

DOCTORAL THESIS

Human motion detection and action recognition

Liu, Chang

Date of Award:
2010

[Link to publication](#)

General rights

Copyright and intellectual property rights for the publications made accessible in HKBU Scholars are retained by the authors and/or other copyright owners. In addition to the restrictions prescribed by the Copyright Ordinance of Hong Kong, all users and readers must also observe the following terms of use:

- Users may download and print one copy of any publication from HKBU Scholars for the purpose of private study or research
- Users cannot further distribute the material or use it for any profit-making activity or commercial gain
- To share publications in HKBU Scholars with others, users are welcome to freely distribute the permanent URL assigned to the publication

Human Motion Detection and Action Recognition

LIU Chang

A thesis submitted in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

Principal Supervisor: Prof. YUEN Pong Chi

Hong Kong Baptist University

January 2010

Abstract

Human action analysis has been received increasing attentions from researchers in the last decade. The objective of human action analysis is to detect and recognize human actions from videos so that the computer system is able to understand human behaviors and make further semantic description of the scene. Computer systems understand human actions from the scene involving two major steps: human motion detection and human action recognition. There are challenges in both of these two research areas. Generally speaking, the main challenge of human motion detection from video is to detect humans with different moving speeds in complex background clusters, and under different illumination changes. For human action recognition, there have been a number of action classifiers proposed, but the crucial factors are how to give effective and efficient representations of high dimensional human actions for categorization or recognition, and how to employ unlabeled video data to train and enhance the performance of the action recognition system. In this thesis, new algorithms are proposed from spatio-temporal based approach to solve these problems in human action detection and recognition.

This thesis proposes to employ the visual saliency for human motion detection via direct analysis from videos. Object saliency is represented by an Information Saliency Map (ISM), which is calculated from spatio-temporal volumes. Both spatial saliency and temporal saliency are calculated and a dynamic fusion method is developed for

incorporation. Principal component analysis and kernel density estimation are used to develop an efficient information theoretic based procedure for constructing the ISM. The ISM is then used for measuring visual saliency and detecting foreground objects. Experimental results on publicly available video surveillance databases show that the proposed method is robust for both detecting fast and slow moving object under illumination changes.

This thesis further explores the use of ISM for human action recognition. A Boosting EigenAction algorithm is proposed to recognize human action from video. A human action is segmented into a set of primitive periodic motion cycles from information saliency curve. Each cycle of motion is represented by a Saliency Action Unit (SAU), which is used to determine the EigenAction using principal component analysis. A human action classifier is developed using multi-class Adaboost algorithm with Bayesian hypothesis as the weak classifier. Given a human action video sequence, the proposed method effectively and efficiently locates the SAU(s) in the video, trains an action classifier, and recognizes the human actions by categorizing these SAU(s).

This thesis develops a semi-supervised algorithm for human action recognition, as labeled data are costly to obtain whereas unlabeled data are abundantly available. A boosted Co-Training algorithm for human action recognition is proposed. Two confidence measures namely inter-view confidence and intra-view confidence are proposed and estimated to solve the two main problems in the Co-Training method, namely view dependency and view insufficiency, and are dynamically fused into one semi-supervised learning process. Mutual information measure is employed to quantify the inter-view uncertainty and measure the independence among respective views. Intra-view confidence is estimated from boosted hypotheses to measure the total data inconsistency among labeled data and unlabeled data. Two discriminative views from temporal and spatial information of the video, namely action saliency view and action eigen-projection view, are proposed as input data in practice. Given a small set of labeled videos and

a large set of unlabeled videos, the proposed semi-supervised learning algorithm trains a classifier by maximizing the inter-view confidence and intra-view confidence, and dynamically incorporating unlabeled data into the labeled data set, the performance of the classifier will be improved in each iteration. The final classifier is able to classify different human actions with action video clips as the input data.

The proposed methods have been extensively evaluated using publicly available databases such as CAVIAR, PETS, OTCBVS-Bench video surveillance databases, and Weizmann, KTH human action recognition databases. Comparison between the proposed methods and existing state-of-the-art methods are also reported in this thesis.

In short, the major contributions of this thesis are summarized as follows:

- An Information Saliency Map (ISM) is proposed to detect human motions. The ISM is robust for object detection under illumination changes.
- Salient Action Unit (SAU) is proposed to represent primitive human actions, the SAU can be efficiently extracted from information saliency curve, and used for training the human action classifier.
- A boosted Co-Training algorithm for human action recognition is proposed. Inter-view confidence and intra-view confidence are proposed and estimated to solve the view dependency and view insufficiency problems in Co-Training.

