggio, 2003) by development under perspective of slowness and fastness context within two parallel pathways of mammalian visual system response for a sequence of operations (Figure (3.7)). It

University of Malaya

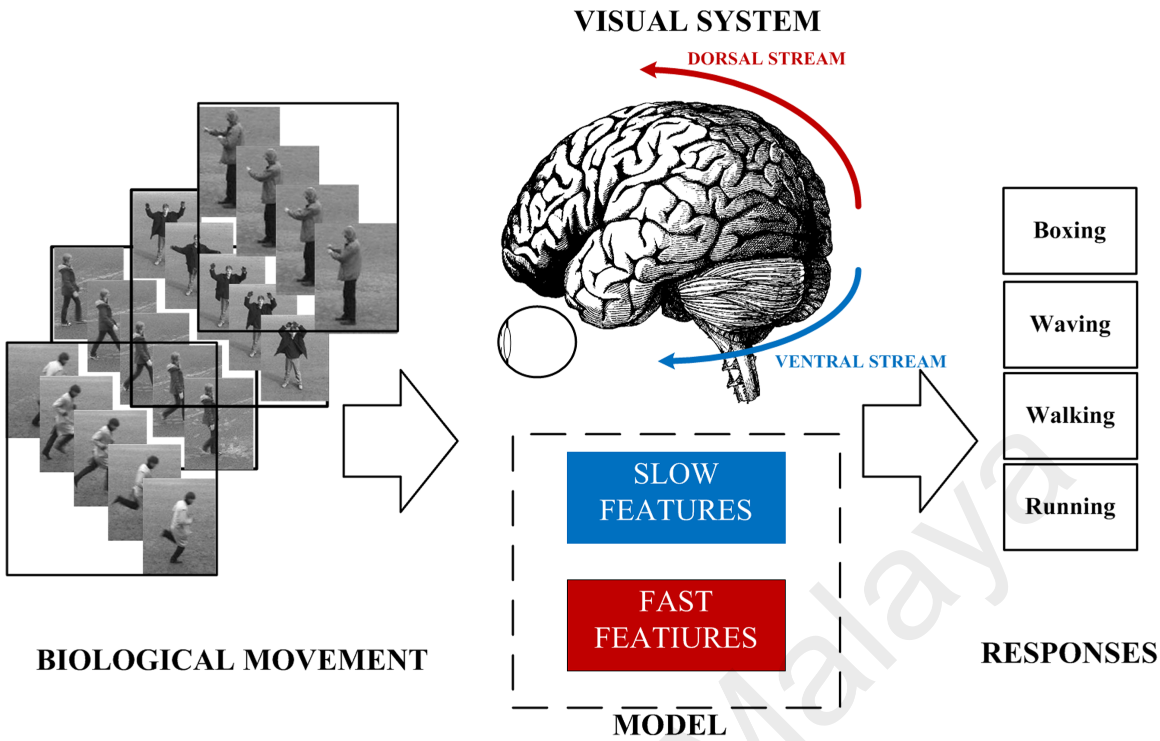

Figure 3.6: Structure overall, visual system analytical models. The approach pursues to develop the computational models for recognition of biological movements and characterize the responses for different actions. This model is the perspective of the original model consists of particular computations of slow and fast feature data. The model can operate in wide range of high-dimension of input and the outcome is a combination of the ventral and dorsal processing stream.

is mostly motivated through prediction by using Active Basis Model (ABM) (Wu et al., 2010) similar with V1-like Gabor filters to the luminance image (in V1) which has normalization of filter outputs, summation of energy which is in contrast (by SUM-MAX in ABM) (Kay et al., 2013). The novel component of this development is utilization of the slowness principle over ventral pathway. This model like the original model has two separated pathways for form and motion information. To motivate and elaborate the concept of temporal features, it seems essential to consider the original model of biological movements which is follow four reliable assumptions by physiological, anatomical information and imaging experiments, and several cortical areas (Giese & Poggio, 2003; Cloutman, 2012; Janssen et al., 2012; Mather et al., 2013). The model splits to two corresponding pre-processing streams (Johansson, 1976; Oram & Perrett, 1996; Thorpe et al., 1996; Riesenhuber & Poggio, 1999) parallel to dorsal and ventral streams which are specified for analysis of motion and shape information, respectively. The proposed approach urges that motion is defined contained by spontaneous temporal variations though biological movement shape is temporally order information and achieved throughout entire movements. Researches have scrutinized the laminar creation profiles and time course of the ventral and dorsal streams have gave several supports that two networks might connect in straight cross-connection at several phases along their pathways, at least inside the visual area (Cloutman, 2012). Modulating the separated pathways processing along with the presence of recurrent feedback loops (Cloutman, 2012)and mutual links have been suggested a recurrent processing loops that permits interaction of top-down and bottom-up processing (Laycock et al., 2007; Cloutman, 2012). However, recent studies are provided better explanations for original model (Janssen et al., 2012). Kruger et al. (2013) presents functional principles in primate visual cortex and relevant biological principles for further advancement in computer vision researches following new findings in neurophysiology. Two streams has used neural detectors for motion and form feature extraction hierarchically allows the in-dependency in size and style in both pathways, and classification of generated features from both feed-forward pathways to categorize the biological movements. The corresponding results on the stationary biological motion recognition revealed that discrimination can be accomplished through particularly small latencies, constructing an important role of top–down unlikely signals (Thorpe et al., 1996). The body shapes are determined by set of patterns like sequences of 'snapshots' (Giese & Poggio, 2003), which has constant feature within whole action episode. The proposed approach expands an earlier model used for the stationary objects (Riesenhuber & Poggio, 1999; Riesenhuber & Poggio, 2002; Giese & Poggio, 2003; Schindler and Van Gool, 2008; Danafar et al., 2010; Somayeh, Alessandro et al., 2010) recognition by adding and combining over the temporal information in pathways. It can be a good relating to quantity tool for organizing, summarizing and interpreting existent information based on the data provided by neurophysiological. The proposed approach develops the original model quantitatively for temporal analysis and even in computer simulations by respect to previous model architecture. Interaction of these two streams is done at few levels in the mammalian brains (Kourtzi & Kanwisher, 2000; Saleem et al., 2000) however many

University of Malaya

neurobiological, physiological and psychological evidences show the slow and fast information coupling occur in many places for instance in STS level (Giese & Vaina, 2001) and in different ways *i.e.* recurrent feedback loops (Cloutman, 2012) and mutual links have been suggested a recurrent processing loops that permits interaction of top-down and bottom-up processing (Laycock et al., 2007; Cloutman, 2012). Although, current neuroscience and psychophysics research specifies that the more extensive form signals, slow features, influence on motion processing, fast features, than previously assumption (Mather et al., 2013). Motion pathway, as representative of dorsal stream in mammalian brain and biologically inspired model, involves information of optical flow, which has fast temporal varying nature. It has consistency with neurophysiological data from neural detectors. Faster varying features due to its achievements within short changes between **Frame(t)** and **Frame(t+1)** rather than whole episode, like ABM based Incremental-SFA in form path. Local detector of optical flow is connected with motion patterns and the model comprises population of neurons four directed neurons in area of MT, however there is a connection between MT and V4 for motion and direction selection. Also, motion edges selectors in two opposite directions that it is found in areas of MT, MSTd, MSTl (Dow et al., 1981; Eifuku & Wurtz, 1998) and many parts of the dorsal steams and probably in the kinetic occipital area (KO) (Giese & Poggio, 2003) also the motion selective edges which can be like MT (Dow et al., 1981) and MSTl (Eifuku & Wurtz, 1998) in a macaque monkey. Few models have been proposed for recognition of human body shape, which is plausible and neurophysiological about recognizing stationary form (for instance in Riesenhuber & Poggio, 1999). The proposed approach follows an object recognition model (Riesenhuber & Poggio, 1999), which is composed of form, features detectors, and make them involve with slowness through ABM based IncSFA. It has reliability and follows the data obtained from neurophysiological information concerning scale, position and sizes invariance, in case of adaptive ABM, which need further computational load along with hierarchy. The methods, which have Gabor, like filters to modeling the detectors have good constancy by simple cells (Jones & Palmer, 1987). The complex-like cells in V1 area or in V2 and V4 are invariant in terms of position varying responses (Giese & Poggio, 2003) and size in-dependency is typically in the area V4. V2, V4 are more selective for difficult form features *e.g.* junctions and corners while are not appropriate for motion recognition because of temporal dependency in these two pathways. The snapshots detectors use to finding the shape models similar with area IT (inferotemporal cortex) of the monkey where the view-tuned neurons located and model of complex shapes tune (Dow et al., 1981). Snapshot neurons are like view-tuned neurons in area IT gives independent scale and position. Previous models used Gaussian Radial Basis functions for modeling and it adjusted in training which performed a key frame regarding training sequences. This modification elaborates that key frame, which is considered efficient, has fast features concept whereas shape of specific human movement is defined in the whole episode and it is independent from fast temporal change. Slowness features, which have Gabor like features, can be a better representative regarding form information of biological movements. A perspective of a model, which follows the original models by utilizing Active Basis Model (ABM), based incSFA is introduced and explained in the method section. The computational simulation along with testing the method is presented in the results section. Finally, concluding remarks that the biological motion perception in human visual system comprises of fast and slow feature association, which makes the recognition of biological movements. For examination of the proposed model on a broader range of high-dimensional video streams, responses are measured to separate parallel pathways of the visual system. Results for an instance patterns model in ventral path are revealed (Figure (3.9)). The proposed model does a sensible job catching the constant pattern of responses of ventral pathway to the human movements (Figure (3.9), upper processing stream). Though, the model does not undervalue the responses of dorsal covering almost half portion of the visual system. The slowness characteristic in the ventral stream has been hidden and its response underestimated in the recognition of biological movement model. It can be clearly indicated through response to different actions. Here, this pattern of slowness features by respect to visual system model regarding application of Gabor like stimuli for the object as an object recognition task throughout the ventral stream. ABM as Gabor based supervised method can boost the responses of

University of Malaya

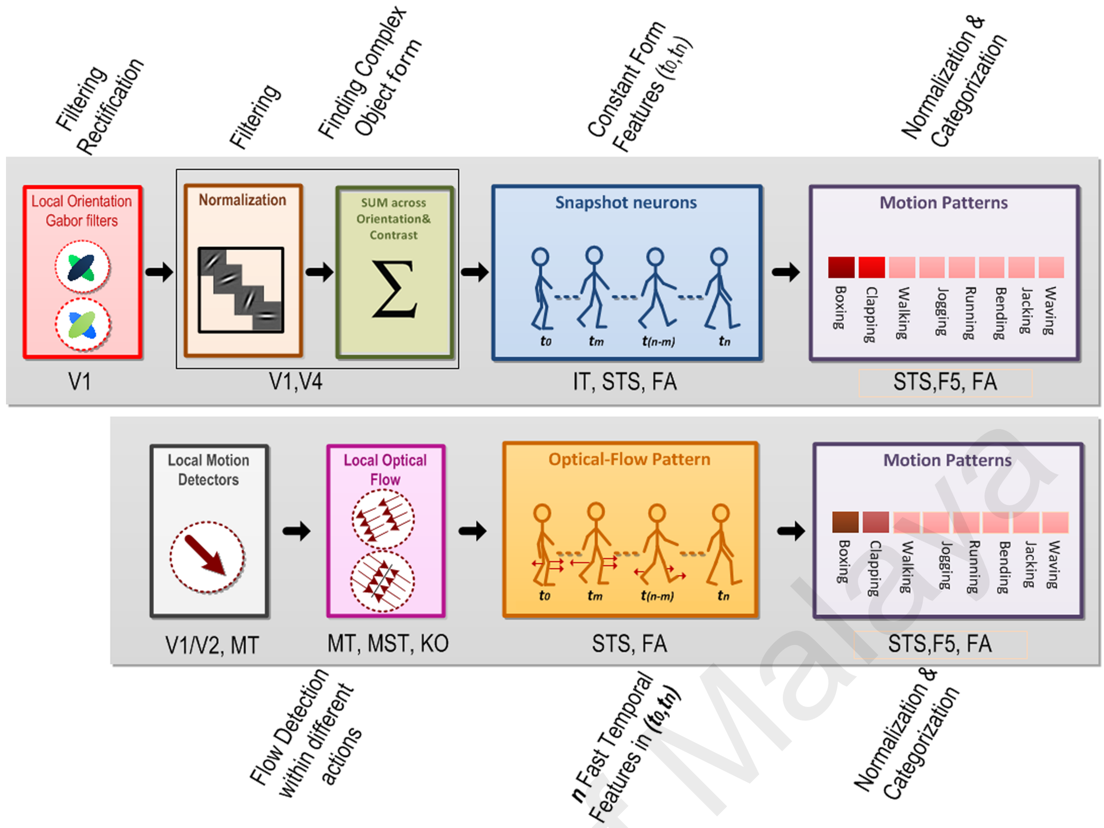

Figure 3.7: The hierarchical model follows the original model and interpretation of the data is in the perspective of combination of slowness and fast features provided from ventral and dorsal processing stream. An overview of these two pathways, form and motion pathways is revealed. Insert depicts the various types of neural detectors in diverse parts of hierarchy. V1 and IT represent primary visual cortex and inferotemporal cortex also KO and STS are kinetic occipital cortex and superior temporal sulcus respectively. These abbreviations along with others indicate visual cortex in monkey and human (Giese & Poggio, 2003).

the stream directive and can be excellent interpreted as providing the human object. The proposed model attempted to increase the performance of the recognition of biological movement model by incorporating of the slowness features with fast features form dorsal stream in the previous and original model (Giese & Poggio, 2003).

Here, this pattern of slowness features by respect to visual system model regarding application of Gabor like stimuli for the object as an object recognition task throughout the ventral stream. ABM as Gabor based supervised method can boost the responses of the stream directive, and can be excellent interpreted as providing the human object. Proposed model attempted to increase the performance of the recognition of biological movement model by incorporating of the slowness features with fast features form dorsal

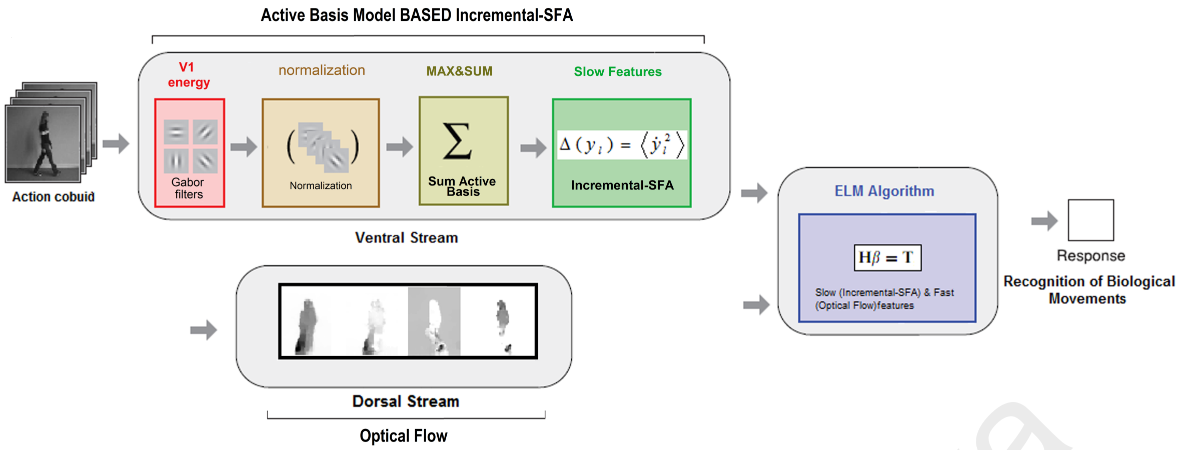

Figure 3.8: The schematic of model is presented here for both pathways. In ventral processing stream, form pathway, a set of Gabor filters have been applied at different orientations, positions and phases; outputs of V1 part is outcomes of quadrature-phase pairs, summed, squared, and square-rooted. Then outputs of the filter normalize considering local population. Afterwards filter outcomes are max pools and summed across space. The MAX, SUM operations are based on the attained active bases of the object form. This initial part of the schematic are done through using ABM (Si et al., 2010) as Gabor based object recognition operation. Finally, the output of the ABM is utilized into the slowness principle method (incremental slow feature analysis) (Kompella et al., 2011; 2012) for extraction of form slow features. On the other hand in the dorsal processing stream, which helps to obtain motion information throughout the high-dimensional input stream. Motion pathway is attained using Optical Flow. Average of these flows within the episode $(t_0, t_1, ..., t_n)$ plays the fast features in this hierarchy which temporally ventral stream requires for utilizing IncSFA for generation slow features. However, each pathways can has its own decision in categorization and it justifies two patients (DF and RV) performances (Goodale et al., 1994).

stream in the previous and original model (Giese & Poggio, 2003).

3.4.2 Theoretical Methods

Based on the biological inspired model complication mentioned in the previous section, it looks difficult to providing analytical statement about the model. In this section the mathematical framework relevant to actual prediction will be introduced. The model will consider a perspective of recognition of biological movement model which is concerns the task throughout two parallel pathways in slowness and fastness theoretical and conceptual outline.

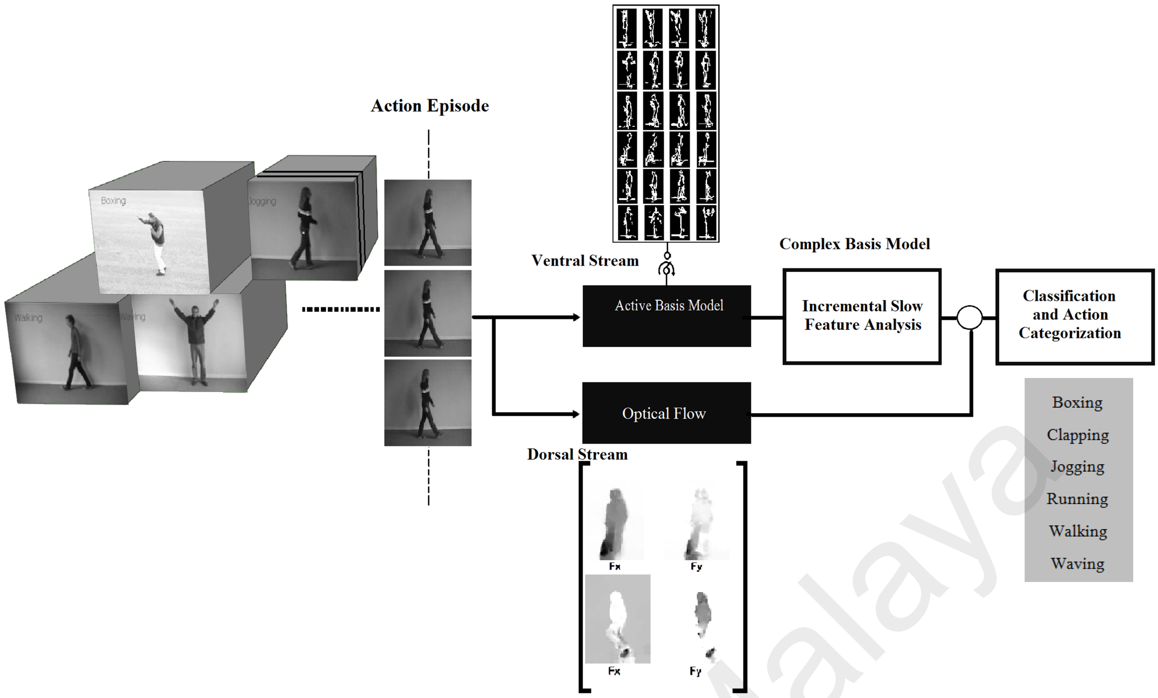

Figure 3.9: Different way to present the hierarchical model in terms of theoretical and computation of the form information, slow features, and motion information, fast features is shown in thirds figure. A supervised Gabor based object recognition method, ABM,gives this property to have human object and computation of the slowness features, performs with IncSFA which is episodic, gives slow form features within the episode that it combination with optical flow information, fast features, creates interaction between the pathways.

3.4.3 Slowest features for ventral processing streams

The perception of slow feature analysis is connected to the hypothesis where the input information (e.g. actions or activities) is included in a 2D signal sequence (e.g. a video) does not vary rapidly, although gradually over time (Liwicki et al., 2013). Whereas, the input signal has normally high difference (e.g. due to variation in environment and different lighting conditions or noise) the separation between informative changes is generally hidden in the rarely changing sequence features. The video attributes, which vary least over time, can be extracted using slowness features. Slowness features are recently entered to the computer vision task (Kompella et al., 2011; Zhang & Tao, 2012; Liwicki et al., 2013) and usually connected to the visual cortex (Wiskott, 2002; Franzius et al., 2007). Incremental learning algorithm is used when applying slow feature analysis for each time step in an unknown video input. Incremental Principle Component Analysis

University of Malaya

(PCA) in closely related to incremental SFA (Olshausen, 1996; Kay et al., 2013), because PCA and Minor Component Analysis (MCA) can solve SFA. Slowness features, which have the information of active basses from multidimensional input along with involving the fast features, can solve the recognition of biological movement task. SFA gives instantaneous scalar input-output functions, which generate signal output (2D signal) that carry the important information and change as slow as possible. Slow Feature Analysis (SFA) is one of the unsupervised learning methods. The functions which plan the input stream to the most slowly changing outcomes are characteristic of a number of elementary representatives of world possessions, summarizing away unrelated details selected up by the sensors (Franzius et al., 2007; Zito et al., 2008; Kompella et al., 2011). Moreover, considering a mobile agent that has high-dimensional video input can be a searching an otherwise stationary room and encode the data by using the combining the situation and direction by slow features (Jolliffe, 1986). SFA is typically concerned with the optimization of complexity, it is common that for the identification of $x(t)$ as input by the D dimension,$x(t) = [x_1(t),...,x_D(t)]$, there is a set of functions similar to $f(x)$ that have L dimension, $g(t) = [g_1(t),...,g_L(t)]$,or that can produce the output for L dimension as $y(t)$ so that $y(t) = [y_1(t),...,y_L(t)]$. Thus, the relationship between these sets is $y_l(t) := g_l(t)$.

$$\Delta_l := \Delta(y_l) := \langle \dot{y}_l^2 \rangle. \quad \textit{is minimal} \tag{3.31}$$

$$\langle y_l \rangle = 0. \quad (Zero\, mean), \tag{3.32}$$

$$\langle y_l^2 \rangle = 1. \quad (Unit\, variance), \tag{3.33}$$

$$\forall d < 1 : \langle y_d, y_l \rangle = 0. \quad (decorrelation\, and\, order), \tag{3.34}$$

These general definitions, similar to **??** and **??**, are the restrictions for having insignificant constants in the output, and **??** is for decorrelation restrictions for features that are the

University of Malaya

same but are not coded. A representation of the evaluation for the derivative of y and the sequential average are considered, correspondingly. The problem is defined by identifying the $f(x)$ for generating the slow varying output. It is noticeable that for the solution of this problem, the optimization of variation calculus is not applicable, but it is predominately straightforward especially for the eigenvector method. Considering that f_l is constrained to be a linear function that consists of a combination of a finite set of nonlinear functions p, the output function will be:

$$y_l(t) = f_l(x(t)) = w_d^T p(x(t)). \tag{3.35}$$

Then, will have $z(t) = p(x(t))$. Based on the changes previously incorporated, the optimization problem will be introduced by minimizing 3.35, the w_l.

$$\Delta(y_l) = \langle \dot{y}_l^2 \rangle = w_l^T \langle \dot{z}\dot{z}^T \rangle w_l. \tag{3.36}$$

If the p functions are selected such that z has a unit of covariance matrix and a zero mean, then the three restrictions will be satisfied if and only if the weight vectors have an orthonormal difference. Whitening is a very common technique that is used for identifying p. For whitening, the principle component of the input data are required; thus, considering the zero mean and the individuality covariance matrix, put the x to z and by this z,the SFA problem will be converted to the linear problem. Equation (6) should be considered for minimizing the L-normed set of eigenvectors of $\langle \dot{z}\dot{z}^T \rangle$. The desired features will be obtained from the set of principle components of $\dot{z}$. The objective was to calculate the temporal slowness, $\delta-$value, features and $g(x)$ as instantaneous functions of the input 2D-signal. This eigenvector-based algorithm is guaranteed to obtain the global optimum and learn biologically plausible rules for the existing optimization problem (Hashimoto, 2003; Franzius et al., 2007; Sprekeler et al., 2007).

