Chapter 15

Human Face Region Detection Driving Activity Recognition in Video

Anastasios Doulamis
Technical University of Crete, Greece

Athanasios Voulodimos
National Technical University of Athens, Greece

Theodora Varvarigou
National Technical University of Athens, Greece

ABSTRACT

Automatic recognition of human actions from video signals is probably one of the most salient research topics of computer vision with a tremendous impact for many applications. In this chapter, the authors introduce a new descriptor, the Human Constrained Pixel Change History (HC-PCH), which is based on PCH but focuses on the human body movements over time. They propose a modification of the conventional PCH that entails the calculation of two probabilistic maps based on human face and body detection, respectively. These HC-PCH features are used as input to an HMM-based classification framework, which exploits redundant information from multiple streams by employing sophisticated fusion methods, resulting in enhanced activity recognition rates.

INTRODUCTION

Identification of events from visual cues is in general a very arduous task because of complex motion, cluttered backgrounds, occlusions, and geometric and photometric variances of the physical objects. Services regarding identification of events from visual signals are of vital importance for large-scale enterprises like industrial plants or public infrastructure organizations. For example, an event identification service is used for quality assurance, i.e. adherence to predefined procedures for production or services, or security and safety purposes, namely prevention of actions that may lead to hazardous situations (Doulamis et al., 2008).
Recently several supervision systems have been presented; however, for most of them the supervision service is manually performed which is insufficient and subjective. The inefficiency stems from the fact that the videos from many cameras are displayed on monitors that switch between cameras, thus no 100% monitoring is possible, even if we assume that the operators are constantly concentrated on their task. Regarding subjectivity, recent studies have proven that the attention of the operators of current surveillance systems is mainly attracted by the appearance of monitored individuals and not by their behaviour (Smith, 2004).

The recent research advances in computer vision and pattern recognition have stimulated the development of a series of innovative algorithms, tools and methods for salient object detection and tracking in still images/video streams. All these research methods can be considered as initial steps towards the ultimate goal for behaviour/event understanding. However, automatic comprehension of someone’s behaviour within a scene or even automatic supervision of workflows (e.g., industrial processes) is a complex research field of great attention but with limited results so far, since we need to map the extracted low level visual features to high level concepts, such as human actions performed within a scene. An example of an architecture, able to recognize events from visual signals, is presented in (Doulamis et al., 2008) and developed by the European Union funded project SCOVIS (“Self Configurable Cognitive Video Supervision- www.scovis.eu). This research was one of the first results for large scale automatic video supervision of complex industrial processes.

Apart from the research work of SCOVIS, other approaches have been also proposed in the literature for automatic event identification from video information, as described in the Related Work Section. The common point of all these works is the extraction of a set of visual descriptors that capture spatial and temporal variations in an image sequence [such as Motion History Image (Davis, 2001) or Pixel Change History (Xiang & Gong, 2006)], which are then fed to classifier to detect the events and human actions, such as Hidden Markov Models or Neural Networks. However, these descriptors are very generic and thus classification accuracy can be robust and reliable for event detection only in cases of well structured actions executed under noise-free environments, or in cases where the visual recordings are restricted to specific visual domains (like sports and news) [see the Background Section]. To improve reliability of human actions recognition process under complex industrial workflows but structured environments, (Doulamis et al., 2008) modifies the traditional Pixel Change History (PCH) descriptor of (Xiang and Gong, 2006), which is in fact a visual map, to incorporate Zernike moments on PCH descriptors (Zernike, 1934). Still, however, the results suffer from accuracy especially when abrupt background changes take place stemming from manufacturing assembly processes. This is mainly due to the fact that PCH considers whole image alterations meaning that abrupt luminosity changes or severe motion in the background can affect the event identification process.

In order to overcome this problem, (Voulodimos et al., 2012) and (Kosmopoulos et al., 2012) propose a rectification mechanism that ameliorates the performance of the classifiers by taking into account information coming from expert users. The latter evaluate the supervision outcomes of the classifier and the evaluation results are fed back to the system to improve its performance at next iterations of the algorithm. Such an approach, however, moves the complexity from the system to the user side involving the user in the loop and therefore making the supervision process semi-automatic (semi-supervised).

In this Chapter, we propose a new methodology that is suitable for detecting human actions and events from video recordings. Our algorithm introduces a Human Constrained Pixel Change History descriptor, hereby called HC-PCH, in order to eliminate background effect on human