HUMAN ACTION RECOGNITION UTILIZING VARIATIONS IN SKELETON DIMENSIONS

By

Mona Mohamed Mahmoud Moussa

A Thesis Submitted to the
Faculty of Engineering at Cairo University
in Partial Fulfillment of the
Requirements for the Degree of
DOCTOR OF PHILOSOPHY
in

Computer Engineering

FACULTY OF ENGINEERING, CAIRO UNIVERSITY
GIZA, EGYPT
2016

HUMAN ACTION RECOGNITION UTILIZING VARIATIONS IN SKELETON DIMENSIONS

By

Mona Mohamed Mahmoud Moussa

A Thesis Submitted to the
Faculty of Engineering at Cairo University
in Partial Fulfillment of the
Requirements for the Degree of
MASTER OF SCIENCE (or DOCTOR OF PHILOSOPHY)
in
Computer Engineering

Under the Supervision of

Prof. Dr. Magda B. Fayek	Prof. Dr. Elsayed E. Hemayed
……………………….	…………………………..
Professor Computer Engineering Department Faculty of Engineering, Cairo University	Professor Computer Engineering Department Faculty of Engineering, Cairo University

Assoc. Prof. Heba A. Elnemr

……………………………
Associate Professor
Computers and Systems Department
Electronics Research Institute

FACULTY OF ENGINEERING, CAIRO UNIVERSITY
GIZA, EGYPT
2016

HUMAN ACTION RECOGNITION UTILIZING VARIATIONS IN SKELETON DIMENSIONS

By

Mona Mohamed Mahmoud Moussa

A Thesis Submitted to the
Faculty of Engineering at Cairo University
in Partial Fulfillment of the
Requirements for the Degree of
DOCTOR OF PHILOSOPHY
in
COMPUTER ENGINEERING

Approved by the
Examining Committee

____
Prof. Dr. Magda B. Fayek, Thesis Main Advisor

____
Prof. Dr. Elsayed E. Hemayed, Member

____
Assoc. Prof. Heba A. Elnemr, Member
Associate professor, Computers and Systems Department, Electronics Research Institute

____
Prof. Dr. Reda Abd Elwahab Ahmed, Examiner
Professor, Faculty of Computers and Information, Cairo university

____
Prof. Dr. Samia Abdel Razek Mashaly, Examiner
Professor, Computers and Systems Department, Electronics Research Institute

FACULTY OF ENGINEERING, CAIRO UNIVERSITY
GIZA, EGYPT
2016

Engineer's Name: Mona Mohamed Mahmoud Moussa
Date of Birth: 2/3/1982
Nationality: Egyptian
E-mail: mona.moussa@gmail.com
Phone: 01001479222
Address: 9055 El Merag city- El Maadi- Cairo
Registration Date: 1/3/2010
Awarding Date: / /2016
Degree: Doctor of Philosophy
Department: Computer Engineering

Supervisors:

Prof. Magda B. Fayek
Prof. Elsayed E. Hemayed
Assoc. Prof. Heba A. Elnemr

Examiners:

Prof. Magda B. Fayek (Thesis main advisor)
Prof. Elsayed E. Hemayed (Member)
Assoc. Prof. Heba A. Elnemr (Member)
Associate professor, Computers and Systems Department, Electronics Research Institute
Prof. Reda Abdel Wahab Ahmed (Examiner)
Professor, Faculty of Computers and Information, Cairo university
Prof. Samia Abdel Razek Mashaly (Examiner)
Professor, Computers and Systems Department, Electronics Research Institute

Title of Thesis:
Human action recognition utilizing variations in skeleton dimensions

Key Words:
Human action recognition, skeletal data, action high-level representation, computer vision

Summary:
The proposed work is a human action recognition system that relies on the amount and shape of change of different body parts to recognize a given action in a recorded video. The system can deal with videos recorded using traditional cameras as well as depth-sensing cameras. Newly proposed features are extracted and encoded to describe the visual way of change of human body parts. The first step in the technique is skeleton extraction for the subject person. Then, novel features are extracted from this skeleton and encoded to obtain a limited short length code that represents the whole video. Training and testing step were performed using benchmark datasets, namely: KTH, Weizmann, Berkeley, MSR-action3D.

ACKNOWLEDGMENTS

I would like to express my gratitude to all those who gave me the possibility to complete this thesis. Above all, I would like to offer my heartfelt thankfulness to Prof. Magda B. Fayek, Prof. Elsayed E. Hemayed and Assoc. Prof. Dr. Heba A. Elnemr who made this thesis possible. Throughout, they provided me with valuable support and gave generously of their time and experience. I would like to extend my gratitude to all people who taught me during the past years.

I would like also to thank my parents who were my first instructors in this world. Since I was a child, they have been doing their best to help and encourage me towards success. Feeling especially indebted to them, I wish them all the best and do hope that one day, I will be able to equally reward them (if an equal reward is ever possible). I also want to thank my family for their constant support.