The modified optimization problem for the high-dimensional visual input utilizes the information of biological movements and the human object through an ABM as a Gabor-based kernel. Then, this pathway information is combined with fast features by optical

University of Malaya

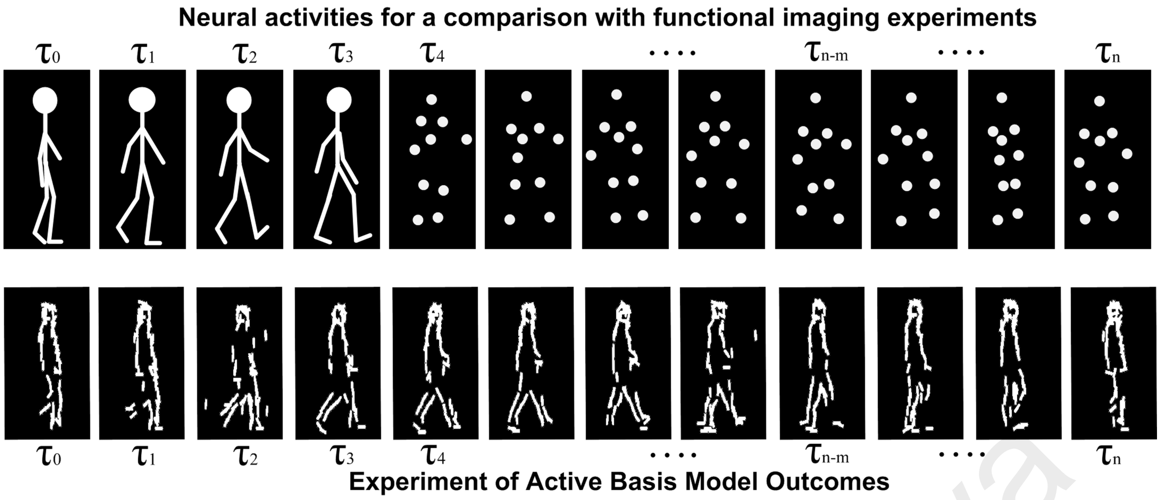

Figure 3.10: Comparison of the functional imaging experiments with the outcome of the ABM regarding features of active basis before generation of the slowness features in the form pathway. The biological movements through the great research experiments of Gunnar Johnsson (1973) regarding ten light bulbs on the joint and recording of the actor within performing complex movements.recognition of the action within episode of the actions. In addition, the dots were spontaneously interpreted as a human. Like points light technique, which presents as static pictures, ABM has very good representative for biological movements which adding it into IncSFA can be very good tool for increasing the ventral pathway in the recognition task.

flow in the motion pathway with respect to the original model (Giese & Poggio, 2003; Schindler & Van Gool, 2008; Danafar et al., 2010; Somayeh et al., 2010).

3.4.4 ABM based IncSFA

The ABM is a supervised learning Gabor wavelet model that has been successfully used for object recognition tasks. It is motivated to apply Olshausen (1996) Field's (1995) representation to model the particular image object category collections. Although the Olshausen (1996) and Field's (1995) model were proposed to provide an explanation of the role of simple cells in the primary visual cortex (V1), Riesenhuber and Poggio's (1999) theory grasps that the local maximum pooling of simple cell responses has been performed in the V1 complex cells. Thus, the local perturbations for the orientations and locations of linear basis elements in the model of Olshausen and Field can be derived to a deformable template from the active basis and, prior to that, the linear basis (Yuille et al., 1992). Riesenhuber and Poggio's (1999) local maximum pooling represents the active basis deforming for the image data explanation. Multiple active bases are used for

University of Malaya

more articulate shape representations being the simplest example of the and-or graph in a compositional framework (Zhu & Mumford, 2007; Wu et al., 2010). Furthermore, the model of Gabor wavelets is very similar to the receptive field profiles of cortical simple cells (Schölkopf et al., 1998). Previously, kernel PCA conquered several restrictions of its linear characteristics by nonlinearly transferring to a space of high-dimensional features from the input space. Kernel PCA derives low-dimensional feature space and is nonlinear in the space of input (Schölkopf et al., 1998). It originates from Cover's theorem regarding pattern separability and represents that in the input space, nonlinear separable patterns are linearly distinguishable with high possibility if the input space is nonlinearly converted to a high dimensional feature space. From the perspective of computation, kernel PCA receives the Mercer equivalence condition benefit as well as feasibility because the inner products in the high dimensional feature space are returned by those in the input space, whereas the complexity of computation is connected to the training sample numbers moderately compared with the feature space dimension (Liu, 2004). Here, the ABM as a subset of the Gabor wavelet kernel for incremental slow feature analysis is introduced. The inference behind this model was motivated by a schematic model of the visual cortex. The set of Gabor wavelet filters on the various phases, orientations, and positions is initially filtered by the input stimulus, the quadrature-phase output are squared, summed and square-rooted (energy of V1) and division normalization and summation occur across orientations (Kay et al., 2013). The ABM has approximately similar operations and a significant consistency with this procedure. Due to the supervised learning object recognition property, it will identify and focus on human objects more precisely and robustly than does the original Gabor wavelet. However, the Gabor wavelet has been successfully used as a kernel and the ABM is one subset of the Gabor kernel. The ABM as a Gabor wavelets-based function was established to analyze images due to their computational and biological significance (Marelja, 1980; Daugman ,1985; Jones & Palmer, 1987; Daugman, 1988). Such kernels are similar to the profile of the 2D receptive field in the mammalian cortical simple and complex cells; the orientation and selectivity exhibit desirable characteristics of spatial locality and are spatially localized in the optimal

University of Malaya

positions and domains of frequency. Previously, Gabor wavelet has been widely utilized as a kernel and for various applications such as in face recognition (Liu, 2004). The Gabor wavelet (kernels filter) has been defined in previous works (Daugman, 1980; Jones & Palmer,,1987; Lades et al., 1993; Liu, 2004):

$$\psi_{\mu,\nu}(z) = \frac{\|\kappa_{\mu,\nu}\|^2}{\sigma^2} e^{\frac{\|\kappa_{\mu,\nu}\|^2\|z\|^2}{2\sigma^2}} e^{i\kappa_{\mu,\nu}z} - e^{-\frac{\sigma^2}{2}}$$

where μ and ν are the orientation and scale, respectively, of the Gabor kernels, $z=(x,y)$,$\|.\|$ is the norm operator, and the wave vector $\kappa_{\mu,\nu}$ is defined as follows:

$$\kappa_{\mu,\nu}, \nu z - e^{\|2} = \kappa_\nu e^{i\phi_\mu} \tag{3.38}$$

where $\kappa_\nu = \kappa_{max}/f^\nu$ and $\phi_\mu = \pi\mu/8$. κ_{max} is the frequency of the maximum, and f is the spacing factor between kernels in the frequency domain (Moghaddam, 2002). The active basis model is a supervised learning Gabor wavelet, and it is considered a kernel that is related to the bases within it. Unlike the Gabor wavelet kernel that was required to define the scales, the orientations and pixels in the ABM, these parameters and attained in a way that is dependent upon the training. The ABM represents the image obtained by the summation of the active base families, which are obtained through the Gabor wavelet dictionary and match scoring function. Let $I(x,y)$ be the gray level distribution of an image; the image convolution I and a Gabor kernel $\psi_{\mu\nu}$ are defined as follows:

$$B_{z_i,\mu,\nu} = \psi_{\mu,\nu}(Z) * I(Z) \quad i = 1,2,...,n \tag{3.39}$$

where $z=(x,y)$ denotes the convolution operator, and $B_{z_i,\mu,\nu}$ is the active base that corresponds to match scoring at the proper orientation and scale. Consequently, the set $S = \{B_{z_i,\mu,\nu} : \mu \in M, \nu \in O\}$forms the Gabor wavelet representation of the image $I(Z)$ along with M and Owhich represent the orientations and scales of the Gabor wavelet dictionary. To include the various spatial localities, spatial frequencies (scales) and orien-

University of Malaya

tation selectivity, concentration is done on all depiction results to obtain a supplemented feature vector X. X is defined as a set of active bases that have the highest matching scores based on the training sets that were used together to make the object form. This method prefers the integration of simple cells to make complex cells.

3.4.5 ABM based Slow Feature Analysis

A method for nonlinearity utilizes the fact that SFA is solved by two-fold PCA and entirely based on second-order statistics. Therefore, SFA is capable of being kerneled in line with the extension of PCA to Kernel-PCA by Schlkopf et al. (1998), thus in the case of incremental SFA, kernelled incremental PCA must be considered (Nickisch, 2006). The presentation and implementation of a kernel based in the principle of temporal slowness has been performed by Bray and Martinez (2003) using the Stone (1996) objective function, which was in some ways not similar to SFA. Incremental SFA is needed, while SFA is used for every time step. As the SFA solution can be reached through PCA and Minor Components Analysis (MCA), it is closely relevant to incremental PCA (Levy & Lindenbaum, 1998; Ross et al., 2008; Kompella et al., 2011).

3.4.6 Motion information from dorsal pathway

In the motion pathway, biological movements are recognized by patterns of optical flow. The optical flow identifies the movement pattern, which is consistent with the neurophysiological information from the hierarchy of neural detectors. In the MT and V1 areas, there are some neurons for motion and direction selection, respectively, in the first level of the motion pathway. For the motion of the subject, the layer-wise optical flow estimation has been utilized. A mask that reveals each layer's visibility is the main difference between the estimation of traditional and layer-wise optical flow. The mask shape is able be fractal and arbitrary, and matching only applies for the pixels that fall inside the mask (Liu, 2009). The layer-wise optical flow method in (Liu, 2009; Liwicki et al., 2013) is used which has a previously described baseline optical flow algorithm (Oram & Perrett, 1996; Thorpe et al., 1996; Mather et al., 2013). After optimization of the objective function, the use of outer and inner fixed-point iterations, image warping

University of Malaya

and a coarse to fine search, bidirectional flow is obtained (more explanation mentioned in section 3.1.2). Compressed optic flow for all frames was calculated by straight matching the template to the earlier frame by applying the summation of the absolute difference ($L1 - norm$). Although optic flow is particularly noisy, lesser smoothing techniques have been performed (one of the good one is pyramid) with it because the field of flow will be blurred in gaps and, in particular, in the locations where information regarding motion is significant (Kompella et al., 2011). To obtain the proper response of the optical flow with regard to its application in the proposed model, the optical flow is applied to adjust the ABM and increase its efficiency. To achieve a reliable representation through the form pathway, the optic flow estimates the velocity and flow direction. The response of the filter based on the local matching of velocity and direction will be maximal as these two parameters are continuously changing.

3.4.7 Extreme Learning Machine (ELM)

Neural networks have been widely utilized in several research areas because of their capability to estimate difficult nonlinear mappings straight from the input sample as well as offering models for a large class of artificial and natural phenomena that are problematic to model via classical parametric techniques. Recently, Huang and his team (Huang et al., 2004; Wang & Huang, 2005; Huang et al., 2006) presented a novel algorithm for learning, i.e., a single layer feed-forward neural network structural design named Extreme Learning Machine (ELM). ELM solves the problems initiated through algorithms that use gradient descent, e.g., the back-propagation used in ANNs. ELM considerably diminishes the time quantity required for training in the neural network and has greatly enhanced faster learning and generalization performance (these are the reasons that it has been used for our approaches). It requires fewer human interventions and can run significantly faster than conventional techniques. It routinely concludes the parameters of the entire network, which evades unimportant external interventions by humans and is more effective in real-time and applications. Several advantages of ELM include the simplicity of usage, quicker speed of learning, greater generalization performance, appropriateness for several nonlinear kernel functions, and activation function (Rajesh & Prakash, 2011).

University of Malaya

The Single Hidden Layer Feed-forward Neural Network (SLFN) function with hidden nodes (Huang et al., 2006; Liang et al., 2006) can be shown by mathematical explanation of the SLFN, which integrates additive and Sigmoid hidden nodes together in a joined method as follows.

$$f_L(x) = \sum_{i=1}^{L} \beta_i G(s_1, b_i, x) \quad x \in \Re^n, a_i \in \Re^n \tag{3.40}$$

Let a_i and b_i represent the parameters of learning in hidden nodes and β_i represent the connecting weight of the i^{th} output node of the hidden node. $G(s_1, b_i, x)$ is the output of the i^{th} hidden node with respect to the input x. For the additive hidden node with the activation function $G(x) : \Re \to \Re$ (*e.g.* sigmoid and threshold), $G(s_1, b_i, x)$ is provided by

$$G(a_i, b_i, x) = g(a_i, x + b_i) \quad b_i \in \Re \tag{3.41}$$

Let a_i represent the connecting weight vector of the input layer to the i^{th} hidden node and b_i represent the i^{th} hidden node bias. For N, arbitrary different examples are indicated by $(x_i, t_i) \in \Re^n \times \Re^m$. Now, x_i is a n vector of contribution, and t_i in a m vector of target. If the SLFN by L hidden nodes can be estimated, then these N samples have zero error. This relationship infers a β_i, a_i and b_i such that

$$f_L(x_j) = \sum_{i=1}^{L} \beta_i G(a_i, b_i, x) \quad j = 1, 2, ..., N. \tag{3.42}$$

The equation above is described in a compacted form as follows:

$$H\beta = T \tag{3.43}$$

where

$$H(\hat{a}, \hat{b}, \hat{x}) = \begin{bmatrix} G(a_1, b_1, x_1) & G(a_L, b_L, x_1) \\ G(a_1, b_1, x_N) & G(a_L, b_L, x_N) \end{bmatrix}_{N \times L} \tag{3.44}$$

University of Malaya

with $\hat{a} = a_1, ..., a_L$; $\hat{b} = b_1, ..., b_L$; $\hat{x} = x_1, ..., x_N$.

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{bmatrix}_{L\times m} \qquad T = \begin{bmatrix} t_1^T \\ \vdots \\ t_L^T \end{bmatrix}_{N\times m} \tag{3.45}$$

Let H represent the hidden layer of the SLFN output matrix, with the i^{th} column of H being the i^{th} hidden node's output with respect to the inputs $x_1, x_2, ..., x_N$.

3.5 Chapter summary

The methods discussed here were based on the stated problems mentioned in the literature. First, the ABM employed into the ventral processing stream as a supervised Gabor based function and increases the robustness of the form pathway in terms of object finding. Secondly, the optical flow division also increases the robustness in motion information along with updating the interaction between the pathways. And finally, slow Feature analysis provides the slowest features which enhances the recognition ability in form information path independent to the style of different actions (it usually occurs among the various people doing same action).

University of Malaya

CHAPTER 4:

Evaluations and Results

This chapter contends with details of implementation and empirical considerations. It addresses the details on the various approaches that have been conducted throughout this research, along with the modifications present in the original model of biological movements. First, the application of the ABM as a supervised Gabor-based object recognition technique in the ventral processing stream is introduced. Then, the modification outcomes of dorsal processing stream using fuzzy optical flow divisions are presented. Finally, results from the combination of the ABM-based slowness principle with fast features are used for the interaction of these two pathways to recognize biological movements. These three main configurations follow the original model for recognition of biological movements in mammalian visual systems.

University of Malaya

4.1 Results of the ABM in the ventral processing stream and balanced/unbalanced synergetic neural network

The experimental results of the involvement of the ABM in the ventral processing stream are presented (Yousefi et al., 2013; Yousefi & Loo, 2014b). The results are extensively presented to reveal the efficiency and estimate the ability of the proposed model to recognize human action.

4.1.1 Biologically Inspired Model and Relation to existing methods

Co-operation among information attained from two processing streams occurs at several levels in the mammalian brain (Kourtzi & Kanwisher, 2000; Saleem et al., 2000), simplifies the aggregation of model (for instance in the STS level (Giese & Vaina, 2001)), and improves recognition performance. Holonomical features consider both pathways for predefined action templates. In the form pathway, the proposed approach followed Karl Pribram's holonomic theory, which is based on evidence that dendritic receptive fields in sensory cortexes are described mathematically by Gabor functions (Pribram) that are vastly utilized by the ABM (Wu et al., 2010). As previously mentioned, the primary stage includes local (in V1 cell) and model detectors (Gabor-like filters) in 16 (including eight preferred) orientations, and the proper scale depends on the receptive field (Dow et al., 1981; Riesenhuber & Poggio, 2002). The ABM also served as snapshot detectors for human body shape models to determine the area IT (inferotemporal cortex) of monkeys where view-tuned neurons are located and the model of complex shapes are tuned (Logothetis et al., 1995); these processes are implemented by applying the synergetic neural network.

In unbalanced synergetic neural networks, tuning optimized attention parameters works as view-tuned neurons in area IT and snapshot neurons to provide independency in scale and position. The proposed model is adjusted through training as key frames. Utilizing optical flow outcome and inferring it with information obtained from the form pathway, the presented approach covers high-level integration of snapshot neurons to obtain information on motion pattern neurons. Furthermore, ABM uses computational mechanism on recognized human object form, which follows up the neurobiological model

University of Malaya

in dorsal stream located in the visual cortex (V1) (Giese & Poggio, 2003; Jhuang et al., 2007; Schindler & Van Gool, 2008). Local direction has been organized in the initial level of form pathway and Gabor-like modeling detector methods, that is, the ABM constantly models cells in the mentioned areas (Jones and Palmer, 1987). Sixteen directions and two spatial scales were obtained from two differentiators, and information on local direction in the pathway and complex-like cells with independent form features appropriate for form pathways was gathered using the proposed neurophysiological plausible model. In the motion pathway, biological movement is consistent with the neurophysiological information of neural detectors in MT and V1, where motion and direction were determined by applying optical flow (Giese & Poggio, 2003). Estimation of local motion is also directly computed from optical flow in response to motion-selective neurons in areas of MT. In the MT areas, MSTd, MSTl, other parts of the dorsal streams, and probably the kinetic occipital (KO) area of motion were selected by opposite directions (Giese & Poggio, 2003) that were modeled by $Fx^-, Fy^-, Fx^+, andFy^+$. The maximum pooling motion and its amount in Gaussian membership function for each direction of optical flow were considered. The product decision membership can be an excellent presenter for form pathway and third-level motion pathway by snapshot neurons. This membership is also a good combination of two pathways for the vertical stream in V2, V4 projection, and primary visual cortex (V1), all of which have been reduced here. The proposed model in the current techniques follows hierarchical feed-forward designs such as those of Jhuang et al. (2007). In particular, a model that follows neurobiological motion processing in the visual cortex was developed and basically follows that of Giese and Poggio (2003). The object recognition task in the form pathway has been changed within the researchers' work from spatiotemporal features (Jhuang et al., 2007) and original Gabor filter (Schindler & Van Gool, 2008) to the model using ABM. However, ABM has the basic characteristic of previous features and basically uses Gabor wavelet while reducing matching operation. ABM is activated by the limited clutters and ensures the important amount of points of interest, which falls on the person subject. In the aspect of used features, layer-wise optical flow (Liu, 2009) which is simply silhouette form regarding motion and form of subject and better combi-

University of Malaya

Figure 4.1: The figure depicts Weizmann and KTH human action datasets. To test the recognition of biological movements, two well-known human action recognition datasets were utilized. Here, the left set of image samples demonstrates actions from the Weizmann dataset; the second set, right-hand side, shows the KTH human action dataset. It is noticeable that the KTH dataset is one of the largest human action datasets, including six various human actions in four different scenarios.

nation of two pathways using fuzzy inference theory and classifying by synergetic neural network tuned by quantum particle swarm optimization that it makes the model more biological.

