DOCTORAL THESIS

Human motion detection and action recognition

Liu, Chang

Date of Award:
2010

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 extensive 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