Table of Contents

Declaration	i
Abstract	ii
Acknowledgements	v
Table of Contents	vi
List of Tables	xiii
List of Figures	xv
List of Symbols	xxiii
List of Abbreviation	xxiv
1 Introduction	1

1.1	Background	1
1.2	Human Motion Detection	5
1.2.1	Existing Human Motion Detection Methods	5
1.2.2	Existing Problems	8
1.3	Human Action Recognition	8
1.3.1	Existing Human Action Recognition Methods	9
1.3.2	Existing Problems	11
1.4	Motivations	12
1.5	Contributions of this Thesis	13
1.6	Thesis Overview	15
2	Review on Existing Works and Public Available Databases	17
2.1	Review on Human Motion Detection Algorithms	18
2.1.1	Distribution based Approach	19
2.1.2	Orientation based Approach	21
2.1.3	Contour based Approach	22
2.1.4	Region-of-interest based Approach	23
2.2	Review on Human Action Recognition Algorithms	25

2.2.1	Space-Time based Approach	25
2.2.2	Human Model based Approach	28
2.2.3	State-Space based Approach	30
2.3	Testing Databases Used in This Thesis	32
2.3.1	Visual Surveillance Databases	32
2.3.1.1	CAVIAR	32
2.3.1.2	PETS	33
2.3.1.3	OTCBVS-Bench	34
2.3.2	Human Action Recognition Databases	35
2.3.2.1	Weizmann	35
2.3.2.2	KTH	36
3	Human Motion Detection Using Information Theoretic Spatio-Temporal Saliency	38
3.1	Introduction	38
3.2	Proposed Method	40
3.2.1	Information Content	40
3.2.2	Computing Information Saliency Map	41

3.2.2.1	Computing Temporal Saliency	43
3.2.2.2	Computing Spatial Saliency	46
3.2.2.3	Spatio-Temporal Saliency Fusion	47
3.2.3	Analysis of Illumination Effect on ISM	49
3.3	Experimental Results	51
3.3.1	Experimental Settings	52
3.3.2	Evaluation of the Proposed Method	52
3.3.2.1	Results on CAVIAR Database	52
3.3.2.2	ISM Parameter Analysis	57
3.3.3	Comparing the Proposed Method with Existing Methods	59
3.3.3.1	Results on PETS2001 Database	59
3.3.3.2	Results on OTCBVS-Bench Database	61
3.4	Discussion	61
4	Human Action Recognition Using Boosted EigenAction	65
4.1	Introduction	65
4.2	Proposed Method	66
4.2.1	Salient Action Unit	66

4.2.1.1	Information Saliency Curve (ISC)	67
4.2.1.2	Constructing Salient Action Unit from ISC	67
4.2.2	Boosting EigenActions	68
4.2.2.1	EigenActions Representation	68
4.2.2.2	Adaboost Classifier	70
4.3	Experimental Results	76
4.3.1	Experimental Settings	76
4.3.2	Evaluation of the Proposed Method	78
4.3.2.1	Results on Weizmann Classification Database and KTH Database	78
4.3.3	Comparative Results	86
4.3.3.1	Results on Weizmann Classification Database and KTH Database	86
4.3.4	Robustness Test	88
4.3.4.1	Results on Weizmann Robustness Database	88
4.4	Discussion	93
5	Human Action Recognition From Boosted Co-Training	94

5.1	Introduction	94
5.2	Previous Works on Semi-Supervised Learning	95
5.3	Boosted Co-Training for Human Action Recognition	98
5.3.1	Overview of the Proposed Method	98
5.3.2	Inter-view Confidence	100
5.3.2.1	Mutual Information	100
5.3.2.2	Computing Inter-View Confidence	102
5.3.3	Intra-view Confidence	103
5.3.3.1	Multi-Class SemiBoost	104
5.3.3.2	Computing Intra-View Confidence	107
5.3.4	Single View Classifier	108
5.4	Experimental Results	110
5.4.1	Experimental Settings	110
5.4.2	Performance Evaluation	110
5.4.3	Comparative Results	113
5.4.4	Robustness Test	119
5.4.5	Visual Surveillance	122

5.5 Discussion	123
6 Conclusions	126
6.1 Summary	126
6.2 Future Work	127
Curriculum Vitae	153