4.1.2 Data Sets

The KTH action dataset (Liu, 2009) is the largest human action dataset that includes 598 action sequences and comprises six types of single-person actions, such as boxing, clapping, jogging, running, walking, and waving. A total of 25 people performed the actions under different conditions, namely, outdoors (s1), outdoors with scale variation (s2), outdoors with different clothes (s3), and indoors with lighting variation (s4). In this study, the sequence resolutions were downsampled to 200×142 pixels. For this approach, we used five random cases (subjects) for training and developing predefined form and motion templates. As previously mentioned in the literature, KTH is a robust intra-subject variation with a large set; a camera with some shacking is used for taking videos during the preparation, which renders working with this database difficult. Moreover, four independent, separately trained, and tested scenarios are involved (i.e., four visually different databases, all of which share the same classes). Both alternatives have been tested. To consider the symmetry problem of human actions, a mirror function for sequences along the vertical axis is available for testing and training sets. Here, all possible overlaps of human actions within the training and testing sets have been considered (e.g., one video

University of Malaya

has 32 and 24 action frames). The Weizmann human action database (Gorelick et al., 2007) comprises nine types of single person actions, and 83 video streams reveal nine actions, namely, running, galloping sideways, jumping in place on two legs, walking, jumping, jack, jumping forward on two legs, waving one hand, waving two hands, and bending. Tracking and stabilizing figures were obtained using background subtraction masks that come with this data set. Sample frames of this data set are shown in Figure 4.1. The above-mentioned data sets have been widely utilized to estimate methods/techniques designed in action recognition. However, the dataset only concentrates on recognizing single-person actions, such as clapping and walking. To understand the advantages of the proposed approach in the testing data sets, the experimental results were illustrated using synergetic neural network in balanced and unbalanced modes; moreover, previous studies that proposed biological human action models were compared. The balanced and unbalanced classification of form pathway was compared, along with the accuracy of form and motion pathways after the application of fuzzy product between these two pathways. The proposed model is efficient, and computational cost will cover feature extraction of two pathways with form and motion features that apply ABM and optical flow, respectively. After optimizing the tune of attention parameter in unbalanced synergetic neural network for form pathway, the system inference of a new video only requires several seconds in an un-optimized MATLAB implementation, which is combined using existing codes for motion and form pathway in MATLAB/C (Liu, 2009; Wu et al., 2010). Subsequently, the system was trained and tested correspondingly by computing features in both pathways with different settings as previously mentioned. For a specified test sequence, the action label was assigned to the action frames. Then, the accuracy of classification was specified by

$$Acc = \frac{Numbers\,of\,correct\,classified\,frames}{Total\,number\,of\,frames} \tag{4.1}$$

The algorithm correctly classifies most of the actions (the confusion matrices are shown in Figure 4.3). Most of the mistakes are on the recognition of running, jogging, boxing, clapping, and waving. The intuitive reason for this is the similarity between these

University of Malaya

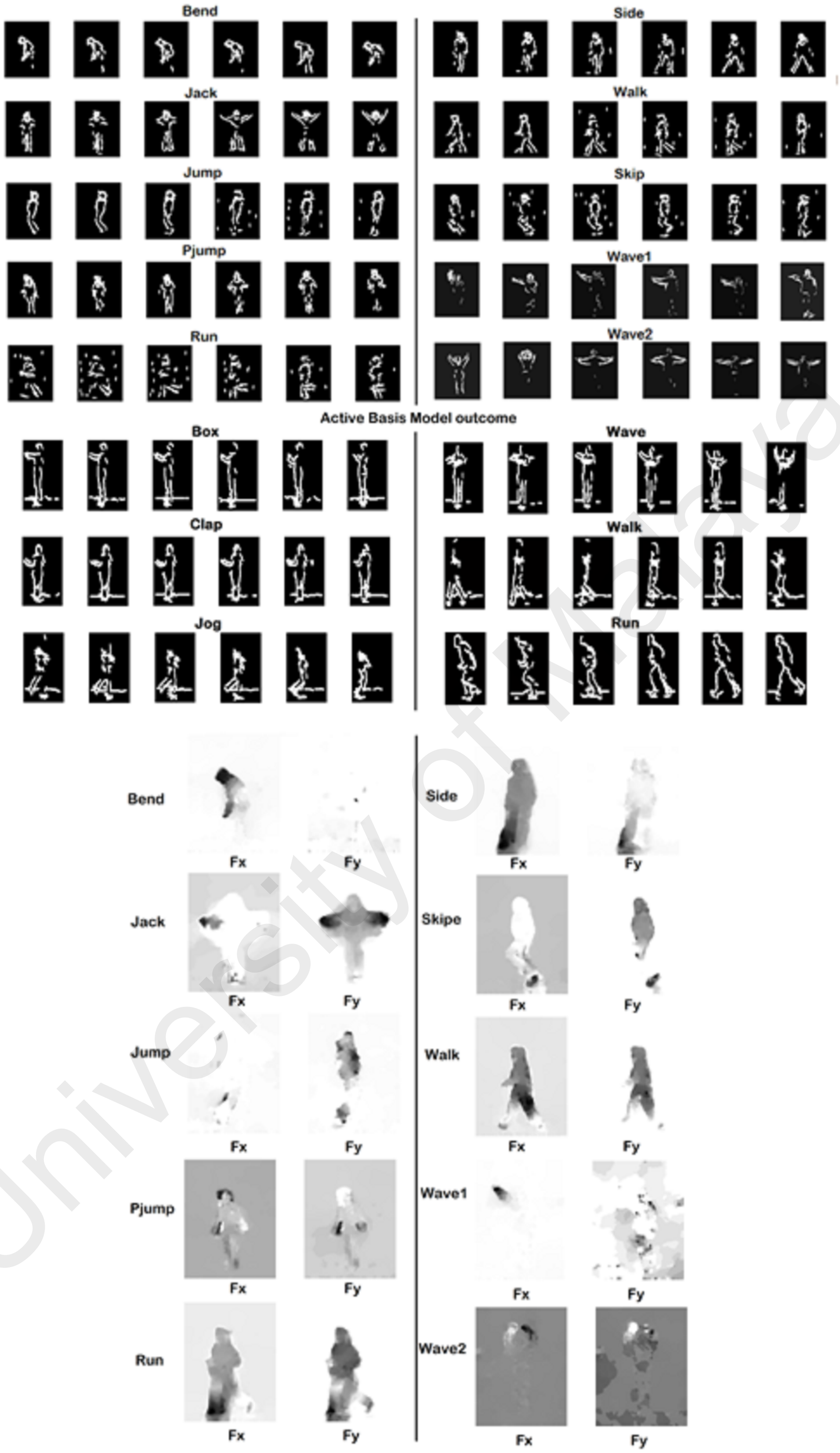

Figure 4.2: Figure represents the outcomes of the different pathways for every action (upper set is ABM-ventral stream and lower set is optical flow-dorsal stream).

University of Malaya

	Bend	Jack	Jump	Pjump	Run	Side	Skip	Walk	Wave1	Wave2
Bend	0.8	0	0	0	0	0	0	0	0.2	0
Jack	0	0.8	0	0	0	0	0	0	0.1	0.1
Jump	0	0.2	0.6	0.1	0	0	0	0	0.1	0
Pjump	0	0.1	0	0.9	0	0	0	0	0	0
Run	0	0.24	0	0.12	0.50	0	0	0.12	0	0
Side	0	0	0	0	0	0.62	0.25	0	0.12	0
Skip	0	0	0.12	0.12	0.12	0	0.50	0	0.12	0
Walk	0	0.22	0	0	0	0	0.11	0.56	0	0
Wave1	0	0	0	0	0	0	0	0	0.67	0.33
Wave2	0	0	0	0	0	0	0	0	0	1.00

	Boxing	Clapping	Jogging	Running	Walking	Waving
Boxing	0.95	0.02	0	0	0	0.03
Clapping	0.16	0.69	0	0	0	0.15
Jogging	0	0	0.66	0.20	0.14	0
Running	0	0	0.10	0.88	0.02	0
Walking	0	0	0.26	0.05	0.70	0
Waving	0.21	0.04	0	0	0	0.74

Figure 4.3: Confusion matrices representing the accuracy of recognition in KTH and Weizmann data set using multi-prototype human action templates.

two groups of action. Upon testing the databases, the confusion matrices were obtained for two proposed scenarios regarding the application of the methods and the overall accuracy of both per-fame and per-video classification. The confusion matrices (per-video or per-frame) of the proposed scenarios have similar patterns; hence, only one confusion matrix is revealed for every dataset. The results of each scenario are mentioned in Table 4.1, which shows the accuracy of the proposed techniques compared with those of previous methods with the same data set. However, this comparison is not precise because of differences in experimental setups, and the presented results are comparable with state-of-the-art techniques. Moreover, considering various methods results in various differences in their setups, such as un-supervision or supervision, with or without tracking, subtraction of the background, and multiple action recognition. In terms of biology, movement contains corticofugal pathways from both peristriate cortex (V2) and striate cortex (V1). The peristriate (V2) and striate (V1) cortices are mutually linked with minor, although important, differences in their receptive properties. In holonomic brain theory, the peristriate (V2) and straite (V1) are a narrowly coupled collaborating system by virtue of both reciprocal cortico-connectivity and connection to the brain stem's tectal region. Upon this carefully joined organism, the additional compound perceptual procedures converge. The convergence locus is the region of the brain stem tectal close to the colliculi. The supe-

University of Malaya

	Bend	Jack	Jump	Pjump	Run	Side	Skip	Walk	Wave1	Wave2
Bend	1.00	0	0	0	0	0	0	0	0	0
Jack	0	0.89	0	0	0	0	0	0.11	0	0
Jump	0.11	0	0.89	0	0	0	0	0	0	0
Pjump	0	0	0	1.00	0	0	0	0	0	0
Run	0	0	0	0	0.75	0	0.12	0.12	0	0
Side	0.13	0	0	0	0	0.75	0	0	0.12	0
Skip	0.23	0.11	0	0	0	0	0.56	0	0	0
Walk	0	0	0.23	0	0	0	0.11	0.56	0	0
Wave1	0	0	0	0	0	0	0	0	1.00	0
Wave2	0	0	0	0	0	0	0	0	0.33	0.67

	Boxing	Clapping	Jogging	Running	Walking	Waving
Boxing	0.89	0	0	0	0	0.11
Clapping	0.05	0.83	0	0	0	0.11
Jogging	0	0	0.79	0.08	0.12	0
Running	0	0	0.05	0.95	0	0
Walking	0	0	0.17	0	0.82	0
Waving	0.14	0.14	0	0	0	0.72

Figure 4.4: Confusion matrices for recognition of human action in KTH and Weizmann data sets applying second scenario.

rior colliculus connections to neurons in the striate cortex (V1) were visualized, showing complex receptive fields that complete the circuit (Stone 1983). Moreover, a set of receptive fields is particularly sensitive to processing movement in the visual input; in specific, the virtual movement of one portion of input with respect to another. This sensitivity to relative movement is critical to the formation of object-centered spaces. Another set is principally sensitive to comparative movement among somatosensory and visual inputs. The receptive fields of these neurons directly comprise the formation of egocentric action spaces (Pribram, 1991). Considering the aforementioned approach in terms of biology, the proposed model considered two structures for V1 information of form pathways to determine the shape and form of human objects by incorporating the original frame after ABM application; in the end, the two configurations were compared.

In V2, the proposed method used local representation and action sequence is selected by its location. The response of ABM is directly used for classification of action.

4.1.3 Multi-prototype human action Templates

In this scenario, the recognition of human action pattern in the form pathway was carried out using one predefined template, which was obtained by applying synergetic neural network prototypes. First, multi-prototype predefined templates were used for each hu-

University of Malaya

man action, which were obtained by applying synergetic neural networks on human action images. To develop a training map for every action, the human action sequences comprising the five primitive movements are divided. A whole action sequence can be created using these five basic actions. In addition, considering the style invariance difficulties of diverse objects in the same action, the proposed training map was obtained using five subjects from targeted human action databases. To simplify the explanation, five snippets in different actions $A_1 - A_5$ and each subject from targeted database $D_1 - D_5$ were considered. First, the synergetic neural network is applied to A_1 in $D_1 - D_5$, and the outcome shown by P_1 serves as the first prototype obtained from first action snippet. The number of prototypes can be completed by applying the synergetic neural network and calculating the residual prototypes called $P_1 - P_5$. Calculated prototype images that consider style invariance represent one action within five snapshots. Afterward, these prototypes melt together using second time synergetic neural network to achieve the final prototypes, each of which represents the specific action within different action snippets and considers style invariance property. Let F_t represent the outcome of melting $P_1 - P_5$ in a specified action. The final prototype images for each human action and the application of synergetic neural network procedure to synthesize a training map is presented in Figure 4.5. The recognition result of the first scenario is revealed in Figure 4.3 (the second scenario showed in Figure 4.4). Two categories use dissimilar paradigms that cannot be directly compared. Here, the experimental result of the proposed approach is presented. The KTH and Weizmann human action databases have been previously used to benchmark the accuracy and consistency of sets of experiments (Schuldt et al., 2004; Jhuang et al., 2007; Niebles et al., 2008; Schindler & Van Gool, 2008; Wang & Mori, 2009; Zhang & Tao, 2012). Thus, a set of training maps was provided, and the proposed technique was tested on the entire data set in which the four video scenarios were combined (for KTH data set). The data set was divided into a set of training-maps with five randomly selected subjects and a test part by residual subjects. Afterward, the performance average of five random splits following their frame numbers was measured.

University of Malaya

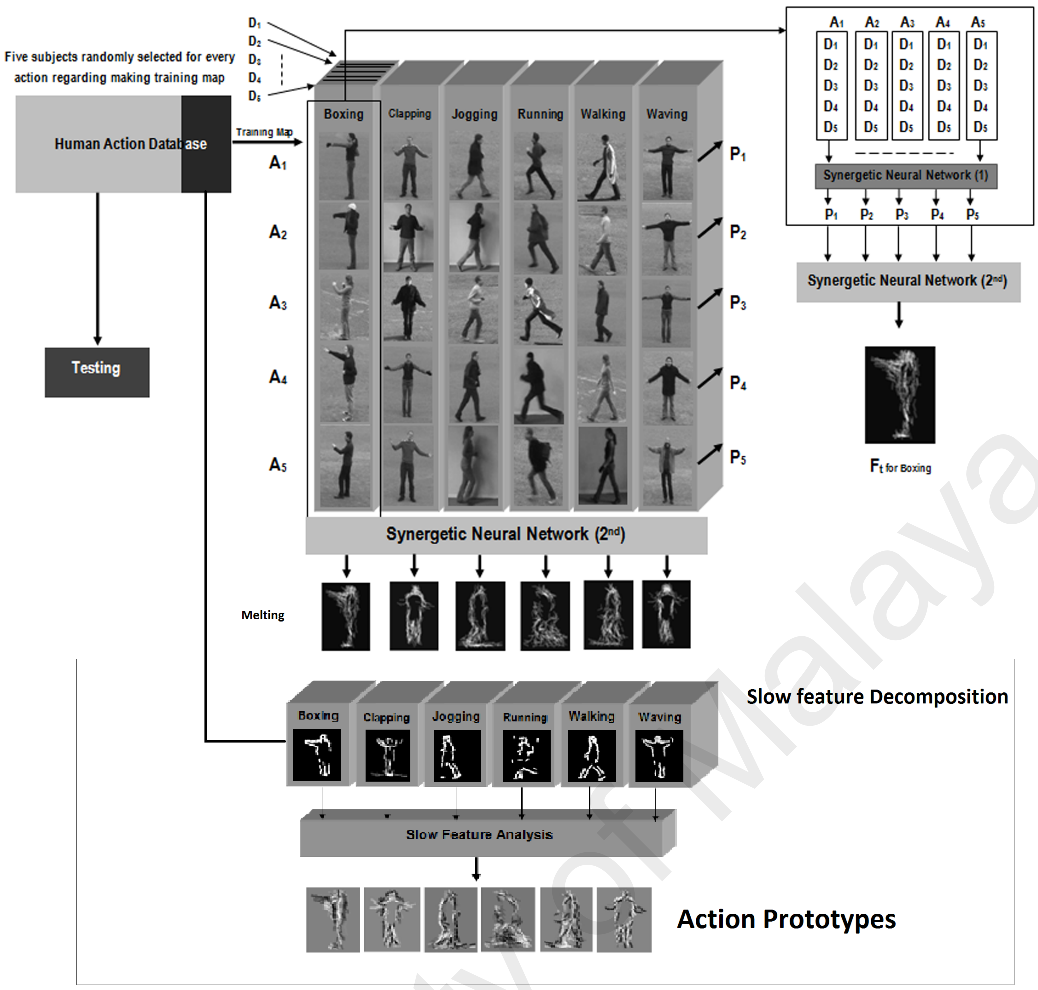

Figure 4.5: Figure shows the procedure of making the action active basis templates by applying two times Melting at SNN or SFA on the training map which calculates from randomly selected video frames from KTH human action database.

4.1.4 Second Scenario for applying action templates

The biologically motivated model in the form pathway was inspired by a computer vision BOW method regarding problems in object recognition. The regular concept of the previously mentioned approaches involves extracting the features in a specific location from a set of image frames for every action, assembling a codebook of visual action words with vector quantization, and constructing an action model by utilizing the four key frames of each action. These models are not certainly correct, and a set of locally selected patches is considered and may ignore many structures; however, these models have been acknowledged as efficient object recognition methods (Grauman Darrell, 2005; Fei-Fei et al., 2006). In the proposed approach, some frames were utilized as key frames (words) to

University of Malaya

recognize human action in whole action frames. Every frame of action video is consigned as one visual word by considering the similarity of each action

Methods	Accuracy(percent)	Year
Schuldt.	71.72	2004
Niebles.	83.33	2006
Jhuang.	91.7	2007
Schindler.	92.7	2008
Wang.	91.2	2009
Zhang.	U-SFA:84.67 S-SFA:88.83 D-SFA:91.17 SD-SFA:93.50	2012
Proposed Model	1^{st} scenario:78.05 2^{nd} scenario:83.34	2013

Table 4.1: The results of recognition by proposed method has presented along with comparison among previous methods on the KTH human action dataset.

codebook. As with problems of object recognition, particular structures were missed by moving this representation (Wang & Mori, 2009). This method exhibits a good performance. Moreover, the local distribution of action sequence is considerably similar to the targeted action but very different from other sample sequences, all of which are in the same action frames but in different categories. Concisely, the variance of intraclass is large, and the variance of interclass is small. In addition, the intraclass variance of the single-person human action recognition is lower than that of the multi-person (Zhang & Tao, 2012). Therefore, its application shows a significant performance in the proposed approach.

4.1.5 Evaluation of Quantum Behavior

Previously mentioned particle swarm optimization results were in the balanced mode of synergetic neural network, which has been performed for better comparison between both scenarios of form pathway. However, quantum particle swarm optimization exhibits an excellent tuning performance for attention parameters, which is constant and equal to one in the balanced mode. While a procedure is working to solve the problem at hand, one of the most significant issues is how to choose its parameters and initiate them. For the

University of Malaya

initial attention parameter, centroidal Voronoi tessellations have been used. The algorithm ran for 500 echoes and for a population size of 20 particles for 20 times.

4.1.6 Evaluation

After converging the algorithm, attention parameters have been used in an unbalanced synergetic neural network to obtain the form pathway. The proposed approach was evaluated through two human action data sets and confusion matrices as previously shown. In this study, the performance of the proposed method is shown and compared with those of previous approaches with the same data sets (Table 4.1 and Table 4.2).

Methods	Accuracy(percent)	Year
Schuldt.	72.8	2004
Niebles.	72.8	2006
Jhuang.	98.8	2007
Schindler.	100	2008
Wang.	100	2009
Zhang.	U-SFA:86.67 S-SFA:86.40 D-SFA:89.33 SD-SFA:93.87	2012
Proposed Model	1^{st} scenario:70 2^{nd} scenario:81.03	2013

Table 4.2: Comparison of the proposed approach and previous methods for Wiezmann human action dataset.

The proposed performances from the saturated state-of-the-art methods used on the KTH and Weizmann data sets achieved good and comparable results. A comparison of methods using biologically inspired state-of-the-art models (with or without biological point of view) is listed in Table 4.1 and Table 4.2. Moreover, the different methods listed in Table 4.1 exhibited numerous variations in their experimental setups, namely, different splits of training/testing data, whether preprocessing is needed or not (e.g., tracking, background subtraction), with or without supervision, whether per-frame classification is feasible, and whether a method handles multiple action classes in a video. The results of the methods were comparable with other state-of-the-art approaches, especially in terms of robustness, although comparing other methods is not absolutely fair; meanwhile, their

University of Malaya

method did not completely cover the biological point of view (Schuldt et al., 2004). However, the biologically inspired technique used by Zhang and Tao (2012) revealed that the proposed model is almost accurate.

University of Malaya

4.2 Results of Fuzzy Optical flow divisions and interactions between two pathways

In the previous section, the results of applying ABM into the ventral processing stream were presented (Yousefi & Loo, 2014a; 2014c). The model was a supervised Gabor-based object recognition technique that exhibits a good performance in finding the human object. However, involving the guidance of the optical flow to the Shared Sketch Algorithm (SSA) in this model results in an excellent interaction on the combination of form pathway and motion pathway information. This configuration has been modified to improve recognition, particularly in human object detection; however, the other path of the visual system, that is, the motion pathway, requires a more robust motion information in terms of instantaneous changes and environmental distortions to help in the decision making. Following the descriptions mentioned in the previous chapter, the optical flow division modified the information of motion by showing resistance in instant variations within the frames. This section presents the results of the optical flow division and its influence on better decision making and robustness.

4.2.1 Selecting video frames

Motion analysis, video processing, and action recognition are based on frame selection for temporal order. Choosing frames based on randomization methods of temporal order can destroy the biological perception of movement (Giese Poggio, 2003). Frame selection through input movie follows the proposed model of form and motion pathway connection from snapshot neurons. Snapshots follow the temporal order regarding configuration motion patterns of different

object activities in both pathways. The proposed model uses feed forward structure for form connection (active basis function) and motion pathway (optical flow). Three frames served as the minimum number of frames for snapshots, which will be taken from the video streams following a temporal order. Moreover, the motion information activates the basis function through feedforward joining in SSA to achieve appropriate connections.

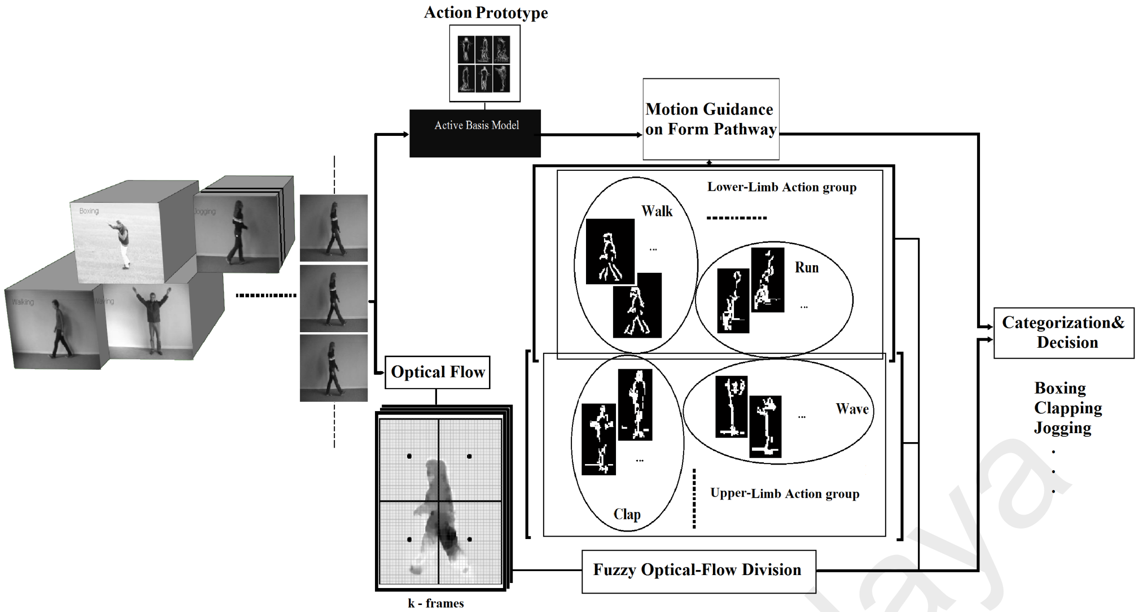

Figure 4.6: Figure reveals the flowchart of our algorithm regarding human action recognition. The flowchart to present the hierarchical model in terms of theoretical and computation for combination of the form information, and motion information is shown. An supervised Gabor based object recognition method, ABM, gives this property to have human object and computation of the form data and its combination with optical flow information, motion information, creates interaction between the pathways.

