A Framework for Dynamic Restructuring of Semantic Video Analysis Systems Based on Learning Attention Control

Mohamad-Hoseyn Sigari, Hamid Soltanian-Zadeh, Hamid-Reza Pourreza

PII: S0262-8856(15)00098-0
DOI: doi: 10.1016/j.imavis.2015.07.004
Reference: IMAVIS 3427

To appear in: *Image and Vision Computing*

Received date: 5 January 2015
Revised date: 8 July 2015
Accepted date: 10 July 2015

Please cite this article as: Mohamad-Hoseyn Sigari, Hamid Soltanian-Zadeh, Hamid-Reza Pourreza, A Framework for Dynamic Restructuring of Semantic Video Analysis Systems Based on Learning Attention Control, *Image and Vision Computing* (2015), doi: 10.1016/j.imavis.2015.07.004

A Framework for Dynamic Restructuring of Semantic Video Analysis Systems Based on Learning Attention Control

Mohamad-Hoseyn Sigari [a,*], Hamid Soltanian-Zadeh [a,b,c], Hamid-Reza Pourreza [d]

[a] *Control and Intelligent Processing Center of Excellence (CIPCE), School of Electrical and Computer Engineering, College of Engineering, University of Tehran, Tehran 14399, Iran*

[b] *Image Analysis Lab, Department of Radiology, Henry Ford Health System, Detroit, MI 48202, USA*

[c] *School of Cognitive Sciences, Institute for Research in Fundamental Sciences (IPM), Tehran 19395, Iran*

[d] *Machine Vision Research Lab, Computer Engineering Department, Faculty of Engineering, Ferdowsi University of Mashhad, Mashhad, 91779, Iran*

Abstract

Current semantic video analysis systems are usually hierarchical and consist of some levels to overcome semantic gaps between low-level features and high-level concepts. In these systems, some features, descriptors, objects or concepts are extracted in each level and therefore, total computational complexity of such systems is huge. In this paper, we present a new general framework to impose attention control on a video analysis system using Q-learning. Thus, our proposed framework restructures a given system dynamically to direct attention to the blocks extracting the most informative features/concepts and reduces computational complexity of the system. In other words, the proposed framework directs flow of processing actively using a learning attention control method. The proposed framework is evaluated for event detection in broadcast soccer videos using a limited numbers of training samples. Experiments show that the proposed framework is able to learn how to direct attention to informative features/concepts and restructure the initial structure of the system dynamically to reach the final goal with less computational complexity.

Keywords: *Attention Control; Broadcast Soccer Video; Event Detection; Q-learning; Semantic Video Analysis.*

1. Introduction

Nowadays, a wide variety of digital videos such as movies, news and sport videos are available on the web, cell-phones, hard disks, and non-volatile memories. On the other hand, almost all users need semi- or fully automated tools for video analysis and management in different applications such as video indexing and retrieval, video summarization and event detection. Therefore, scientists and companies are trying to develop efficient methods and tools for content and semantic video analysis. In this section, we review semantic video analysis systems developed in recent years. Methods used in these systems are divided into two main approaches: (1) methods based on content analysis; and (2) methods based on

* Corresponding author. Tel.: +98 9194099867.
E-mail addresses: hoseynsigari@ut.ac.ir (M. H. Sigari), hszadeh@ut.ac.ir (H. Soltanian-Zadeh), hpourreza@um.ac.ir (H. R. Pourreza).