PUBLICATIONS

Moussa, M. M., Hamayed, E., Fayek, M. B., & El Nemr, H. A. 2013. An enhanced method for human action recognition. Journal of Advanced Research, 6(2) (pp. 163-169)

Moussa, M. M., Hamayed, E., El Nemr, H. A. , & Fayek, M. B. Human action recognition utilizing variations in skeleton dimensions. Under review

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

NOMENCLATURE

AdaBoost	Adaptive boosting
ADI	Average depth image
BM-HAR	Body modeling-human action recognition
BoVW	Bag of Visual Words
BST	Binary shape templates
CFG	Context-free grammars
DBN	Dynamic Bayesian networks
DDI	Depth difference image
DoG	Difference of Gaussians
FIS	Fuzzy Inference System
HAR	Human action recognition
HMM	Hidden Markov model
HOG	Histogram of oriented gradients
LD-HAR	Local descriptors-human action recognition
LDA	Latent dirichlet allocation
LoG	Laplacian of gaussian
MAP	Maximum a posteriori probability
MHI	Motion history image
MHT	Motion history templates
MLE	Maximum likelihood estimation
MMI	Maximization of mutual information
MoCap	Motion capture
MoSift	Motion-scale invariant feature transform
pLSA	Probabilistic latent semantic analysis
SCFGs	Stochastic context-free grammars
SIFT	Scale invariant feature transform
SMIJ	Sequence of the most informative joints
SVM	Support vector machine

ABSTRACT

This thesis presents an integrated automatic human action recognition system that distinguishes between different actions using a new set of features based on global variation in the visual appearance of the subject body. The proposed technique utilizes the changes in human body dimensions, during performing an action, to extract this feature set. These dimension variations are calculated from the human body skeleton performing the action to be recognized. The skeleton can be extracted from a video captured using traditional 2D cameras or depth sensing cameras. Finally, a multi-class linear support vector machine is employed in the classification stage.

Experiments are conducted on Weizmann, Berkeley MHAD, and MSR-Action3D datasets. The results show that the proposed technique achieves an accuracy of 98.9% for Weizmann, 99.63% for Berkeley MHAD, and 94.3% for MSR-Action3D. Moreover, a cross-dataset experiment is held to ensure the generality of the proposed technique, where the system is trained using Berkeley MHAD dataset and tested using MSR-Action3D, achieving accuracy of 88.76%.

The thesis includes as well an experiment that was held to recognize human activities using local descriptors by extracting a group of interesting points from each frame of the video. Scale-invariant feature transform (SIFT) algorithm is used to obtain the group of interesting points. An adapting step is performed to limit the number of interesting points depending on the degree of details. Then, the well-known approach Bag of Visual Words (BoVW) is applied with a new proposed normalization technique. The proposed normalization technique improves the results remarkable. Finally, a multi-class linear support vector machine is used for classification.

When utilizing local descriptors, experiments were held on the KTH and Weizmann datasets, achieving an accuracy of 96.66% for Weizmann and 97.89% for KTH.

CHAPTER 1: INTRODUCTION

1.1. Introduction

Video analysis of human activities is an area with increasingly significant consequences from security and surveillance to entertainment and personal archiving. Human motion analysis can be categorized into three groups: human activity recognition, human motion tracking, and body parts movement analysis.

- **Human activity recognition:** recognizes the actions of one or more person of a group of observations on the person's activity and the surrounding environmental conditions. The aim of this branch is to support different applications (as computer vision and surveillance applications); also, it is connected to a number of fields of study such as human-computer interaction, medicine, and sociology.
- **Human motion tracking:** here the objective is to correlate target objects in consecutive video frames. The correlation is a difficult task if the objects are moving fast relative to the frame rate or if there are changes in the object orientation over time. Two of the standard target representations and localization algorithms are:
  - Kernel-based tracking: an iterative localization process based on maximizing the similarity measure
  - Contour tracking: iteratively evolves an initial contour initialized from each frame to its position in the successive frame. Here contour tracking directly evolves the contour by minimizing the contour energy using gradient descent.
- **Motion analysis of body parts:** tracks the location and orientation of body parts, it becomes an investigative and diagnostic tool in some areas as medicine, sports, video surveillance and kinesiology (the scientific study of human movement).

1.2. Problem Statement

The aim of the presented work is to automatically recognize actions of one or more persons using observations on the their activities. Human action recognition is an important branch of computer vision and pattern recognition due to its broad range of applications such as surveillance video, robot vision, content-based video retrieval, automatic video indexing and retrieval, and human-computer interaction. Videos can be recorded by either 2D cameras or 3D cameras.

Recently RGBD cameras such as Microsoft Kinect are used to detect human activity where they add an extra dimension, which is the depth that the traditional 2D cameras fail to provide. Sensor information captured through depth cameras can be used to generate real-time human skeleton model describing different body positions, this model can be further used to define human activities.