4.2.2 Relation to existing methods

The proposed approach and the current techniques on human action recognition are basically similar to each other. In this part, differences and similarities are discussed. Processing is needed before the core method, and the biologically inspired model in terms of simplicity should be respected (Efros et al., 2003; Giese Poggio, 2003; Jhuang et al., 2007; Schindler and Van Gool, 2008). The approach is based on object recognition following hierarchical feed-forward designs (Jhuang et al., 2007); moreover, a model was developed following neurobiological motion processing in the visual cortex, following the method of Giese and Poggio (2003). The object recognition task in the form pathway was changed within the researchers' work, from spatiotemporal features similar to those of Jhuang et al. (2007) and Jones and Shao (2013), to the original Gabor filter (Schindler & Van Gool, 2008) for the proposed approach using ABM. However, the ABM has the basic characteristics of previous features and basically uses Gabor wavelet that reduces matching operation. The ABM is activated by limited clutters and ensures that the important amounts in points of interest fall on the person subject. In terms of used features,

University of Malaya

	Boxing	Clapping	Jogging	Running	Walking	Waving
Boxing	0.94	0.001	0.02	0.01	0.001	0.02
Clapping	0.26	0.52	0.02	002	0.02	0.15
Jogging	0.02	0.01	0.70	0.06	0.15	0.04
Running	0.015	0	0.08	0.88	0.01	0.01
Walking	0.03	0	0.18	0.01	0.77	0.01
Waving	0.21	0.03	0.05	0.01	0.01	0.69

	Boxing	Clapping	Jogging	Running	Walking	Waving
Boxing	0.86	0	0	0	0	0.13
Clapping	0.07	0.81	0	0	0	0.11
Jogging	0	0	0.77	0.09	0.11	0
Running	0	0	0.09	0.91	0	0
Walking	0	0	0.17	0.02	0.80	0
Waving	0.13	0.16	0	0	0	0.71

Figure 4.7: Confusion matrices SNN classifying KTH data set obtained by adapted active basis model as combination of form and motion pathways. Confusion matrices of the proposed approach has been presented for the case of without fuzzy interference system, left matrix, and after it, right matrix which are achieved from human action movements of KTH dataset (Schuldt et al., 2004). The robustness of the method after adding the fuzzy interference stabilizer is considerably increased. The wrong recognitions in the left confusion matrix have been decrease especially in case of some actions i.e. clapping. Moreover, soar of robustness helps to increase the overall accuracy and better results in classification of biological movement. The accuracy of categorizations using unbalanced SNN reaches to 86.46%.

layer-wise optical flow (Liu, 2009) is simply a silhouette of the subject motion. In this work, the approach was used to help the active basic model to concentrate on the object and prevent the wastage of Gabor beams. Moreover, as previously mentioned, the proposed approach follows a biologically inspired model (Giese & Poggio, 2003) through parallels to the visual cortex.

4.2.3 Experimental Results

A famous human action and the largest databases, such as the KTH human action data set (Schuldt et al., 2004) and the Weizmann human action recognition robustness set (Blank et al., 2005; Gorelick et al., 2007), were implemented in the tests to estimate the ability of the proposed approach to human action recognition. As the largest human action dataset, the KTH action dataset includes 598 action sequences, comprising six types of single-person actions, such as boxing, clapping, jogging, running, walking, and waving.

University of Malaya

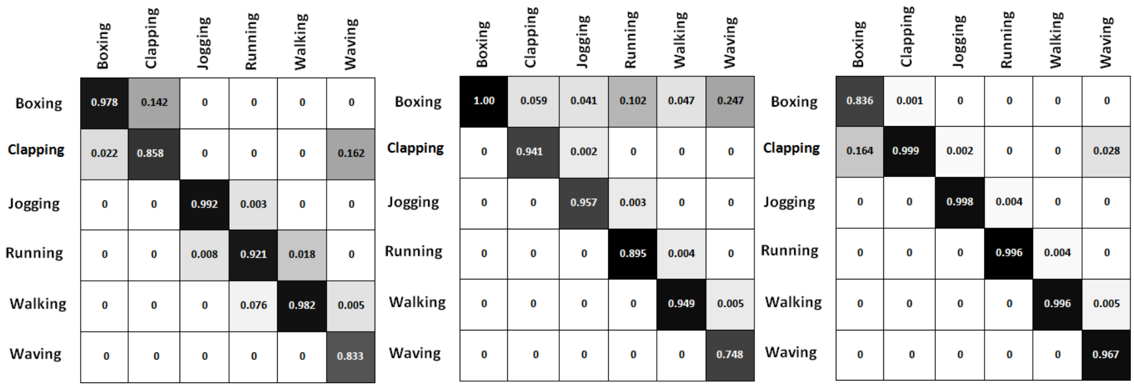

Figure 4.8: Confusion matrices ELM classifying KTH data-set attained by adapted active basis model as combination of form and motion pathways. Confusion matrices of the proposed approach has been presented which is obtained from human action movements of KTH dataset(Schuldt et al. 2004). There are three different kernel have been used to classifying using ELM algorithm(Huang et al. 2004; Huang et al. 2006; Liang et al. 2006; Lehky et al. 2008; Huang, Wang et al. 2011; Rajesh & Prakash 2011) in the decision making and categorization of the biological movement. From left to right, RBF kernel-ELM , Wavelet kernel ELM and Sigmoid-ELM confusion matrices have been depicted which Sigmoid Kernel-ELM has better results in classification of biological movement. The accuracy of categorizations are ELM-Wav = 91.5%, ELM-RBF = 92.7% ,and ELM-Sig = 96.5%.

These actions were performed by 25 people under different conditions: outdoors (s1), outdoors with scale variation (s2), outdoors with different clothes (s3), and indoors with lighting variation (s4). Here, sequence resolutions become 200 $\times$ 142 pixels by using down-sampling. For our approach, we used five random cases (subjects) for training and making the form- and motion-predefined templates. As mentioned in the literature, KTH is a robust intra-subject variation with a large set. However, the video camera during the preparation presented a certain degree of shacking, which rendered the work with this database extremely difficult. Moreover, the four independent scenarios were separately trained and tested (i.e., four visually different databases that share the same classes). Both alternatives were run. For considering the symmetry problem of human actions, a mirror function for sequences along the vertical axis can be available for the testing and training sets. All possible overlaps of human actions within the training and testing sets were considered in this study (for example, one video contained 32 and 24 action frames).

University of Malaya

4.2.4 Contribution between motion and form features

One major strength in comparison with other human action recognition methods is the use of fuzzy optical flow division to guide the share sketch algorithm in ABM. The approach combines the form and motion pathways with respect to the original model. Regarding combination, we may ask whether combining the two pathways is necessary and how the two pathways are combined. In a modification of the original model, the proposed method applied ABM to the form pathway and was adjusted using motion pathway information. The use of optical flow division guidance for ABM is highly successful in preventing Gabor beams wastage, thus offering novelty in comparison with common methods. Optical flow is applied for updating the ABM point of application by evaluating the velocity of the object with the guidance of each optical flow division in the form of fuzzy membership function, implying that the information attained from the motion pathway helps the form pathway. However, the combination of motion and form generally overtakes both motion and form separately. In most conducted experiments, the combined information of these two pathways participates in the final decision part (Giese & Poggio, 2003; Jhuang et al., 2007; Schindler & Van Gool, 2008). In addition, relative feed-forward structure from input data-stream does not change until the final decision and is similar across different data-sets among two independent sets of the computed features (Figure 1 in (Giese & Poggio, 2003) and Figure 2 in (Schindler & Van Gool, 2008)). The proposed approach has been presented previously (Yousefi et al., 2013), but the fuzzy optical flow divisions have yet to be applied. In this study, presentation is conducted with respect to the original model regarding both pathways; the extracted features for each pathway can be relevant. Moreover, feed-forward structure has been modified, and extracted features for both pathways are considered having dependent information.

4.2.5 Results

In this study, the biologically inspired model for human action recognition has been studied. Principally, form features attained from ABM representing the feature form of the pathway are described. We also mentioned that ABM is adjusted by motion pathway

University of Malaya

Table 4.3: The proposed comparison method recognition results has revealed among previous human action recognition method accuracies (bio- or non-bioinspired) on KTH human action dataset.

Methods	Accuracy(%)	Years
Schuldt. Niebles.	71.72 83.33	2004 2006
Jhuang.	91.79	2008
Schindler.	92.79	2009
Wang.	91.29	2009
Zhang.	U-SFA: 86.67 S-SFA: 86.40 D-SFA: 89.33 SD-SFA: 93.87	2012
Proposed Method	SNN : 86.46 ELM: 96.5	2014

information and utilizes fuzzy optical flow division regarding adjustment for increasing recognition accuracy. Afterward, we applied feature selection experiments and prepared an action prototype for every specific movement of human objects applying a synergetic neural network. These templates are established by extracting prototypes for two times from the application of synergetic neural networks on the train set of our human action dataset.

Finally, to examine our proposed approach, we applied it to a popular dataset to determine the accuracy. The confusion matrices of the proposed approaches considered without fuzzy optical flow division and without the division and with recognition of action under consideration using the proposed method are revealed in Figure 4.7. The difference between the two confusion matrices is substantial and can prove the advantage of using fuzzy optical flow division for this context. Furthermore, Table 4.3 reveals a comparison of our method with other methods in terms of recognition accuracy. The accuracy result indicates that the accuracy of the proposed technique is relatively comparable with state-of-the-art techniques by considering two categories with two slightly different paradigms, which cannot be directly compared. In this study, the experimental result of the proposed approach is presented. As the KTH human action database (Schuldt et al., 2004) has been used for benchmarking the accuracy of consistency with a set of experiments used

University of Malaya

in the works of Giese and Poggio (2003), Niebles et al. (2008), Schindler and Van Gool (2008), Wang and Mori (2009), Zhang and Tao (2012), and Jones and Shao (2013), a set of training map and test set for proposed technique was established for the entire data set, in which four-scenario videos were combined. The data set was split to a set of training maps with five randomly selected subjects as well as a test part by residual subjects. Afterward, measurements of the performance were averaged over five random splits. The training map dataset was extremely small, comprising five video-frame snippets that were randomly obtained from the mixture dataset.

Figures 4.7 and 4.8 present classification confusion matrices for the KTH data set. Rows in the confusion matrix represent the corresponding classification results, whereas each column signifies instances to be classified. In the proposed approach, the highest confusion occurs among walking, jogging, and running. These actions are difficult to discriminate because the performance of actions by certain subjects is similar. In addition, another misclassification occurs principally between similar classes, such as previous confusion or hand-clapping and, in consequence, waving (confusion matrices in Figures 4.7, 4.8).

4.2.6 If the fuzzy optical flow division helps to have better accuracy?

Previous sections of the paper focused on obtaining action prototypes through synergetic neural networks during one whole action frame. This method can yield a good abstract from the action video but exhibits a problem (Yousefi et al., 2013; 2014) that decreases our accuracy because of cluttered areas in action prototypes. Following this problem, a similarity was observed among the matched image frames, causing accuracy disparity, as clearly revealed in the confusion matrix (Figure 4.7). The presented approach considered 0.1 s for k = 3, which means three frame times considered as dependency of preventing the membership function value and attained from training on our training set. During the experiment, the upper and lower limb membership functions were applied, and the left and right limb functions can be suggested for complex actions. Disparity was markedly diminished after the application of fuzzy optical flow division. The confusion matrix after the application of this method is shown in Figure 4.7 (second confusion

University of Malaya

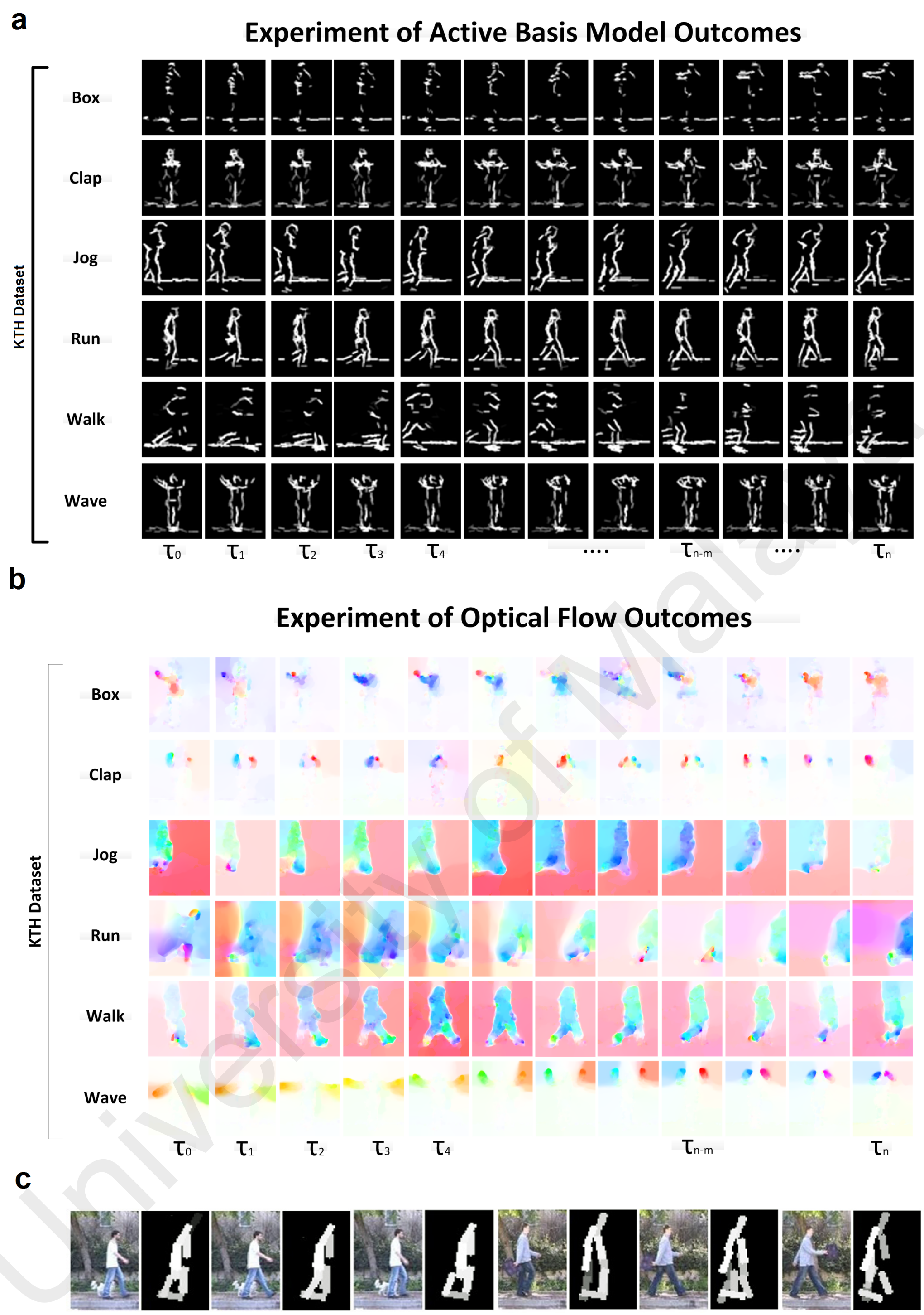

Figure 4.9: Simulation results for simple biological movement paradigm based on ABM (Wu et al., 2010) in the ventral processing stream and optical-flow (Liu, 2009) in dorsal stream are shown. Each row within the panel reveals the response of ABM during the episode as well as flow generated for every different action. The set of biological movements belongs to the biological movements are from KTH dataset (Schuldt et al., 2004). (a) the simulation results of the different actions of KTH dataset. (b) Optical-flow simulation results; (c) The figure depicts some results of Weizmann robustness dataset. It reveals increasing in the robustness of the proposed approach due to utilization of ABM (Wu et al., 2010) in the ventral stream.

matrix).

4.2.7 Related work

Human action recognition tasks are generally categorized as two separated classes. First class prefers to track the part of image where the object (human) exists (Bregler et al., 2004). The mentioned groups of techniques may not be useful in less articulate objects. However, the techniques are considered successful. The other popular class addresses low-resolution videos, high local-resolution images (Dollár et al., 2005), or using spatiotemporal features (Efros et al., 2003; Jones & Shao, 2013). As previously discussed regarding neurobiologically inspired models for the analysis of movement in the dorsal stream visual cortex and psychological and physiological information, our proposed approach was categorized under the second group of methods. The previous method (Giese & Poggio, 2003) shows constant translation lack and a limited handcrafted feature dictionary in intermediate periods (Fanti et al., 2005). Jhuang et al. (2007) and Schindler et al. (2008) presented a biologically inspired method for human action recognition. One major contribution is regarding the combination of both pathways with respect to the neurobiological model. Action prototypes regarding human object recognition can increase the novelty of the method. None of the existing motion processing neurobiological models has used one prototype template dataset to recognize different actions (Hogg et al., 1995; Lee & Loo, 2010). Furthermore, applying the ABM for the form pathway and categorizing the action by utilizing single-action prototype creates certain disparities. This problem can be diminished by using fuzzy optical flow division. For a neuroscience model (Giese & Poggio, 2003; Syrris & Petridis, 2011) into the real world by computer vision algorithm, two important techniques have been altered to make the system more biologically inspired and impart good organization to the proposed approach for object recognition.

University of Malaya

4.3 Results of Combination of slow principle and fast features in interaction between two pathways

This approach applies the mentioned theoretical framework through computer programming and simulations of the patterns of several movements in different environments that appear as typical experiments for benchmarking the system. In the following section, the two human-movement data sets will be introduced as a diverse biological movement paradigm. Simulation results are revealed by mathematical analysis to the mentioned datasets as case experiments for benchmarking. The results of simulation for generating the slowest features in ventral streams and the combination technique for recognizing biological movements are consequently shown. In the following sections, the results of recognition accuracy and confusion matrices are presented. Experimental results are extensively presented to reveal the effectiveness and understanding of the proposed perspective of the biological movement model. For model evaluation, the recognition of different biological movements is rated through model simulation and comparison with state-of-the-art methods (Yousefi & Loo, 2014d; 2015). The aforementioned datasets have been extensively utilized to estimate the recognition performance of biological movement examples for the proposed technique. However, these studies have concentrated on the recognition of single-person actions, such as clapping and walking. The other advantage of using these datasets is its comparison with state-of-the-art methods. Using our testing datasets, we illustrate experimental results by using the kerneled ELM algorithm to classify different kernel modes and a comparison among previous works that proposed biological human action models. In addition, we compared various kernels in the form pathway along with their accuracy. The proposed methods are efficient, and the computational cost will be due to feature extraction regarding two form and motion pathways, slow and fast, features applying ABM-based incremental feature analysis, and optical flow correspondingly. The system infers that a new video only takes a certain time period for our un-optimized MATLAB implementation, which is combined by existing codes for motion and form pathways in MATLAB/C (Huang et al., 2006; Liang et al., 2006; Liu, 2009; Wu et al., 2010; Kompella et al., 2011; 2012; Zhou & Huang, 2012).

University of Malaya

4.3.1 Results of simulation for the slow features of the ventral stream in biological movement paradigm

Ventral stream results must be oriented to surround the concepts of shape and form features, following the original biological movement model (Giese & Poggio, 2003). Numerous approaches are based on a visual system. Cooperation between information attained from two processing streams occurs at few levels in the mammalian brains (Kourtzi & Kanwisher, 2000; Saleem et al., 2000) and can thus simplify the aggregation (for instance in STS level in Giese& Vaina (2001)) and improve the performance of the model. Holonomical features consider both pathways for predefined action templates. In the form pathway, the proposed approach follows Karl Pribram's holonomic theory, which is based on an evidence that dendritic receptive fields in sensory cortexes are described mathematically by Gabor functions (Schuldt et al., 2004) and massively used by the ABM (Wu et al.,2010). As mentioned before, the primary stage includes local (in V1 cell) and model detectors (Gabor-like filters) in 16 (including eight preferred) orientations and by proper scale depending on the receptive field (Dow et al., 1981; Riesenhuber & Poggio, 2002). The ABM also functions as snapshot detectors of human body shapes model finding like with the IT of monkeys, where view-tuned neurons are located and the model of complex shapes is tuned (Logothetis et al., 1995). This model is implemented through ABM-based IncSFA. In particular, slow feature analysis represents the performance of view-tuned neurons in IT and snapshot neurons that provide dependence in scale and position. The proposed model is obtained and adjusted through training as key frames. The optical flow outcome is utilized and inferred with information known as the biological object form. The presented approach covers a high-level integration of snapshot neuron outcomes with motion pattern neuron information. Furthermore, the ABM uses a computational mechanism regarding recognized human object form along with slowness information throughout biological movements, following neurobiological, neuro-computational, and theoretical records (Giese & Poggio, 2003; Schindler & Van Gool, 2008; Liwicki et al., 2013). Local direction was organized at the initial level of the form pathway, and Gabor-like modeling detector methods, such as ABM, exhibited good constancy by modeling cells in the mentioned part (Jones & Palmer, 1987). We determined 16 directions and two spa-

University of Malaya

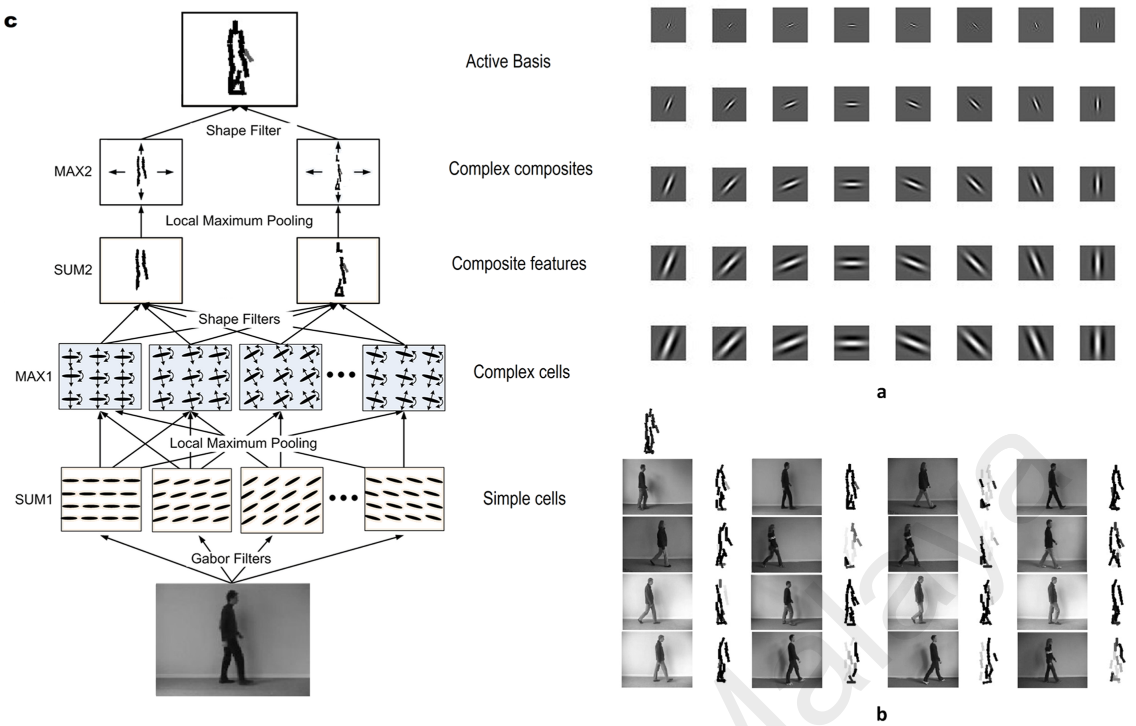

Figure 4.10: The explanation diagram of the ventral processing of the applying ABM (Wu et al., 2010) which represents movements pattern and shape form of biological object within its movement episode. ABM is a Gabor based supervised object recognition method, which is, can learn the object shape in the training stage and can be utilized object recognizer within the action episode. (a) it represents the Gabor bank filter in different scales and orientations. (b) Simulation results for training of ABM system for biological walking movements using KTH human action recognition dataset. At the end, walker shape has been presented in the top of the figure. (c) the processing diagram of the ABM process for finding human object presented. The similarity between the method and biological finding in different level has been mentioned in different stages. Overall, ABM have two stages SUM & MAX which make the hierarchy from simple cells to complex cells and at the end whole human object shape by active bases.

tial scales by two differentiators and found information on local direction in the pathway and complex-like cells with independent form features. These form features are appropriate for the form pathway. This approach will be applied by using the mechanism of the proposed neurophysiological plausible model.

