Chapter 53
Classifying Behaviours in Videos with Recurrent Neural Networks

Javier Abellan-Abenza
University of Alicante, Spain

Sergiu Oprea
University of Alicante, Spain

Alberto Garcia-Garcia
University of Alicante, Spain

David Ivorra-Piqueres
University of Alicante, Spain

Jose Garcia-Rodriguez
University of Alicante, Spain

ABSTRACT

This article describes how the human activity recognition in videos is a very attractive topic among researchers due to vast possible applications. This article considers the analysis of behaviors and activities in videos obtained with low-cost RGB cameras. To do this, a system is developed where a video is input, and produces as output the possible activities happening in the video. This information could be used in many applications such as video surveillance, disabled person assistance, as a home assistant, employee monitoring, etc. The developed system makes use of the successful techniques of Deep Learning. In particular, convolutional neural networks are used to detect features in the video images, meanwhile Recurrent Neural Networks are used to analyze these features and predict the possible activity in the video.

INTRODUCTION

Human Behavior Analysis (HBA) is a big field of interest in the artificial intelligence and computer vision community. It has many application areas like Video Surveillance, Ambient- Assisted Living, Smart Shopping Environments, etc. Supported by relevant companies in this field, the availability of human video data is growing significantly.

This work approaches HBA from the DL perspective. Deep Learning techniques have been a great step in the context of classification in the last few years due to the growth of computational power. Some of these techniques are Convolutional Neural Networks (CNNs) for image understanding, and RNNs

DOI: 10.4018/978-1-7998-0414-7.ch053

for temporary understanding such as video or text. The main purpose of this project is to design and implement an efficient deep learning solution able to predict daily human activities recorded from RGB cameras, by using both CNNs and RNNs architectures. Moreover, an implementation of this project over Graphics Processing Units (GPUs) is aimed for comparison purposes.

The main goal of this work is the development of a human behavior recognition system to assist dependent people. The secondary goal is to take advantage of GPUs and accelerate the system. For this purpose, an extensive state of the art of human behavior recognition systems has been carried out. At the same time, an analysis of the available datasets has been performed in order to choose the most suitable one. For the implementation, existing deep learning frameworks have been used. To accomplish with the secondary goal, this training has been performed using GPUs to exploit parallelism. Finally, a performance analysis of our system has been performed.

After exposing the motivation and goals of this work, this paper is structured as follows: First, we present a detailed state of the art regarding HBA systems and DL techniques. Next, we present the methodology exposing the different techniques, technologies, and datasets used to carry out this work. Then, we describe the proposed solution in detail attaching the main implementation parts of our system. The experiments and discussion for each part of the implementation are detailed throughout this section. Finally, some conclusions were extracted alongside potential research directions and future works.

STATE OF THE ART

Human Behaviour Analysis (HBA) involves a wide range of applications: Video Surveillance, Ambient-Assisted Living, etc. All these applications have in common the need of creating an artificial intelligence that understands the body of a person and its natural movement for different activities. Human activities, such as "walking" or "running," are relatively easy to recognize. On the other hand, more complex activities, such as "peeling an apple," are more difficult to identify. Complex activities may be decomposed into other simpler activities, which are generally easier to recognize. Therefore, it is necessary to understand the different HBA levels that exist. Moeslund, Hilton, and Krüger (2006) defined a classification of the different action taxonomies that have been adopted later in many other works. It defines three levels of abstractions from smallest to biggest: 1. Action primitive: Basic motion recognition that represents the atomic movement out of which actions are built. 2. Action: Composed of different action primitives. 3. Activity: A higher level of abstraction which requires the semantic notion of the context and the involved objects.

Although this taxonomy is highly used among researchers, some of them use their own taxonomies, for example Ji, Liu, Li, and Brown (2008) include a higher level of abstraction called behaviour. They defined behaviour as" human motion patterns involving high-level description of actions and interactions".

Motion recognition is the fundament for detecting human activities or behaviours. Motion is decomposed in a series of poses through time. A pose can be described as the state of the body posture that can be represented by an articulated system of rigid segments connected by joints, like the model described in Andriluka, Roth, and Schiele (2009); Sapp, Toshev, and Taskar (2010).

The work of Gavrila (1999), reveals important applications of "looking at people" and reviews several lower level techniques of detecting human motion. Later, the work made by Moeslund and Granum (2001) was focused in recognizing human movements. He described a functional taxonomy of the phases required