Vision-based human action recognition is the task of labeling videos containing human motion with action classes to identify a person′s action. The task is challenging due to variations in motion performance, recording settings and inter-personal differences. Some methods have been used to achieve vision-based action recognition such as optical flow, Kalman filtering, Hidden Markov models. In addition, multiple aspects are considered on this topic as single agent tracking, group tracking, and detecting dropped objects.

1.3. Human activity recognition hierarchy

Figure 1.1 presents a hierarchal taxonomy for human action recognition as proposed by Aggarwal [1]. Action recognition process is mainly divided into two branches single-layered approaches and hierarchical approaches.

Single-layered approaches includes:

- Space-time approaches
  - Space-time volume
  - Trajectories
  - Space-time features
- Sequential approaches
  - Exemplar-based
  - State-based

Hierarchical approaches includes:

- Statistical
- Syntactic
- Description-based

It is worth noting that, the above taxonomy is not sharply divided; a technique can combine more than one approach to perform human action recognition.

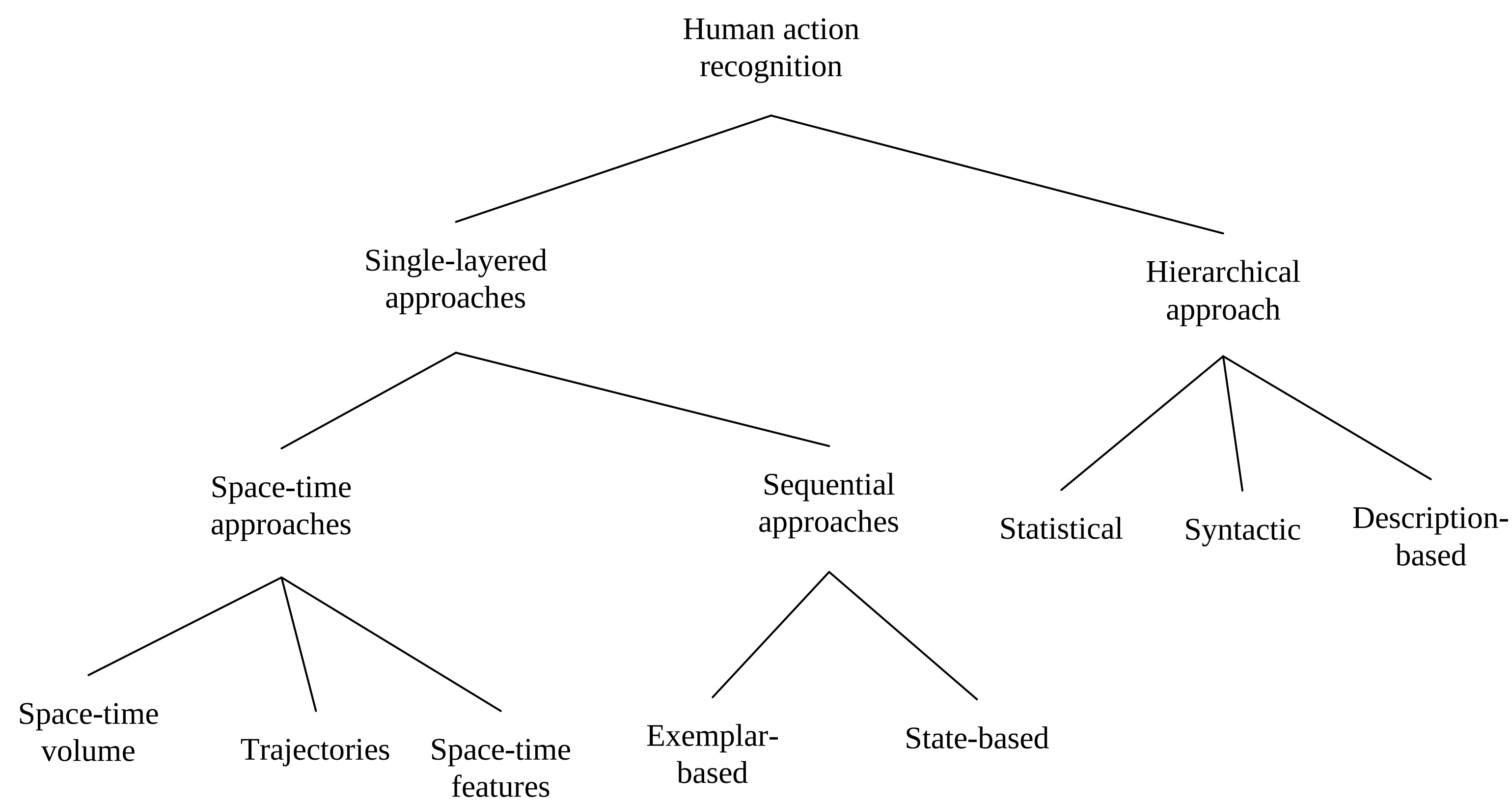

Figure 1.1: Human activity recognition hierarchy