The outcomes of ABM for every possible action and simulations through the ventral stream for human movements have generated complex-like cell outcomes (Figure 4.10), which rendered the system more plausible. However, object-based techniques, such as ABM, direct the system to the human object and decrease a weak attitude, which may occur through a Gabor-based model (non-object detector-based techniques). The slowest

University of Malaya

outcome through IncSFA confirmed that active bases represent the biological object form in the ventral streams (Figure 4.11). In Figure 4.11, the active bases of individual biological human object movements are shown for simulation with a set of 40-direction Gabor beam banks with a 0.7 Gabor scale in 15 orientations. In this simulation, all units resemble the theoretical predictions. In addition, frameworks are similar to biological object recognition tasks but are completed by the extraction of slowness features.

4.3.2 Slowness features in Ventral stream

The recognition of the biological movement patterns in the form pathway depends on the slow features generated by IncSFA (Kompella et al., 2011; 2012). The slowest features of the training set were used as human action prototypes. First, we performed multi-prototype predefined templates for each obtained human action by applying IncSFA on the datasets. To create a training map for every action, we divided every human movement sequence to training and testing sets. These action prototypes are considerably preventative for different biological movements.

The result of slowness features attained through applying IncSFA is revealed in Figure 4.11. Two categories of biological movements were included for every dataset. Different slowness prototypes were required, and actions were not directly comparable. As the KTH and Weizmann human action databases have been used for benchmarking the approach performance, consistency with the set of experiments used in the references (Schuldt et al., 2004; Schindler & Van Gool, 2008; Zhang & Tao, 2012) is required. The training map and test of the proposed technique were defined for each data set, in which the mixture of four scenarios videos was grouped together (for KTH data set). The datasets were split into a set of training maps with randomly selected subjects and a test part by the residual subjects. Afterward, IncSFA was applied to the training sets, obtaining slowness feature prototypes, which function in form movement templates. The slowness features and the slowest feature in the ventral stream of the proposed biological movement model helped prevent the use of computer vision techniques, such as “bag-of-words.” The regular concept of the mentioned approaches comprises extracting the slowest features in the set of image frames for every action. The ABM-based IncSFA in the model performs

University of Malaya

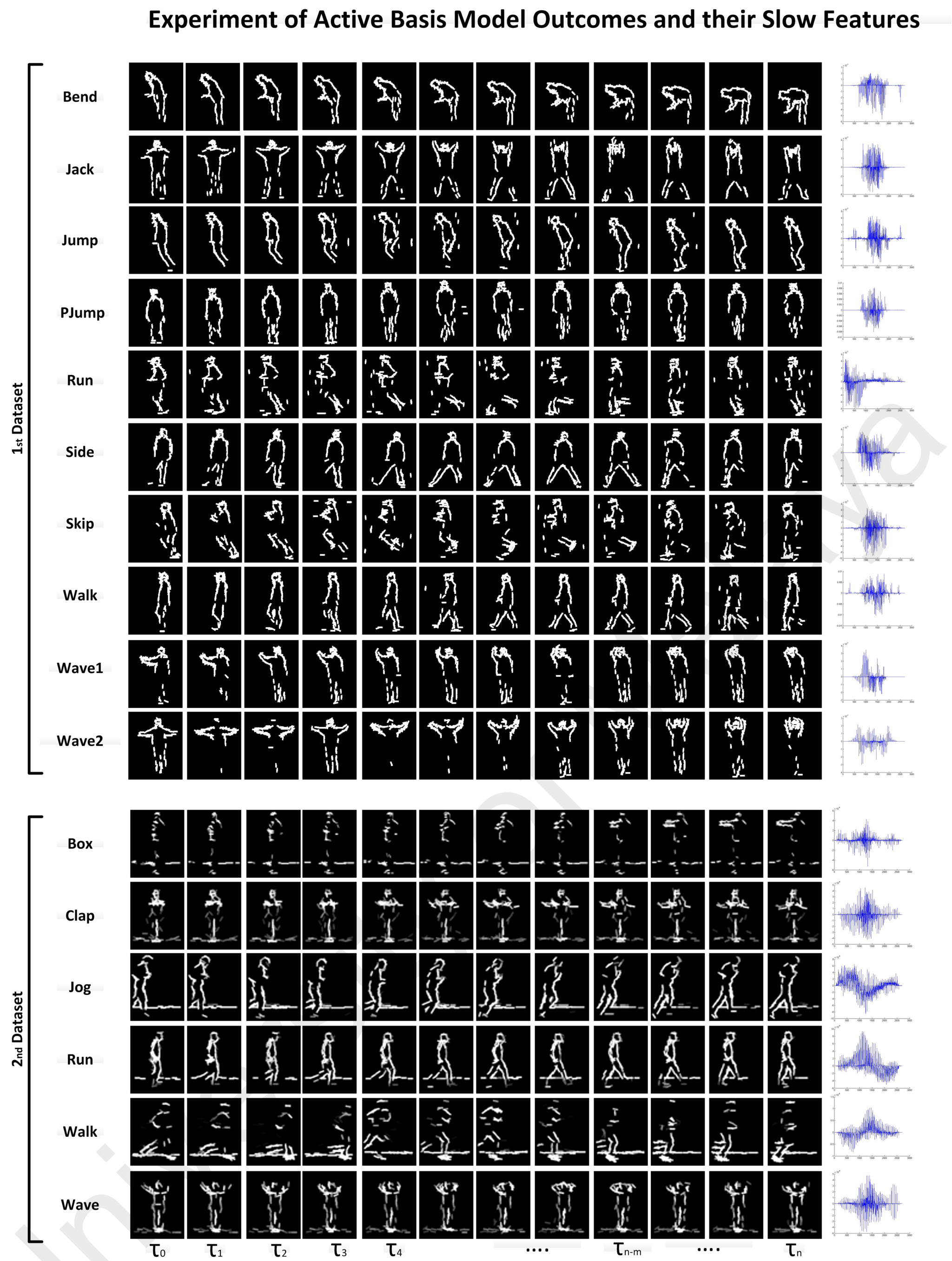

Figure 4.11: Simulation results for simple biological movement paradigm based on ABM based slow features in the ventral processing stream shown. Each row within the panel reveals the response of ABM during the episode as well as slowness features generated for every different action. The first set of biological movements belongs to Weizmann human action recognition dataset (Gorelick et al., 2007) and second group of the biological movements is from KTH dataset (Schuldt et al., 2004). Simulation results of the active bases through incremental slow features follows the theoretical prediction regarding simplification of recognition using ABM based IncSFA and its application in ventral processing stream for opening a new perspective of the original model of recognition of biological movements.

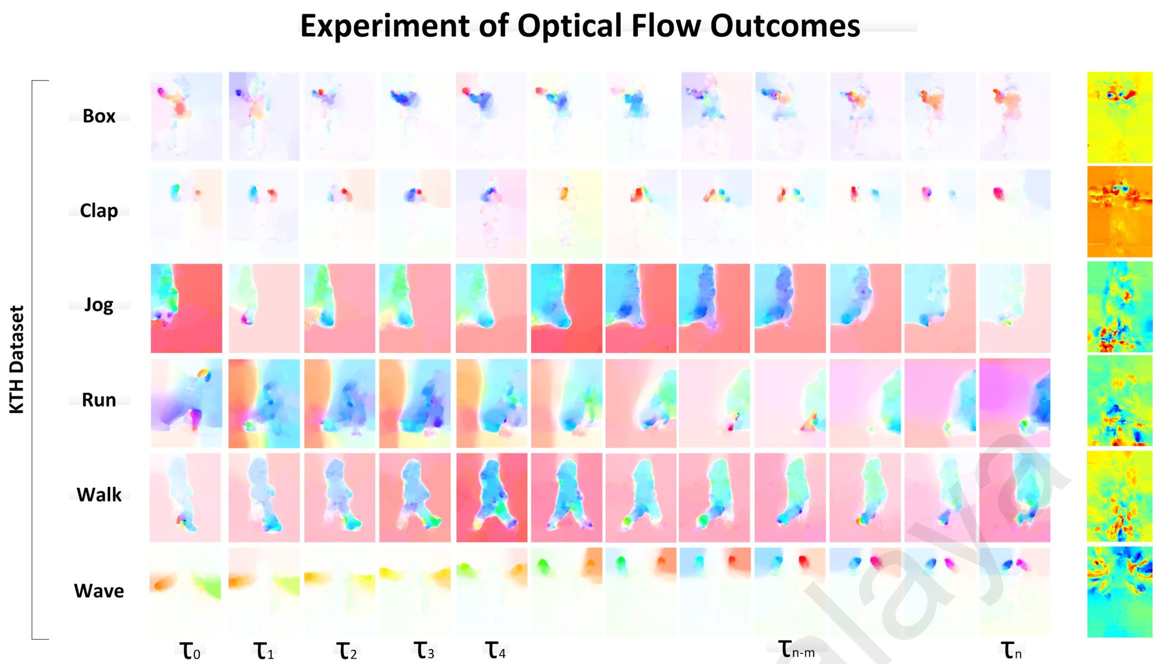

Figure 4.12: Simulation results regarding dorsal processing stream applying Optical Flow (Liu 2009) has been depicted here. As there is episodic operation happening in the ventral processing stream, form information, motion information, fast features, must be considered during the time that ventral stream is performing $(t_0, \ldots, t_n)$. Each row is representing an action during its episode and the average of the flow for whole episode considered at the end of every row. The images present flow in color form, which can depict the biological movement flow in the episode. Average of Optical flow throughout the biological movement considers for recognition of the biological movement by fast features and add a coefficient to form pathway results as the interaction between ventral and dorsal streams. Different actions of simulation presentation are from KTH human action recognition dataset (Schuldt et al., 2004).

considerably well on the mentioned datasets. The whole model outperforms reported state-of-the-art computational methods. Moreover, the proposed model achieves superior performance to some methods, such as bag-of-words and action key. Although the considered patch sets were not locally selected and may ignore numerous structures, the data sets have been acknowledged as an efficient object recognition method. This method exhibits significant performance (Figures 4.13 and 4.14). Meanwhile, the local distribution of the action sequence is highly similar to the targeted action and extremely different from other sample sequences in different categories. Concisely, the interclass variance is large, whereas the interclass variance is small. In the case of single-person human action recognition, the interclass variance is smaller (Zhang & Tao, 2012); consequently, the model has been applied to recognize biological movements.

University of Malaya

4.3.3 Simulation Results for the dorsal stream and information on the motion pathway in the biological movement paradigm

To implement this pathway, we applied optical flow, a common and extremely noticeable tool, in the proposed technique (Liu, 2009) to generate information regarding motion pathway. Motion information for the recognition of biological movements was obtained by analyzing optical flow patterns (Giese & Poggio, 2003). The information contains neural detectors for optical flow features with growing complication, which is related to neurophysiological data (Tanaka, 1996). In the present study, information on motion-processing stream was considered as a fast feature from the perspective of temporal changes through biological movements. These features are generated not as constant representative features throughout a whole episode but instead focus on the temporal order within the current frame. In contrast to the form pathway, motion path possesses temporal-based features, and every feature represents the motion information in specific temporal movements. On the basis of several proposed neurophysiologically plausible models for the approximation of local motion, such as those mentioned in the references (Rodman & Albright, 1989; Sereno, 1993; Nowlan & Sejnowski, 1995; Simoncelli & Heeger, 1998; Grossberg et al., 2001; Giese & Poggio, 2003), the first level of the motion pathway comprises correspondence detectors for local motion, which involve direction-selective neurons (Smith & Snowden, 1994) and motion-selective component neurons in MT (Giese & Poggio, 2003). In the simulation stage, the temporal optical flow patterns are directly calculated, and motion-sensitive neural responses are computed using realistic physiologically parameters (Giese & Poggio, 2003). The size of the receptive field in the range of neurons for direction selection is in V1 (foveal neurons) and MT (Hubel & Wiesel, 1968). The second level of motion pathway includes large receptive fields for flow local structure that induces the movement stimuli.

The selective flow translation and neurons of the motion pattern correspond to the MT area (Rodman & Albright, 1989) with band-passed or low tuning by considering speed. Typically in the original model, four-directional neuron populations are preferred, and local optical flow detectors are considered for motion edges (Giese & Poggio, 2003). The output signals are calculated using a combination of two nearby subfields with con-

University of Malaya

	Boxing	Clapping	Jogging	Running	Walking	Waving
Boxing	1.00	0	0	0	0	0.023
Clapping	0	0.857	0	0	0	0
Jogging	0	0	1.00	0	0	0
Running	0	0	0	1.00	0	0
Walking	0	0	0	0	1.00	0
Waving	0	0.023	0	0	0	0.85

	Boxing	Clapping	Jogging	Running	Walking	Waving
Boxing	0.857	0	0.023	0	0	0.023
Clapping	0	0.857	0	0	0	0
Jogging	0	0	0.857	0	0	0
Running	0	0	0	1.00	0	0
Walking	0	0	0	0	1.00	0
Waving	0.023	0.023	0	0	0	0.85

	Boxing	Clapping	Jogging	Running	Walking	Waving
Boxing	0.875	0	0	0	0.023	0.023
Clapping	0	0.714	0	0	0	0
Jogging	0	0	1.00	0	0	0
Running	0	0	0	1.00	0	0
Walking	0	0	0	0	0.857	0
Waving	0.023	0.023	0	0	0	0.857

Figure 4.13: Confusion metrics of the proposed approach has been presented which is obtained from human action movements of KTH dataset. There are three different kernel have been used to classifying using ELM algorithm in the decision making and categorization of the biological movement. From left to right, RBF kernel-ELM, Wavelet kernel ELM and Sigmoid-ELM confusion matrices have been depicted which Sigmoid Kernel-ELM has better results in classification of biological movement.

tradictory preferred directions. As a reminder, motion-selective neurons and opponents have been found in several areas of dorsal processing streams containing the areas MT, MSTD, and MSTl (Dow et al., 1981; Eifuku & Wurtz, 1998). The optical flow pattern neurons in the third step of the motion path stream correspond to snapshot neurons in another pathway. Optical flow pattern neurons have been found at different locations of the visual cortex (i.e., STS or fusiform and occipital face areas). Temporally optical flow pattern neurons are generated during action cycles and modeled like the form pattern neurons from the form processing stream. The difference is that, in the feature space, motion pattern features are considered as fast features ahead of form features, which are the slowest features. Figure 4.12 reveals the motion patterns features throughout action cycles in consideration of integration in the processing stream.

4.3.4 Evaluation of Interaction between Two Paths

To analyze the model, more than 38000 frame cuboids from different biological human action movements were prepared. The recognition of human movement model is episode based because the incremental slow feature analysis training of this algorithm requires numerous inputs frames; thus, the testing and evaluation of the model are extremely limited. The proposed models are valued through the level of prediction matching with the data. The proposed model subsumes the original models and focuses on slowness

University of Malaya

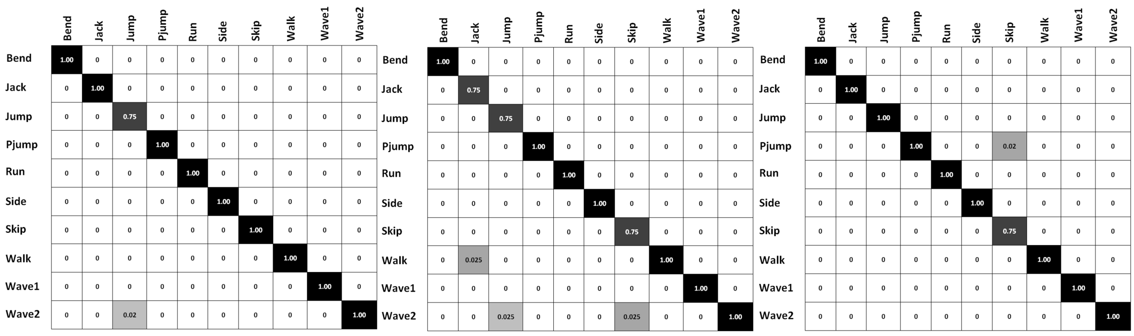

Figure 4.14: Confusion matrices of the proposed approach has been presented which is obtained from human action movements of Wiezmann dataset (Gorelick et al. 2007). There are three different kernel have been used to classifying usig ELM algorithm (Liang et al. 2006) in the decision making and categorization of the biological movement. From left to right, RBF kernel-ELM , Wavelet kernel ELM and Sigmoid-ELM confusion matrices have been depicted which RBF and Sigmoid Kernel-ELM have better results ,97.5%, in classification of biological movement.

features and episodic processing; hence, the proposed model approximates the ideal and yields satisfactory results. However, whether the proposed model reaches a completely accurate level cannot be guaranteed. Subsequently, the system was trained and tested by computing features in both pathways with different, previously mentioned settings. For a specified test sequence, the action label was assigned to the action frames. The proposed model correctly classifies most of the actions (confusion matrices are revealed below). Most of the occurring mistakes in recognition of “running,“ “jogging,” “boxing,” “clapping,” and “waving” used to occur (Danafar et al., 2010; Somayeh et al., 2010; Zhang & Tao, 2012). The intuitive reasoning was because of the resemblance between the two groups of movements. However, the presented model markedly diminishes this issue because of episodic learning and ABM-based slowness features in the ventral processing stream.

The result of each human movement scenario is mentioned in Tables 4.4 and 4.5, which represent the accuracy of the proposed approach in comparison with several earlier methods on the same datasets. However, this comparison is not precise because of differences in experimental setup. The presented results are comparable with state-of-the-art systems, whereas various other methods exhibit multiple differences in their setups, such as the presence or lack of supervision and tracking, subtraction of the background, or con-

University of Malaya

Table 4.4: Comparison of the proposed approach and previous methods for Wiezmann human action dataset.

Methods	Accuracy	Years
Schuldt.	72.8%	2004
Niebles.	72.8%	2008
Schindler.	100%	2008
Wang.	100%	2009
Zhang.	U-SFA:86.67%	
	S-SFA:86.40%	
	D-SFA:89.33%	2012
	SD-SFA:93.87%	
Proposed Model	97.5%	2013

sidering multiple actions recognition. We evaluated the proposed approach through two human action datasets, and the confusion matrices are shown (Figures 4.13 and 4.14). In the present study, we reveal the performance of the proposed model in comparison with previous approaches on the same data set (Tables 4.4 and 4.5). In addition, the different methods listed in Table 4.4 exhibit all types of variations in their experimental setups, including different splits of training/testing data, whether pre-processing (such as tracking and background subtraction) is needed, presence or lack of supervision, whether per-frame classification can be conducted, and whether a method handles multiple action classes in a video. Model results are stable because of slow features in the ventral stream and its combination with fast features; although comparison with some of the other methods is not fair, their method does not completely cover the biological point of view (Schuldt et al., 2004).

4.3.5 Further Analysis

After explaining all approaches and techniques presented previously, we assessed the ability of the systems and analyzed the combination of ABM-based slowness features, as well as its combination with the so-called fast features attained from the dorsal processing stream. Figure (4.15) represents the confusion metric and experimental result accuracy for every processing pathway in the model; however, no interaction occurred between

University of Malaya

Table 4.5: The results of recognition by proposed method has presented along with comparison among previous methods on the KTH human action dataset.

Methods	Accuracy	Years
Schuldt.	71.72%	2004
Niebles.	83.33%	2008
Schindler.	92.7%	2008
Wang.	91.2%	2009
Danafar	93.1%	2010
Zhang.	U-SFA:84.67%	
	S-SFA:88.83%	
	D-SFA:91.17%	2012
	SD-SFA:93.50%	
Proposed Model	90.07%	2013

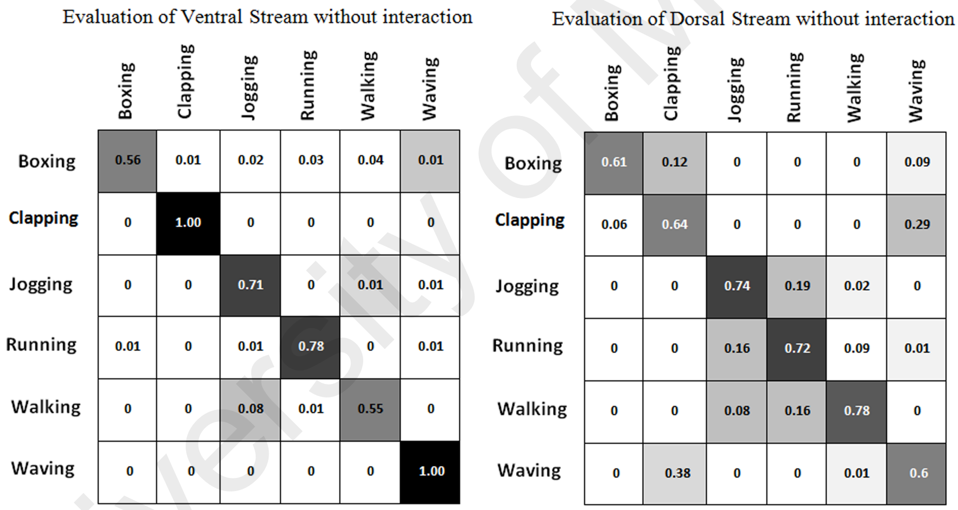

Figure 4.15: Confusion matrices of two pathways separately represent the accuracy of each processing stream considering no interaction between the pathways. These representation is for KTH human action dataset for benchmarking without interaction of the paths. It can be justifies the performances of two patients, DF whom developed visual agnosia (damage to ventrolateral occipital) and RV whom developed optic ataxia (damage on the occipitoparietal cortex (Goodale et al. 1994)).

this information. Biologically inspired methods and applicable methods show considerable challenges. The application of the slowness principle in the extraction of ABM-based slow features or optical flow-based methods could be used to yield higher accuracy than the biological model and computational load. This reasonable justification shows why computer vision methods sometimes may show higher rate of the accuracy. The com-

University of Malaya

bination of the information from both pathways and the interaction mode between these paths render the system authorized and categorized in the recognition of biological movement. Numerous pieces of physiological, neurophysiological, and psychological evidence have proven the consistency of the model with the mammalian visual system. This concept will be argued in the discussion chapter based on new modification on the original model.

4.4 Chapter summary

The results of the proposed approaches mentioned in this chapter and quantitatively compared with the state-of-the-state methods. Here, there are some related points that should be considered for every method:

Applying ABM into ventral stream : applying ABM as a supervised object recognition method in ventral stream increased the robustness of the mechanism for recognition of human (as an object). The selected boundaries of human in the outcomes of ABM (see Figure 4.2) have significant similarity with the results of MLD (used for action perception-Figure 3.10).

Optical Flow division: Optical flow division provided reasonable robustness into the motion pathway (Figure 4.9). It also gives interaction between both the streams.

Slowness principle : involving slow feature analysis into the mechanism for recognition biological movement, defined the concept of fast and slow features corresponding to dynamic and static patterns.

University of Malaya

CHAPTER 5:

Discussion and Conclusions

The previous chapter has provided details on the implementation, presented experimental simulation, and indicated the generated results. The original model has been significantly modified in the proposed approaches to recognize biological movements based on the mammalian visual system. Until this paper, the model of biological action recognition has been modified by using Gabor wavelet in the ventral processing stream to extract form information. Optic flow has been widely used to generate motion information. Approaches have followed psychological and computer-vision evidence with respect to neuroscience and computational intelligence models. The discussion and conclusion for every modification in this research included various subtitles with conclusions and discussions in these approaches.

University of Malaya

5.1 ABM in the ventral processing stream and interaction between the paths

The experimental results of actively applying a basic model in the original model for the recognition of biological movement are extensively presented to reveal the effectiveness and estimate the ability of proposed model in a human action recognition task (Yousefi et al., 2013; 2014b). The consideration of proposed technique in terms of biologically inspired models for recognition along with computer vision performances are integrated in this project. Cooperation among information attained from two processing streams occurs at few levels in mammalian brains (Kourtzi & Kanwisher, 2000; Saleem et al., 2000), simplifies model aggregation (for instance in STS level (Giese & Vaina, 2001), and improves recognition performance. Holonomical features consider both pathways for predefined action. The correctly classified sequences are reported as the highest results in the literature. To place the proposed technique in this context, we presented the sequences along with state-of-the-art systems. This method, similar to other methods, includes frame-based runs for all frames of action sequences. Individual labels obtained from the training map are simply compared with a sequence label through majority voting (similar to the bag-of-frames model and that in a previous study (Schindler & Van Gool, 2008)). Comparison with the state-of-the-art system was conducted, and the results are revealed in Tables 4.1 and 4.2. Notably, original frames are adopted as system input, and the use of different frames results in decreased performance when considering a random location of Gabor beams on human objects in different frames. A training map dataset (in the shape of multi-prototype template set) comprised five frames of action snippet videos that are randomly obtained from the mixed dataset for the case of multi-templates experiment. In the second scenario, four key-frames have been precisely selected from videos randomly selected for every action. Figures 4.3 and 4.4 show the classification confusion matrices for the KTH and Weizmann data sets. The row of confusion matrices represents the corresponding classification results, whereas each column signifies the instances to be classified. In terms of contribution, the application of the ABM in the form pathway is utilized for the first time in a biological model, and the fuzzy inference system involving the combination of the two pathways is the novelty of the proposed model. However, the

University of Malaya

natural questions (Schindler & Van Gool, 2008) arise regarding whether this combination is necessary and how the model can be improved in terms of accuracy. Experiments are conducted for this part of the presented model, in which the form pathway is modified and combined with motion features. The combination results in a complete relation between the two almost independent feature pathways, revealing promising results. This robust approach is superior to other human action recognition methods that use a similar biological model because of the application of the ABM for focusing on the target. In addition, the approach combines the form and motion pathways with respect to the original model. Regarding combination, the necessity of combining the two pathways and which combination form is superior may be in question. These questions have been answered to a certain extent in other approaches that considered different interaction techniques and methods in this context. The interaction between paths using the fuzzy inference system and the information attained from the motion pathway helps form the pathway or the other way around. However, the combination of motion and form generally overtakes both motion and form separately; in most of the conducted experiments, the information of these two pathways is combined in the final decision part (Giese & Poggio, 2003; Jhuang et al., 2007; Schindler & Van Gool, 2008). In addition, the relative feed-forward structure from the input data-stream until the final decision remains unchanged across different data sets between two independent sets of features computed (Figure 1 in Giese and Poggio (2003) and Figure 2 in Schindler and Van Gool (2008)). This work shows that the extracted features for each pathway can be relatively utilized in the other pathway by referring to the original model topology regarding both pathways. Moreover, the configuration of both pathways was modified by using the fuzzy inference technique. We presented this approach, which involves a biologically inspired model based on inter-relevant calculated motion and form features tested, for application in human action recognition tasks. In principle, the approach includes defined form features with the application of the ABM as a form extractor in the form pathway. This pathway was modified, and optical flow was used as flow detector in the motion pathway for the video sequence stream. Unbalanced synergetic neural networks were utilized to classify the shapes and structures of

University of Malaya

human objects along with the tuning of quantum particle swarm optimization (QPSO) by the introduction of Centroidal Voronoi Tessellations, which are utilized and proven as good tools in the form pathway. Finally, a decision was formed through the combination of final outcomes of both pathways in the fuzzy inference domain and fusion of these two brain pathways, considering each feature set to Gaussian membership functions and then fuzzy product inference. Two configurations have been proposed for the form pathway: the first scenario applied multi-prototype human action templates using two time synergetic neural networks for obtaining uniform template; the second scenario used a model motivated from bag-of-words and abstracting human action in four key-frames. The experimental result of the proposed model shows promising accuracy, and robust performance results from the use of the KTH and Weizmann data sets. Furthermore, the model shows good performance on different datasets, and training is done with lower computational load and regarding final prototype template learning. However, initialization of attention parameters requires longer time to find the appropriate attention parameters. Questions that remain for scrutiny include methods for diminishing the computational load for model training and whether improvement is needed. Further works extend the proposed approach for better integration of the present form and motion information in the two pathways. Another extension is to find an accurate method of finding a classifier, which has been improved in subsequent approaches.

University of Malaya

5.2 Fuzzy Optical flow divisions and its helps for interaction between two pathways

In the present study, the approach scrutinized the theory of the interaction of ABM, ventral processing stream, optical flow, and dorsal stream with respect to the original model. The proposed method investigated the influence of interaction of the optical flow division in its fuzzy interference on the SSA in the ABM method (Giese & Poggio, 2003; Schindler & Van Gool, 2008; Danafar et al., 2010; Somayeh et al., 2010; Yousefi & Loo, 2014a; 2014c). We applied a supervised learning Gabor-based method, which has been successfully utilized previously for the task of object recognition (Wu et al., 2010) for the form pathway (Yousefi et al., 2013). The form pathway is considered for ventral stream representation and functions in biological object recognition. An active basis model can learn the human object by considering the prototypes and find the object within frames. This property is highly advantageous for visual system representation in the model; however, this part has been conducted with the Gabor wavelet in the previous models (Giese & Poggio, 2003; Schindler & Van Gool, 2008; Danafar et al., 2010; Somayeh et al., 2010) and similar works (Schindler & Van Gool, 2008). In the visual system, Gabor-like filters mainly function in representing simple and complex cells. The ABM is a suitable preventative model for this part and especially notable in its involvement in object recognition task. The model could follow the involved encoded object shape (Mather et al., 2013). The object shape concern in the form pathway and ventral stream was considered based on the training stage, and its consistency was justified through human prototypes. The ABM considers the Gabor action stimulus for pinning down form processing at two levels of local information on limb angle from Gabor orientations and global body structure signaled by the spatial arrangement of Gabor paths. Conversely, using optical flow for extraction of motion information is followed by the second attribute and involves filtering by direction-selection sensors and its integration for solving the famous aperture problem. Motion information presents the local velocity of both types of motion signal. Joint motion trajectories will function as the signals to the form path by guiding SSA in the ABM (Thurman & Lu, 2013) as a good representation of cross-connection between V4 and MT (Cloutman, 2012; Janssen et al., 2012; Mather et al., 2013). This step follows the predom-

University of Malaya

inant view of form and motion processing in human visual system, which assumes that these two attributes are handled by independent and separate modules (Giese & Poggio, 2003; Schindler & Van Gool, 2008; Danafar et al., 2010; Somayeh et al., 2010). Form signal information can induce the influence processing of motion more extensively than was previously thought (Mather et al., 2013). Moreover, the proposed approach considers a direct effect on motion information during form processing. Connectivity within the visual system is characterized by cross-connections with respect to parallel feed-forward connection (Felleman & Van Essen, 1991; Distler et al., 1993; Cloutman, 2012). The use of the optical flow division technique provides connection and interaction of bottom-up and top-down processing among brain regions along the dual computational streams. Moreover, this approach can be a good representative for the connection between ventral and dorsal streams (that is, V4 and MT). In addition, the dorsal stream is assumed to conduct complementary spatial computation ("where") and ventral stream for performing object recognition (“what”) in the cortical areas V1, V2, V4, and IT along with current evidence in opposition to a complete segregation of the ”where” and “what” information in the brain of macaque (Felleman & Van Essen, 1991; Hung et al., 2005). These findings indicate that information on the position and size of objects is also represented in the IT of macaques as the top layer of the ventral stream. However, the proposed approach involves an early isolation of spatial configuration and identity into divided processing pathways that require heavy hardware computation. Given their low resolution, optical flow divisions (four divided parts) could be a good parameter for diminishing this computational load. The accurate classified sequences are reported as the highest results in the literature. To place the proposed technique in this context, we presented the technique with a state-of-the-art system. Our method is similar to other methods that are frame based for all action sequence frames. Then, the individual labels obtained from the training map are simply compared with a sequence label through majority voting (like in (Schindler & Van Gool, 2008; Danafar et al., 2010; Jones & Shao, 2013)). The results of the comparison with the state-of-the-art system are revealed in Table 4.3. Its accuracy in comparison with other methods indicates the relative compatibility of this approach. In

University of Malaya

terms of contribution among motion and form features, the ABM, being a Gabor-based model, is a modified form pathway. In addition, the approach can learn that the object increases system robustness, as tested using the Weizmann robustness dataset. Moreover, the approach provides the optical flow guidance for SSA as a cross connection among the dual computational streams and realizes the prevention of application of Gabor beams for non-targeted objects. Considering that the fuzzy optical flow division maintains system robustness, this method can be categorized as an improvement in this field. The presented method was tested using experiments, in which a form pathway was modified and combined with a motion path, resulting in a relationship between the two independent feature sets. The connection revealed promising results. Overall, a human action recognition method has been proposed. This method is an extension of the previous approaches (original model and presented model of section before). This method is based on inter-relevant calculated motion and form information following the biologically inspired system. The ABM is applied for generating form information, and optical flow guides the share sketch algorithm for better concentration on human object in video frames; thus, the model can represent the cross-connection of V4 and MT in the brain (Schindler & Van Gool, 2008). The synergetic neural network was used twice on the training set to find action prototypes for each action. The approach was tested for the KTH and robustness Weizmann human action datasets, and experimental assessment of the proposed technique showed promising results comparable with those of state-of-the-art methods. Moreover, the results are beneficial for the proposed cross-connection into the feed-forward method on biological movement. Moreover, the method exhibited good performance on different datasets, and training was conducted with lower computational load regarding final action prototype learning and computational cost. One limitation of the proposed approach is the lack of mechanisms for invariance against rotation and viewpoint changes despite the multiscale capability of the mechanism. Several questions arise, such as in which frame does motion sequences consistently represent the recognition of video stream and the extent to which the two pathways clearly follow the biologically inspired movement of mammalian brain. Further approach has shown better episodic recognition of the biological movement by

University of Malaya

using an ABM-based slowness principle.

5.3 Applying slow principle

How should the proposed model with a slow and fast processing stream be gauged in the task of recognizing biological movement? The presented results underline that the combination of form and motion or of slow and fast information was performed by ABM-based IncSFA, respectively (Yousefi & Loo 2014d; 2015). The temporal features in terms of episodic or frame consideration of the features attained through pathways are our instincts regarding "slow" and "fast" features, which often approach the original model. The proposed method offers a different perspective for overlooking the original recognition of biological movement (Giese & Poggio, 2003). In addition, the model is presented to destroy a model viewpoint to perform under slowness and fastness features, which are satisfactory according to the mammalian visual system, and the motion analysis temporal response (Shioiri & Cavanagh, 1990) along with sensory information was gathered over diverse time scales (Hasson et al., 2008). Consequently, with respect to the original model, the model was able to achieve good performance in the targeted databases (Schindler & Van Gool, 2008; Danafar et al., 2010; Somayeh et al., 2010) and should be objectively considered in the recognition task. From a biological viewpoint, an understanding of biological movement containing both pathways and the cooperation between information attained from two processing streams occur at several levels in mammalian brain (Kourtzi & Kanwisher, 2000; Saleem et al., 2000). The approach can simplify the aggregation of model (for instance in STS level (Giese & Vaina, 2001)) and improve recognition performance. Although current neuroscience and psychophysics research shows that the effect of form signals on motion processing is more widespread than earlier supposed (more details in (Mather et al., 2013)), the holonomical features consider both pathway features to recognize biological movements. In the form pathway, the approach followed Pribram's holonomic theory, which is based on evidence that the dendritic receptive fields in sensory cortexes are described mathematically by Gabor functions (Pribram, 1991) that are vastly utilized by the ABM. From this part of stream, visual information is treated incrementally in a cortical stages series (e.g., motion and orientation as local features in neurons

University of Malaya

at early levels, such as V1 (Hubel & Wiesel, 1968). In ABM (Wu et al., 2010)-based IncSFA (Kompella et al., 2011; 2012), basically all slow feature analysis methods can be an important tool for extracting slow features of modeling form processing streams in the ventral stream. As previously mentioned, the primary stage includes local (in V1 cell) and model detectors (Gabor like filters) in 16 (including eight preferred) orientations and by the proper scale depending on the receptive field (Dow et al., 1981; Schindler & Van Gool, 2008; Danafar et al., 2010). Conversely, wide invariance behavior is referred from the neuron response of the central nervous system (Sprekeler et al., 2007) (e.g., early vision complex cells phase invariance in (Hubel and Wiesel, 1968)) and in the hippocampal place cells of head direction invariance (Muller et al. 1994). Human action cycles are unlikely to possess invariant poses that are independent from environment, different lighting conditions, and pose of the actions. These invariance forms of actions can be an important criterion that represents form processing stream information. The slowness principle applies the perception and inferences on which neurons are trained by invariances by favoring slowly changing outputs, 2D (more details in (Muller et al., 1994)). A good implementation of this principle is SFA, which is the mean square from the temporal derivative of output and can be a good model for the physiological properties of complex cells in the visual cortex (Berkes & Wiskott, 2005) and other invariances in the visual system (Wiskott & Sejnowski, 2002). This combination can also function as snapshot detectors for human body shape models to determine the area IT (inferotemporal cortex) of monkeys where view-tuned neurons are located and the model of complex shapes are tuned (Logothetis et al., 1995), which is implemented by applying a synergetic neural network. IncSFA can function as view-tuned neurons in area IT and Snapshot neurons that provide independence in scale and form. The proposed model follows the modeling through slowest feature of ventral stream and adjustment through unsupervised learning methods. Optical flow was used by fast features and inferred with slowness information of the other pathway, which can represent a high level of integration of the snapshot neuron outcomes with information on motion pattern neuron information. In conclusion, the presented perspective recognizes the biological movement model following the original

University of Malaya

model (Giese & Poggio, 2003) and human visual system, which include two distinct pathways. The form and motion pathways are represented by ventral and dorsal processing streams. Slowness principle using ABM-based IncSFA has been utilized to extract form information (denoted as slow features). Optical flow generates motion information that is considered as fast features. The model analyzed the original recognition of biological movements in view of combining fast and slow features. Furthermore, the integration of these features, namely, slow and fast features, showed good performance in terms of recognition of human actions, which were evaluated through KTH and Weizmann human action for benchmarking.

5.4 Overview

This chapter concludes the whole processes have been done through presented approaches. The contributions we have achieved by these approaches are listed as below:

ABM in ventral stream

* It improved the abilities of ventral streams in terms of detection of human body shape (object recognition task improvement)(*it corresponded to the first objective and solved first stated problem*).

* It implies the interaction between pathways (*related to the second objective and second problem*).

Optical Flow division

* It improved the process in dorsal streams using the divisions of optical flow which helped more robust outcome (*Corresponding to objective 3, problem 2*).

* The optical flow division updated the interaction between two pathways which was performed in previous method (*Corresponded to objective 2 and problem 2*).

Slow features(SF) action Prototypes

* It improved the process of prototype generation, using slow feature analysis (*fulfilled the* 4^{th} *objective and problem 3*).

University of Malaya

* It updated interaction between the pathways(*it responded to objective 2 and solved problem 2*).

Dual Slow and Fast features

* It developed the computational model slow and fast features interaction (*Objective 2 and problem 2*).

* It used the slowness principle into the mechanism and converted the form process into a incremental procedure (Objective 4^{th} and problem 3).

5.5 Future work

Several open problems should be solved to allow us to develop the method further and use the proposed approaches and more applicable systems. This research involves various directions and makes a feasible system. One direction could involve further investigation of neuroscience evidence for a better perception of the biology of the brain. Evidently, the presented approaches carefully follow the existing evidence in the field, and further framework requires explicit details in biological studies.

The current framework includes different methods to cover the requirements of the different portions of the model. The computational load of these combinations should also be considered (currently the computational load is high which is considered as a limitation of the approaches). This step would allow adapting or extending into more complex analyses, which provide advancements in the current mechanism.

Finally, in terms of machine learning, another possibility would be to create another machine learning framework (such as deep learning, further fuzzy analysis, with respect to biological limitations) and modify the system from episodic recognition (particularly in the last approach) into the frame recognition. My particular interest would be on applying the framework to learn more complexity in biological movements as a means of carrying out further general analysis that is required for machine vision applications (considering the human visual system) and more complex datasets.

University of Malaya

REFERENCES

Adelson, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. JOSA A, 2(2), 284-299.

Aggarwal, J. K., & Cai, Q. (1997, June). Human motion analysis: A review. In Nonrigid and Articulated Motion Workshop, 1997. Proceedings., IEEE (pp. 90-102). IEEE.

Ahveninen, J., Huang, S., Ahlfors, S. P., Hämäläinen, M., Rossi, S., Sams, M., & Jääskelainen, I. P. (2016). Interacting parallel pathways associate sounds with visual identity in auditory cortices. NeuroImage, 124, 858-868.

Al-Absi, H. R., & Abdullah, A. B. (2010, June). Biologically inspired object recognition system. In Information Technology (ITSim), 2010 International Symposium in (Vol. 1, pp. 1-5). IEEE.

Ali, S., Basharat, A., & Shah, M. (2007, October). Chaotic invariants for human action recognition. In Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on (pp. 1-8). IEEE.

Alvarez, L., Deriche, R., Papadopoulo, T., & Sánchez, J. (2002). Symmetrical dense optical flow estimation with occlusions detection. In Computer Vision—ECCV 2002 (pp. 721-735). Springer Berlin Heidelberg.

Arbib, M. A. (2005). From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. Behavioral and brain sciences, 28(02), 105-124.

Babaeian, A., Babaee, M., Bayestehtashk, A., & Bandarabadi, M. (2015). Nonlinear subspace clustering using curvature constrained distances. Pattern Recognition Letters, 68, 118-125.

Babaeian, A., Bayestehtashk, A., & Bandarabadi, M. (2015). Multiple manifold clustering using curvature constrained path. PloS one, 10(9), e0137986.

Banquet, J. P., Gaussier, P., Quoy, M., Revel, A., & Burnod, Y. (2005). A hierarchy of associations in hippocampo-cortical systems: cognitive maps and navigation strategies. Neural Computation, 17(6), 1339-1384.

Becker, S., & Hinton, G. E. (1992). Self-organizing neural network that discovers surfaces in random-dot stereograms. Nature, 355(6356), 161-163.

Beintema, J. A., & Lappe, M. (2002). Perception of biological motion without local image motion. Proceedings of the National Academy of Sciences, 99(8), 5661-5663.

Berkes, P., & Wiskott, L. (2005). Slow feature analysis yields a rich repertoire of complex cell properties. Journal of Vision, 5(6), 9.

Billard, A., & Matarić, M. J. (2000, June). A biologically inspired robotic model for learning by imitation. In Proceedings of the fourth international conference on Autonomous agents (pp. 373-380). ACM.

Billard, A., & Matarić, M. J. (2001). Learning human arm movements by imitation: Evaluation of a biologically inspired connectionist architecture. Robotics and Autonomous Systems, 37(2), 145-160.

Blais, C., Arguin, M., & Marleau, I. (2009). Orientation invariance in visual shape perception. Journal of Vision, 9(2), 14-14.

Blake, R., & Shiffrar, M. (2007). Perception of human motion. Annu. Rev. Psychol., 58, 47-73.

Blank, M., Gorelick, L., Shechtman, E., Irani, M., & Basri, R. (2005, October). Actions as space-time

University of Malaya

shapes. In Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on (Vol. 2, pp. 1395-1402). IEEE.

Bohmer, W., Grünewälder, S., Nickisch, H., & Obermayer, K. (2011). Regularized sparse kernel slow feature analysis. In Machine Learning and Knowledge Discovery in Databases (pp. 235-248). Springer Berlin Heidelberg.

Bourke, M. M., & Fisher, D. G. (1996, September). A predictive fuzzy relational controller. In Fuzzy Systems, 1996., Proceedings of the Fifth IEEE International Conference on (Vol. 2, pp. 1464-1470). IEEE.

Bourke, M. M., & Fisher, D. G. (1998). Solution algorithms for fuzzy relational equations with max-product composition. Fuzzy sets and Systems, 94(1), 61-69.

Bray, A., & Martinez, D. (2003). Kernel-based extraction of slow features: Complex cells learn disparity and translation invariance from natural images. Advances in neural information processing systems, 269-276.

Breazeal, C., & Scassellati, B. (2002). Robots that imitate humans. Trends in cognitive sciences, 6(11), 481-487.

Bregler, C., Malik, J., & Pullen, K. (2004). Twist based acquisition and tracking of animal and human kinematics. International Journal of Computer Vision, 56(3), 179-194.

Brox, T., Bruhn, A., Papenberg, N., & Weickert, J. (2004). High accuracy optical flow estimation based on a theory for warping. In Computer Vision-ECCV 2004 (pp. 25-36). Springer Berlin Heidelberg.

Bruhn, A., Weickert, J., & Schnörr, C. (2005). Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods. International Journal of Computer Vision, 61(3), 211-231.

Cadieu, C. F., & Olshausen, B. A. (2012). Learning intermediate-level representations of form and motion from natural movies. Neural computation, 24(4), 827-866.

Cai, B., Xu, X., & Qing, C. (2014, July). Bio-inspired model with dual visual pathways for human action recognition. In Communication Systems, Networks Digital Signal Processing (CSNDSP), 2014 9th International Symposium on (pp. 271-276). IEEE.

Carlsson, S., & Sullivan, J. (2001, December). Action recognition by shape matching to key frames. In Workshop on Models versus Exemplars in Computer Vision (Vol. 1, p. 18).

Casile, A., & Giese, M. A. (2005). Critical features for the recognition of biological motion. Journal of vision, 5(4), 6-6.

Cavina-Pratesi, C., Large, M. E., & Milner, A. D. (2015). Reprint of: Visual processing of words in a patient with visual form agnosia: A behavioural and fMRI study. Cortex, 72, 97-114.

Cedras, C., Shah, M. (1995). Motion-based recognition a survey. Image and Vision Computing, 13(2), 129-155.

Chen, B., & Liu, X. (2005). Delay-Dependent Robust Control for T–S Fuzzy Systems With Time Delay. Fuzzy Systems, IEEE Transactions on, 13(4), 544-556.

Chin, T. J., & Suter, D. (2007). Incremental kernel principal component analysis. Image Processing, IEEE Transactions on, 16(6), 1662-1674.

Clarke, S., Bellmann, A., Meuli, R. A., Assal, G., & Steck, A. J. (2000). Auditory agnosia and auditory spatial deficits following left hemispheric lesions: evidence for distinct processing pathways. Neuropsychologia, 38(6), 797-807.

University of Malaya

Cloutman, L. L. (2013). Interaction between dorsal and ventral processing streams: where, when and how?. Brain and language, 127(2), 251-263.

Cohen, E. H., & Singh, M. (2006). Perceived orientation of complex shape reflects graded part decomposition. Journal of Vision, 6(8), 4-4.

Cook, J., Saygin, A. P., Swain, R., & Blakemore, S. J. (2009). Reduced sensitivity to minimum-jerk biological motion in autism spectrum conditions. Neuropsychologia, 47(14), 3275-3278.

Daems, A., & Verfaillie, K. (1999). Viewpoint-dependent priming effects in the perception of human actions and body postures. Visual cognition, 6(6), 665-693.

Danafar, S., Gretton, A., & Schmidhuber, J. (2010). Characteristic kernels on structured domains excel in robotics and human action recognition. In Machine Learning and Knowledge Discovery in Databases (pp. 264-279). Springer Berlin Heidelberg.

Daugman, J. G. (1980). Two-dimensional spectral analysis of cortical receptive field profiles. Vision research, 20(10), 847-856.

Daugman, J. G. (1985). Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. JOSA A, 2(7), 1160-1169.

Daugman, J. G. (1988). Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression. Acoustics, Speech and Signal Processing, IEEE Transactions on, 36(7), 1169-1179.

Dean, T., Washington, R., & Corrado, G. (2009, December). Recursive sparse, spatiotemporal coding. In Multimedia, 2009. ISM'09. 11th IEEE International Symposium on (pp. 645-650). IEEE.

Decety, J., Grezes, J. (1999). Neural mechanisms subserving the perception of human actions. Trends in cognitive sciences, 3(5), 172-178.

Demiris*, Y., & Johnson†, M. (2003). Distributed, predictive perception of actions: a biologically inspired robotics architecture for imitation and learning. Connection Science, 15(4), 231-243.

Demiris, Y., & Khadhouri, B. (2006). Hierarchical attentive multiple models for execution and recognition of actions. Robotics and autonomous systems, 54(5), 361-369.

Distler, C., Boussaoud, D., Desimone, R., & Ungerleider, L. G. (1993). Cortical connections of inferior temporal area TEO in macaque monkeys. Journal of Comparative Neurology, 334(1), 125-150.

Dollár, P., Rabaud, V., Cottrell, G., & Belongie, S. (2005, October). Behavior recognition via sparse spatio-temporal features. In Visual Surveillance and Performance Evaluation of Tracking and Surveillance, 2005. 2nd Joint IEEE International Workshop on (pp. 65-72). IEEE.

Dow, B. M., Snyder, A. Z., Vautin, R. G., & Bauer, R. (1981). Magnification factor and receptive field size in foveal striate cortex of the monkey. Experimental Brain Research, 44(2), 213-228.

Du, Q., Faber, V., & Gunzburger, M. (1999). Centroidal Voronoi tessellations: applications and algorithms. SIAM review, 41(4), 637-676.

Duncker, K. (1929). Uber induzierte bewegung. Psychologische Forschung, 12(1), 180-259.

Efros, A. A., Berg, A. C., Mori, G., & Malik, J. (2003, October). Recognizing action at a distance. In Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on (pp. 726-733). IEEE.

Eifuku, S., & Wurtz, R. H. (1998). Response to motion in extrastriate area MSTl: center-surround interactions. Journal of Neurophysiology, 80(1), 282-296.

University of Malaya

Escobar, M. J., & Kornprobst, P. (2012). Action recognition via bio-inspired features: The richness of center–surround interaction. Computer Vision and Image Understanding, 116(5), 593-605.

Escobar, M. J., Masson, G. S., Vieville, T., & Kornprobst, P. (2009). Action recognition using a bio-inspired feedforward spiking network. International Journal of Computer Vision, 82(3), 284-301.

Esmaili, S. S., Maghooli, K., & Nasrabadi, A. M. (2015). Combining two visual cortex models for robust face recognition. Optik-International Journal for Light and Electron Optics, 126(21), 2818-2824.

Fanti, C., Zelnik-Manor, L., & Perona, P. (2005, June). Hybrid models for human motion recognition. In Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on (Vol. 1, pp. 1166-1173). IEEE.

Fathi, A., & Mori, G. (2008, June). Action recognition by learning mid-level motion features. In Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on (pp. 1-8). IEEE.

Fei-Fei, L., Fergus, R., & Perona, P. (2006). One-shot learning of object categories. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 28(4), 594-611.

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. Cerebral cortex, 1(1), 1-47.

Fielding, K. H., Ruck, D. W. (1995). Recognition of moving light displays using hidden Markov models. Pattern Recognition, 28(9), 1415-1421.

Fleischer, F., Caggiano, V., Thier, P., & Giese, M. A. (2013). Physiologically inspired model for the visual recognition of transitive hand actions. The Journal of Neuroscience, 33(15), 6563-6580.

Franzius, M., Sprekeler, H., & Wiskott, L. (2007). Slowness and sparseness lead to place, head-direction, and spatial-view cells. PLoS Computational Biology, 3(8), e166.

Gölcü, D., & Gilbert, C. D. (2009). Perceptual learning of object shape. The Journal of Neuroscience, 29(43), 13621-13629.

Gallese, V., Fadiga, L., Fogassi, L., Rizzolatti, G. (1996). Action recognition in the premotor cortex. Brain, 119(2), 593-609.

Gao, J., Bao, J., Chen, D., Yang, Y., & Yang, X. (2001, April). Optical-electronic shape recognition system based on synergetic associative memory. In Photonics West 2001-Electronic Imaging (pp. 138-148). International Society for Optics and Photonics.

Gavrila, D. M. (1999). The visual analysis of human movement: A survey. Computer vision and image understanding, 73(1), 82-98.

Gawne, T. J., & Martin, J. M. (2002). Responses of primate visual cortical neurons to stimuli presented by flash, saccade, blink, and external darkening. Journal of neurophysiology, 88(5), 2178-2186.

Giese, M. A. (2014). Biological and body motion perception. Oxford Handbook of Perceptual Organization.

Giese, M. A., & Lappe, M. (2002). Measurement of generalization fields for the recognition of biological motion. Vision research, 42(15), 1847-1858.

Giese, M. A., & Poggio, T. (1999). Synthesis and recognition of biological motion patterns based on linear superposition of prototypical motion sequences. In Multi-View Modeling and Analysis of Visual Scenes, 1999.(MVIEW'99) Proceedings. IEEE Workshop on (pp. 73-80). IEEE.

Giese, M. A., & Poggio, T. (2000). Morphable models for the analysis and synthesis of complex motion patterns. International Journal of Computer Vision, 38(1), 59-73.

University of Malaya

Giese, M. A., & Poggio, T. (2003). Neural mechanisms for the recognition of biological movements. Nature Reviews Neuroscience, 4(3), 179-192.

Giese, M. A., & Vaina, L. M. (2001). Pathways in the analysis of biological motion: computational model and fMRI results. Perception, 30, ll6.

Giese, M. A., Jastorff, J., & Kourtzi, Z. (2002, January). Learning of the discrimination of artificial complex biological motion. In Perception, Vol 31; Abstract Supplement (Vol. 31).

Goddard, N. H. (1989, March). The interpretation of visual motion: Recognizing moving light displays. In Visual Motion, 1989., Proceedings. Workshop on (pp. 212-220). IEEE.

Goddard, N. H. (1989, March). The interpretation of visual motion: recognizing moving light displays. In Visual Motion, 1989., Proceedings. Workshop on (pp. 212-220). IEEE.

Goddard, N. H. (1992). The perception of articulated motion: recognizing moving light displays (No. TR-405). ROCHESTER UNIV NY DEPT OF COMPUTER SCIENCE.

Goodale, M. A., Milner, A. D. (1992). Separate visual pathways for perception and action. Trends in neurosciences, 15(1), 20-25.

Goodale, M. A., Westwood, D. A. (2004). An evolving view of duplex vision: separate but interacting cortical pathways for perception and action. Current opinion in neurobiology, 14(2), 203-211.

Goodale, M. A., Meenan, J. P., Bülthoff, H. H., Nicolle, D. A., Murphy, K. J., & Racicot, C. I. (1994). Separate neural pathways for the visual analysis of object shape in perception and prehension. Current Biology, 4(7), 604-610.

Gorelick, L., Blank, M., Shechtman, E., Irani, M., & Basri, R. (2007). Actions as space-time shapes. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 29(12), 2247-2253.

Grauman, K., & Darrell, T. (2005, October). The pyramid match kernel: Discriminative classification with sets of image features. In Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on (Vol. 2, pp. 1458-1465). IEEE.

Grezes, J., Fonlupt, P., Bertenthal, B., Delon-Martin, C., Segebarth, C., & Decety, J. (2001). Does perception of biological motion rely on specific brain regions?. Neuroimage, 13(5), 775-785.

Grossberg, S., Mingolla, E., & Viswanathan, L. (2001). Neural dynamics of motion integration and segmentation within and across apertures. Vision research, 41(19), 2521-2553.

Grossman, E. (2010). fMR-adaptation reveals invariant coding of biological motion on human STS. Frontiers in human neuroscience.

Grossman, E. D., & Blake, R. (2002). Brain areas active during visual perception of biological motion. Neuron, 35(6), 1167-1175.

Guha, T., & Ward, R. K. (2012). Learning sparse representations for human action recognition. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 34(8), 1576-1588.

Guthier, T., Willert, V., & Eggert, J. (2014). Topological sparse learning of dynamic form patterns. Neural computation.

Guthier, T., Willert, V., Schnall, A., Kreuter, K., & Eggert, J. (2013, August). Non-negative sparse coding for motion extraction. In Neural Networks (IJCNN), The 2013 International Joint Conference on (pp. 1-8). IEEE.

Haghighi, H., Abdollahi, F., & Gharibzadeh, S. (2016). Brain-inspired self-organizing modular structure

University of Malaya

to control human-like movements based on primitive motion identification. Neurocomputing, 173, 1436-1442.

Haken, H. (1996). Principles of brain functioning: A synergetic approach to brain activity, behaviour and cognition. Springer.

Haken, H. (2004). Synergetic computers and cognition: A top-down approach to neural nets (Vol. 50). Springer.

Harvey, B. M., & Dumoulin, S. O. (2016). Visual motion transforms visual space representations similarly throughout the human visual hierarchy. NeuroImage, 127, 173-185.

Hashimoto, W. (2003). Quadratic forms in natural images. Network: Computation in Neural Systems, 14(4), 765-788.

Hasson, U., Yang, E., Vallines, I., Heeger, D. J., & Rubin, N. (2008). A hierarchy of temporal receptive windows in human cortex. The Journal of Neuroscience, 28(10), 2539-2550.

Hesse, C., & Schenk, T. (2014). Delayed action does not always require the ventral stream: A study on a patient with visual form agnosia. Cortex, 54, 77-91.

Hill, H., & Pollick, F. E. (2000). Exaggerating temporal differences enhances recognition of individuals from point light displays. Psychological Science, 11(3), 223-228.

Hogg, T., Rees, D., & Talhami, H. (1995, November). Three-dimensional pose from two-dimensional images: a novel approach using synergetic networks. In Neural Networks, 1995. Proceedings., IEEE International Conference on (Vol. 2, pp. 1140-1144). IEEE.

Holcombe, A. O. (2009). Seeing slow and seeing fast: two limits on perception. Trends in cognitive sciences, 13(5), 216-221.

Hu, B., Kane-Jackson, R., & Niebur, E. (2016). A proto-object based saliency model in three-dimensional space. Vision research, 119, 42-49.

Huang, G. B., Chen, L., & Siew, C. K. (2006). Universal approximation using incremental constructive feedforward networks with random hidden nodes. Neural Networks, IEEE Transactions on, 17(4), 879-892.

Huang, G. B., Wang, D. H., & Lan, Y. (2011). Extreme learning machines: a survey. International Journal of Machine Learning and Cybernetics, 2(2), 107-122.

Huang, G. B., Zhu, Q. Y., & Siew, C. K. (2004, July). Extreme learning machine: a new learning scheme of feedforward neural networks. In Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on (Vol. 2, pp. 985-990). IEEE.

Huang, G. B., Zhu, Q. Y., & Siew, C. K. (2006). Extreme learning machine: theory and applications. Neurocomputing, 70(1), 489-501.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. The Journal of physiology, 195(1), 215-243.

Hung, C. P., Kreiman, G., Poggio, T., & DiCarlo, J. J. (2005). Fast readout of object identity from macaque inferior temporal cortex. Science, 310(5749), 863-866.

Ikizler, N., Duygulu, P. (2009). Histogram of oriented rectangles: A new pose descriptor for human action recognition. Image and Vision Computing, 27(10), 1515-1526.

Jellema, T., Baker, C. I., Wicker, B., & Perrett, D. I. (2000). Neural representation for the perception of the

University of Malaya

intentionality of actions. Brain and cognition, 44(2), 280-302.

Jhuang, H., Serre, T., Wolf, L., & Poggio, T. (2007, October). A biologically inspired system for action recognition. In Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on (pp. 1-8). Ieee.

Ji, S., Xu, W., Yang, M., & Yu, K. (2013). 3D convolutional neural networks for human action recognition. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 35(1), 221-231.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. Perception & psychophysics, 14(2), 201-211.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. Perception psychophysics, 14(2), 201-211.

Johansson, G. (1975). Visual motion perception. Scientific American.

Johansson, G. (1976). Spatio-temporal differentiation and integration in visual motion perception. Psychological research, 38(4), 379-393.

Johnson, M., & Demiris, Y. (2005, April). Hierarchies of coupled inverse and forward models for abstraction in robot action planning, recognition and imitation. In Proceedings of the AISB 2005 Symposium on Imitation in Animals and Artifacts (pp. 69-76).

Johnston, A., Arnold, D. H., & Nishida, S. (2006). Spatially localized distortions of event time. Current Biology, 16(5), 472-479.

Jolliffe, I. (1986). "Principal component analysis." Spring-verlag, New York.

Jones, J. P., & Palmer, L. A. (1987). An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. Journal of neurophysiology, 58(6), 1233-1258.

Jones, S., & Shao, L. (2013). Content-based retrieval of human actions from realistic video databases. Information Sciences, 236, 56-65.

Ju, L., Du, Q., & Gunzburger, M. (2002). Probabilistic methods for centroidal Voronoi tessellations and their parallel implementations. Parallel Computing, 28(10), 1477-1500.

Jung, C., Sun, T., & Gu, A. (2015). Content adaptive video denoising based on human visual perception. Journal of Visual Communication and Image Representation, 31, 14-25.

Karnath, H. O., Rüter, J., Mandler, A., & Himmelbach, M. (2009). The anatomy of object recognition—visual form agnosia caused by medial occipitotemporal stroke. The Journal of neuroscience, 29(18), 5854-5862.

Kay, K. N., Winawer, J., Rokem, A., Mezer, A., & Wandell, B. A. (2013). A two-stage cascade model of BOLD responses in human visual cortex. PLoS computational biology, 9(5), e1003079.

Kennedy, J. (2010). Particle swarm optimization. In Encyclopedia of Machine Learning (pp. 760-766). Springer US.

Kilner, J. M., Paulignan, Y., & Blakemore, S. J. (2003). An interference effect of observed biological movement on action. Current Biology, 13(6), 522-525.

Kompella, V. R., Luciw, M. D., & Schmidhuber, J. (2011, July). Incremental Slow Feature Analysis. In IJCAI (Vol. 11, pp. 1354-1359).

Kompella, V. R., Luciw, M., & Schmidhuber, J. (2012). Incremental slow feature analysis: Adaptive low-

University of Malaya

complexity slow feature updating from high-dimensional input streams. Neural computation, 24(11), 2994-3024.

Körding, K. P., Kayser, C., Einhäuser, W., & König, P. (2004). How are complex cell properties adapted to the statistics of natural stimuli?. Journal of neurophysiology, 91(1), 206-212.

Kourtzi, Z., & Kanwisher, N. (2000). Activation in human MT/MST by static images with implied motion. Journal of cognitive neuroscience, 12(1), 48-55.

Kozlowski, L. T., Cutting, J. E. (1977). Recognizing the sex of a walker from a dynamic point-light display. Perception Psychophysics, 21(6), 575-580.

Krigolson, O. E., Cheng, D., & Binsted, G. (2015). The role of visual processing in motor learning and control: Insights from electroencephalography. Vision research, 110, 277-285.

Kruger, N., Janssen, P., Kalkan, S., Lappe, M., Leonardis, A., Piater, J., ... & Wiskott, L. (2013). Deep hierarchies in the primate visual cortex: What can we learn for computer vision?. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 35(8), 1847-1871.

Kumar, S. (2004). Neural networks: a classroom approach. Tata McGraw-Hill Education.

Lades, M., Vorbruggen, J. C., Buhmann, J., Lange, J., von der Malsburg, C., Wurtz, R. P., & Konen, W. (1993). Distortion invariant object recognition in the dynamic link architecture. Computers, IEEE Transactions on, 42(3), 300-311.

Lampl, I., Ferster, D., Poggio, T., & Riesenhuber, M. (2004). Intracellular measurements of spatial integration and the MAX operation in complex cells of the cat primary visual cortex. Journal of neurophysiology, 92(5), 2704-2713.

Lange, J., & Lappe, M. (2006). A model of biological motion perception from configural form cues. The Journal of Neuroscience, 26(11), 2894-2906.

Lange, J., Georg, K. & Lappe, M. (2006). Visual perception of biological motion by form: A template-matching analysis. Journal of vision,6(8),6–6.

Laptev, I., & Lindeberg, T. (2006). Local descriptors for spatio-temporal recognition. In Spatial Coherence for Visual Motion Analysis (pp. 91-103). Springer Berlin Heidelberg.

Laycock, R., Crewther, S. G., & Crewther, D. P. (2007). A role for the 'magnocellular advantage'in visual impairments in neurodevelopmental and psychiatric disorders. Neuroscience & Biobehavioral Reviews, 31(3), 363-376.

Lee, G. C., & Loo, C. K. (2010, December). Facial pose estimation using modified synergetic computer. In Nature and Biologically Inspired Computing (NaBIC), 2010 Second World Congress on (pp. 322-327). IEEE.

Leek, E. C., Cristino, F., Conlan, L. I., Patterson, C., Rodriguez, E., & Johnston, S. J. (2012). Eye movement patterns during the recognition of three-dimensional objects: Preferential fixation of concave surface curvature minima. Journal of Vision, 12(1), 7-7.

Lehky, S. R., Kiani, R., Esteky, H., & Tanaka, K. (2011). Statistics of visual responses in primate inferotemporal cortex to object stimuli. Journal of Neurophysiology, 106(3), 1097-1117.

Lehky, S. R., Peng, X., McAdams, C. J., & Sereno, A. B. (2008). Spatial modulation of primate inferotemporal responses by eye position. PloS one, 3(10), e3492.

Levy, A., & Lindenbaum, M. (1998, October). Sequential Karhunen-Loeve basis extraction and its application to images. In Image Processing, 1998. ICIP 98. Proceedings. 1998 International Conference on

University of Malaya

(Vol. 2, pp. 456-460). IEEE.

Liang, N. Y., Huang, G. B., Saratchandran, P., & Sundararajan, N. (2006). A fast and accurate online sequential learning algorithm for feedforward networks. Neural Networks, IEEE Transactions on, 17(6), 1411-1423.

Libet, B., Wright, E. W., & Gleason, C. A. (1983). Preparation-or intention-to-act, in relation to pre-event potentials recorded at the vertex. Electroencephalography and clinical Neurophysiology, 56(4), 367-372.

Lin, C. J., Peng, C. C., & Lee, C. Y. (2011). Identification and prediction using neuro-fuzzy networks with symbiotic adaptive particle swarm optimization. Informatica: An International Journal of Computing and Informatics, 35(1), 113-122.

Liu, C. (2004). Gabor-based kernel PCA with fractional power polynomial models for face recognition. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 26(5), 572-581.

Liu, C. (2009). Beyond pixels: exploring new representations and applications for motion analysis (Doctoral dissertation, Massachusetts Institute of Technology).

Liu, C. (2009). Optical flow matlab/c++ code. h ttp://people. csail. mit. edu/celiu/OpticalFlow.

Liwicki, S., Zafeiriou, S., & Pantic, M. (2013). Incremental slow feature analysis with indefinite kernel for online temporal video segmentation. In Computer Vision–ACCV 2012 (pp. 162-176). Springer Berlin Heidelberg.

Loetamonphong, J., & Fang, S. C. (1999). An efficient solution procedure for fuzzy relation equations with max-product composition. Fuzzy Systems, IEEE Transactions on, 7(4), 441-445.

Logothetis, N. K., Pauls, J., & Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. Current Biology, 5(5), 552-563.

MacQueen, J. (1967, June). Some methods for classification and analysis of multivariate observations. In Proceedings of the fifth Berkeley symposium on mathematical statistics and probability (Vol. 1, No. 14, pp. 281-297).

Marčelja, S. (1980). Mathematical description of the responses of simple cortical cells*. JOSA, 70(11), 1297-1300.

Marr, D. (1982). Vision: A computational investigation into the human representation and processing of visual information.

Marr, D., Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. Proceedings of the Royal Society of London B: Biological Sciences, 200(1140), 269-294.

Mathe, S., & Sminchisescu, C. (2015). Actions in the eye: dynamic gaze datasets and learnt saliency models for visual recognition. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 37(7), 1408-1424.

Mather, G., Pavan, A., Marotti, R. B., Campana, G., & Casco, C. (2013). Interactions between motion and form processing in the human visual system. Frontiers in computational neuroscience, 7.

Matsumoto, Y., Takahashi, H., Murai, T., & Takahashi, H. (2015). Visual processing and social cognition in schizophrenia: Relationships among eye movements, biological motion perception, and empathy. Neuroscience research, 90, 95-100.

Maunsell, J. H., & Newsome, W. T. (1987). Visual processing in monkey extrastriate cortex. Annual review

University of Malaya

of neuroscience, 10(1), 363-401.

McLeod, P. (1996). Preserved and Impaired Detection of Structure From Motion by a'Motion-blind" Patient. Visual Cognition, 3(4), 363-392.

Meso, A. I., & Masson, G. S. (2015). Dynamic resolution of ambiguity during tri-stable motion perception. Vision research, 107, 113-123.

Milner, A. D., & Goodale, M. A. (2008). Two visual systems re-viewed. Neuropsychologia, 46(3), 774-785.

Moayedi, F., Azimifar, Z., & Boostani, R. (2015). Structured sparse representation for human action recognition. Neurocomputing, 161, 38-46.

Moeslund, T. B., & Granum, E. (2001). A survey of computer vision-based human motion capture. Computer vision and image understanding, 81(3), 231-268.

Moghaddam, B. (2002). Principal manifolds and probabilistic subspaces for visual recognition. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 24(6), 780-788.

Moutoussis, K., & Zeki, S. (1997). A direct demonstration of perceptual asynchrony in vision. Proceedings of the Royal Society of London. Series B: Biological Sciences, 264(1380), 393-399.

Muller, R. U., Bostock, E., Taube, J. S., & Kubie, J. L. (1994). On the directional firing properties of hippocampal place cells. Journal of Neuroscience, 14(12), 7235-7251.

Nayak, N. M., & Roy-Chowdhury, A. K. (2014, October). Learning a sparse dictionary of video structure for activity modeling. In Image Processing (ICIP), 2014 IEEE International Conference on (pp. 4892-4896). IEEE.

Neri, P., Morrone, M. C., & Burr, D. C. (1998). Seeing biological motion. Nature, 395(6705), 894-896.

Nickisch, H. (2006). "Eidesstattliche Erklärung."

Niebles, J. C., Wang, H., & Fei-Fei, L. (2008). Unsupervised learning of human action categories using spatial-temporal words. International journal of computer vision, 79(3), 299-318.

Nigmatullina, Y., Arshad, Q., Wu, K., Seemungal, B. M., Bronstein, A. M., & Soto, D. (2015). How imagery changes self-motion perception. Neuroscience, 291, 46-52.

Nowlan, S. J., & Sejnowski, T. J. (1995). A selection model for motion processing in area MT of primates. The Journal of Neuroscience, 15(2), 1195-1214.

Olshausen, B. A. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature, 381(6583), 607-609.

Oram, M. W., & Perrett, D. I. (1996). Integration of form and motion in the anterior superior temporal polysensory area (STPa) of the macaque monkey. Journal of neurophysiology.

Owens, J., Hunter, A., & Fletcher, E. (2002). A fast model-free morphology-based object tracking algorithm.

Perkins, D. (1995). Outsmarting IQ: The emerging science of learnable intelligence. Simon and Schuster.

Perlin, K. (1995). Real time responsive animation with personality. Visualization and Computer Graphics, IEEE Transactions on, 1(1), 5-15.

Perrett, D. I., Harries, M. H., Bevan, R., Thomas, S., Benson, P. J., Mistlin, A. J., ... & Ortega, J. E. (1989). Frameworks of analysis for the neural representation of animate objects and actions. Journal

University of Malaya

of Experimental Biology, 146(1), 87-113.

Perrett, D. I., Smith, P. A. J., Mistlin, A. J., Chitty, A. J., Head, A. S., Potter, D. D., ... & Jeeves, M. A. (1985). Visual analysis of body movements by neurones in the temporal cortex of the macaque monkey: a preliminary report. Behavioural brain research, 16(2), 153-170.

Pitzalis, S., Sdoia, S., Bultrini, A., Committeri, G., Di Russo, F., Fattori, P., ... & Galati, G. (2013). Selectivity to translational egomotion in human brain motion areas. PLoS One, 8(4), e60241.

Pollick, F. E., Paterson, H. M., Bruderlin, A., & Sanford, A. J. (2001). Perceiving affect from arm movement. Cognition, 82(2), B51-B61.

Poppe, R. (2010). A survey on vision-based human action recognition. Image and vision computing, 28(6), 976-990.

Pribram, K. H. (1991). Brain and perception: Holonomy and structure in figural processing. Psychology Press.

Prinz, W. (2014). Action representation: crosstalk between semantics and pragmatics. Neuropsychologia, 55, 51-56.

Pyles, J. A., Garcia, J. O., Hoffman, D. D., & Grossman, E. D. (2007). Visual perception and neural correlates of novel 'biological motion'. Vision research, 47(21), 2786-2797.

Rajesh, R., & Prakash, J. S. (2011). Extreme learning machines-a review and state-of-the-art. International journal of wisdom based computing, 1(1), 35-49.

Rangarajan, K., Allen, W., & Shah, M. (1992, January). Recognition using motion and shape. In Pattern Recognition, 1992. Vol. I. Conference A: Computer Vision and Applications, Proceedings., 11th IAPR International Conference on (pp. 255-258). IEEE.

Rao, C., Yilmaz, A., & Shah, M. (2002). View-invariant representation and recognition of actions. International Journal of Computer Vision, 50(2), 203-226.

Regan, D. (1989). Orientation discrimination for objects defined by relative motion and objects defined by luminance contrast. Vision research, 29(10), 1389-1400.

Richards, M., & Ventura, D. (2004). Choosing a starting configuration for particle swarm optimization.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. Nature neuroscience, 2(11), 1019-1025.

Riesenhuber, M., & Poggio, T. (2000). Models of object recognition. Nature neuroscience, 3, 1199-1204.

Riesenhuber, M., & Poggio, T. (2002). Neural mechanisms of object recognition. Current opinion in neurobiology, 12(2), 162-168.

Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. Nature Reviews Neuroscience, 2(9), 661-670.

Rodman, H. R., & Albright, T. D. (1989). Single-unit analysis of pattern-motion selective properties in the middle temporal visual area (MT). Experimental Brain Research, 75(1), 53-64.

Ross, D. A., Lim, J., Lin, R. S., & Yang, M. H. (2008). Incremental learning for robust visual tracking. International Journal of Computer Vision, 77(1-3), 125-141.

Rubin, E. (1927). Visuell wahrgenommene wirkliche Bewegungen. Zeitschrift für Psychologie, 103, 384-392.

University of Malaya

Ryoo, M. S., & Aggarwal, J. K. (2009, September). Spatio-temporal relationship match: Video structure comparison for recognition of complex human activities. In Computer vision, 2009 ieee 12th international conference on (pp. 1593-1600). IEEE.

Saleem, K. S., Suzuki, W., Tanaka, K., & Hashikawa, T. (2000). Connections between anterior inferotemporal cortex and superior temporal sulcus regions in the macaque monkey. The Journal of Neuroscience, 20(13), 5083-5101.

Santofimia, M. J., Martinez-del-Rincon, J., & Nebel, J. C. (2014). Episodic reasoning for vision-based human action recognition. The Scientific World Journal, 2014.

Schölkopf, B., Smola, A., & Müller, K. R. (1998). Nonlinear component analysis as a kernel eigenvalue problem. Neural computation, 10(5), 1299-1319.

Schüldt, C., Laptev, I., Caputo, B. (2004, August). Recognizing human actions: a local SVM approach. In Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on (Vol. 3, pp. 32-36). IEEE.

Schaal, S., Ijspeert, A., & Billard, A. (2003). Computational approaches to motor learning by imitation. Philosophical Transactions of the Royal Society of London B: Biological Sciences, 358(1431), 537-547.

Schenk, T. (2012). No dissociation between perception and action in patient DF when haptic feedback is withdrawn. The Journal of neuroscience, 32(6), 2013-2017.

Schenk, T. (2012). Response to Milner et al.: Grasping uses vision and haptic feedback. Trends in Cognitive Sciences, 16(5), 258.

Schindler, A., & Bartels, A. (2016). Motion parallax links visual motion areas and scene regions. NeuroImage, 125, 803-812.

Schindler, K., & Van Gool, L. (2008). Combining densely sampled form and motion for human action recognition. In Pattern Recognition (pp. 122-131). Springer Berlin Heidelberg.

Schindler, K., & Van Gool, L. (2008, June). Action snippets: How many frames does human action recognition require?. In Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on (pp. 1-8). IEEE.

Sereno, M. E. (1993). Neural computation of pattern motion: modeling stages of motion analysis in the primate visual cortex. Mit Press.

Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M., & Poggio, T. (2007). Robust object recognition with cortex-like mechanisms. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 29(3), 411-426.

Servos, P., Osu, R., Santi, A., & Kawato, M. (2002). The neural substrates of biological motion perception: an fMRI study. Cerebral Cortex, 12(7), 772-782.

Shabani, A. H., Zelek, J. S., & Clausi, D. A. (2010, May). Human action recognition using salient opponent-based motion features. In 2010 Canadian Conference Computer and Robot Vision (pp. 362-369). IEEE.

Shioiri, S., & Cavanagh, P. (1990). ISI produces reverse apparent motion. Vision Research, 30(5), 757-768.

Shioiri, S., & Matsumiya, K. (2009). Motion mechanisms with different spatiotemporal characteristics identified by an MAE technique with superimposed gratings. Journal of Vision, 9(5), 30.

Shu, N., Tang, Q., & Liu, H. (2014, July). A bio-inspired approach modeling spiking neural networks of

University of Malaya

visual cortex for human action recognition. In Neural Networks (IJCNN), 2014 International Joint Conference on (pp. 3450-3457). IEEE.

Sigal, L., Fleet, D. J., Troje, N. F., & Livne, M. (2010). Human attributes from 3d pose tracking. In Computer Vision–ECCV 2010 (pp. 243-257). Springer Berlin Heidelberg.

Simoncelli, E. P., & Heeger, D. J. (1998). A model of neuronal responses in visual area MT. Vision research, 38(5), 743-761.

Smith, A. T., & Snowden, R. J. (1994). Visual detection of motion. Academic Pr.

Somayeh, D., Alessandro, G., & Jürgen, S. (2010). Novel kernel-based recognizers of human actions. EURASIP Journal on Advances in Signal Processing, 2010.

Song, Y., Goncalves, L., Di Bernardo, E., & Perona, P. (2001). Monocular perception of biological motion in johansson displays. Computer Vision and Image Understanding, 81(3), 303-327.

Sprekeler, H., Michaelis, C., & Wiskott, L. (2007). Slowness: An objective for spike-timing–dependent plasticity?. PLoS Computational Biology, 3(6), e112.

Stone, J. (1983). Parallel processing in the visual system: the classification of retinal ganglion cells and its impact on the neurobiology of vision (p. 438). New York and London: Plenum Press.

Stone, J. V. (1996). Learning perceptually salient visual parameters using spatiotemporal smoothness constraints. Neural Computation, 8(7), 1463-1492.

Strasburger, H., Rentschler, I., & Jüttner, M. (2011). Peripheral vision and pattern recognition: A review. Journal of vision, 11(5), 13-13.

Sun, D., Roth, S., & Black, M. J. (2010, June). Secrets of optical flow estimation and their principles. In Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on (pp. 2432-2439). IEEE.

Sun, J., Fang, W., Wu, X., Palade, V., & Xu, W. (2012). Quantum-behaved particle swarm optimization: Analysis of individual particle behavior and parameter selection. Evolutionary Computation, 20(3), 349-393.

Syrris, V., & Petridis, V. (2011). A lattice-based neuro-computing methodology for real-time human action recognition. Information Sciences, 181(10), 1874-1887.

Tadin, D. (2015). Suppressive mechanisms in visual motion processing: From perception to intelligence. Vision research, 115, 58-70.

Tanaka, K. (1996). Inferotemporal cortex and object vision. Annual review of neuroscience, 19(1), 109-139.

Tarr, M. J., & Bülthoff, H. H. (1998). Image-based object recognition in man, monkey and machine. Cognition, 67(1), 1-20.

Theusner, S., de Lussanet, M., & Lappe, M. (2014). Action recognition by motion detection in posture space. The Journal of Neuroscience, 34(3), 909-921.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. nature, 381(6582), 520-522.

Thurman, S. M., & Lu, H. (2013). Complex interactions between spatial, orientation, and motion cues for biological motion perception across visual space. Journal of vision, 13(2), 8.

Tian, Y., Ruan, Q., An, G., & Xu, W. (2015). Context and locality constrained linear coding for human

University of Malaya

action recognition. Neurocomputing, 167, 359-370.

Tlapale, É., Dosher, B. A., Lu, Z. L. (2015). Construction and evaluation of an integrated dynamical model of visual motion perception. Neural Networks, 67, 110-120.

Troje, N. F. (2002). Decomposing biological motion: A framework for analysis and synthesis of human gait patterns. Journal of vision, 2(5), 2-2.

Troje, N. F., & Westhoff, C. (2006). The inversion effect in biological motion perception: evidence for a "life detector"?. Current Biology, 16(8), 821-824.

Tsai, P. S., Shah, M., Keiter, K., Kasparis, T. (1993). Cyclic motion detection. Computer Science Technical Report, University of Central Florida, Orlando.

Turaga, P., Chellappa, R., Subrahmanian, V. S., & Udrea, O. (2008). Machine recognition of human activities: A survey. Circuits and Systems for Video Technology, IEEE Transactions on, 18(11), 1473-1488.

Ungerleider, L. G. (1982). Two cortical visual systems. Analysis of visual behavior, 549-586.

Ungerleider, L. G., & Pasternak, T. (2004). Ventral and dorsal cortical processing streams. The visual neurosciences, 1(34), 541-562.

Vaina, L. M., Solomon, J., Chowdhury, S., Sinha, P., & Belliveau, J. W. (2001). Functional neuroanatomy of biological motion perception in humans. Proceedings of the National Academy of Sciences, 98(20), 11656-11661.

Valstar, M. F., & Pantic, M. (2006, July). Biologically vs. logic inspired encoding of facial actions and emotions in video. In Multimedia and Expo, 2006 IEEE International Conference on (pp. 325-328). IEEE.

Vander Wyk, B. C., Voos, A., & Pelphrey, K. A. (2012). Action representation in the superior temporal sulcus in children and adults: an fMRI study. Developmental cognitive neuroscience, 2(4), 409-416.

Venezia, J. H., Fillmore, P., Matchin, W., Isenberg, A. L., Hickok, G., & Fridriksson, J. (2016). Perception drives production across sensory modalities: A network for sensorimotor integration of visual speech. NeuroImage, 126, 196-207.

Wachter, S., & Nagel, H. H. (1997, June). Tracking of persons in monocular image sequences. In Nonrigid and Articulated Motion Workshop, 1997. Proceedings., IEEE (pp. 2-9). IEEE.

Wang, D., & Huang, G. B. (2005, August). Protein sequence classification using extreme learning machine. In Neural Networks, 2005. IJCNN'05. Proceedings. 2005 IEEE International Joint Conference on (Vol. 3, pp. 1406-1411). IEEE.

Wang, L., & Suter, D. (2007, June). Recognizing human activities from silhouettes: Motion subspace and factorial discriminative graphical model. In Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on (pp. 1-8). IEEE.

Wang, Y., & Mori, G. (2009). Human action recognition by semilatent topic models. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 31(10), 1762-1774.

Wang, Y., & Zhang, K. (2010). Decomposing the spatiotemporal signature in dynamic 3D object recognition. Journal of vision, 10(10), 23-23.

Ward, E. J., MacEvoy, S. P., & Epstein, R. A. (2010). Eye-centered encoding of visual space in scene-selective regions. Journal of vision, 10(14), 6-6.

Webb, B. S., Roach, N. W., & Peirce, J. W. (2008). Masking exposes multiple global form mechanisms.

University of Malaya

Journal of vision, 8(9), 16-16.

Weinland, D., Ronfard, R., & Boyer, E. (2011). A survey of vision-based methods for action representation, segmentation and recognition. Computer Vision and Image Understanding, 115(2), 224-241.

Whitney, D., & Murakami, I. (1998). Latency difference, not spatial extrapolation. Nature neuroscience, 1(8), 656-657.

Whitwell, R. L., Milner, A. D., Cavina-Pratesi, C., Barat, M., & Goodale, M. A. (2015). Patient DF's visual brain in action: visual feedforward control in visual form agnosia. Vision research, 110, 265-276.

Whitwell, R. L., Milner, A. D., Cavina-Pratesi, C., Byrne, C. M., & Goodale, M. A. (2014). DF's visual brain in action: the role of tactile cues. Neuropsychologia, 55, 41-50.

Wiley, D. J., & Hahn, J. K. (1997). Interpolation synthesis of articulated figure motion. Computer Graphics and Applications, IEEE, 17(6), 39-45.

Willert, V., & Eggert, J. (2009, September). A stochastic dynamical system for optical flow estimation. In Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on (pp. 711-718). IEEE.

Willert, V., & Eggert, J. (2011). Modeling short-term adaptation processes of visual motion detectors. Neurocomputing, 74(9), 1329-1339.

Willert, V., Toussaint, M., Eggert, J., & Körner, E. (2007, December). Uncertainty optimization for robust dynamic optical flow estimation. In Machine Learning and Applications, 2007. ICMLA 2007. Sixth International Conference on (pp. 450-457). IEEE.

Wiskott, L. (2003). Slow feature analysis: A theoretical analysis of optimal free responses. Neural Computation, 15(9), 2147-2177.

Wiskott, L., & Sejnowski, T. (2002). Slow feature analysis: Unsupervised learning of invariances. Neural computation, 14(4), 715-770.

Wong, W. K., Loo, C. K., Lim, W. S., & Tan, P. N. (2010). Thermal condition monitoring system using log-polar mapping, quaternion correlation and max-product fuzzy neural network classification. Neurocomputing, 74(1), 164-177.

Wu, Y. N., Si, Z., Gong, H., & Zhu, S. C. (2010). Learning active basis model for object detection and recognition. International journal of computer vision, 90(2), 198-235.

Yacoob, Y., & Black, M. J. (1998, January). Parameterized modeling and recognition of activities. In Computer Vision, 1998. Sixth International Conference on (pp. 120-127). IEEE.

Yamamoto, K., & Miura, K. (2015). Effect of motion coherence on time perception relates to perceived speed. Vision research.

Yamato, J., Ohya, J., & Ishii, K. (1992, June). Recognizing human action in time-sequential images using hidden markov model. In Computer Vision and Pattern Recognition, 1992. Proceedings CVPR'92., 1992 IEEE Computer Society Conference on (pp. 379-385). IEEE.

Yau, J. M., Pasupathy, A., Fitzgerald, P. J., Hsiao, S. S., & Connor, C. E. (2009). Analogous intermediate shape coding in vision and touch. Proceedings of the National Academy of Sciences, 106(38), 16457-16462.

Yousefi, B., Loo, C. K., & Memariani, A. (2013, April). Biological inspired human action recognition. In Robotic Intelligence In Informationally Structured Space (RiiSS), 2013 IEEE Workshop on (pp. 58-65). IEEE.

University of Malaya

Yousefi, B., & Loo, C. K. (2014a). Development of biological movement recognition by interaction between active basis model and fuzzy optical flow division. The Scientific World Journal, 2014.

Yousefi, B. & Loo, C. K. (2014b). Bio-Inspired Human Action Recognition using Hybrid Max-Product Neuro-Fuzzy Classifier and Quantum-Behaved PSO, Informatica 1574.

Yousefi, B., & Loo, C. K. (2014c). Comparative study on interaction of form and motion processing streams by applying two different classifiers in mechanism for recognition of biological movement. The Scientific World Journal, 2014.

Yousefi, B., & Loo, C. K. (2014d). Slow feature Action Prototypes Effect Assessment in mechanism for Recognition of Biological Movement Ventral Stream. Int. J. of Bio-Inspired Computation (in-press).

Yousefi, B. & Loo, C. K. (2016) Applied Soft Computing (Accepted with revision-Applied Soft Computing Journal).

Yousefi, B. & Loo, C. K. (2016) Mechanism for the recognition of biological movements: A review. Under submission.

Yue, S., & Rind, F. C. (2013). Redundant neural vision systems—Competing for collision recognition roles. Autonomous Mental Development, IEEE Transactions on, 5(2), 173-186.

Yuille, A. L., Hallinan, P. W., & Cohen, D. S. (1992). Feature extraction from faces using deformable templates. International journal of computer vision, 8(2), 99-111.

Zadeh, L. A. (1965). Fuzzy sets. Information and control, 8(3), 338-353.

Zhang, Y., & Weng, J. (2001). Convergence analysis of complementary candid incremental principal component analysis. Michigan State University.

Zhang, Z., & Tao, D. (2012). Slow feature analysis for human action recognition. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 34(3), 436-450.

Zhou, H.-M. and G.-B. Huang (2012). "Kernelled ELM Algorithms ". from http://www.ntu.edu.sg/home/egbhuang/elm$_{c}odes.html$.

Zhu, S. C., & Mumford, D. (2007). A stochastic grammar of images. Now Publishers Inc.

Zito, T., Wilbert, N., Wiskott, L., & Berkes, P. (2008). Modular toolkit for Data Processing (MDP): a Python data processing framework. Frontiers in neuroinformatics, 2.

University of Malaya