Chapter 14
Directional Multi-Scale Stationary Wavelet-Based Representation for Human Action Classification

M. N. Al-Berry
Ain Shams University, Egypt

H. M. Ebeid
Ain Shams University, Egypt

Mohammed A.-M. Salem
Ain Shams University, Egypt

A. S. Hussein
Arab Open University, Kuwait

Mohamed F. Tolba
Ain Shams University, Egypt

ABSTRACT

Human action recognition is a very active field in computer vision. Many important applications depend on accurate human action recognition, which is based on accurate representation of the actions. These applications include surveillance, athletic performance analysis, driver assistance, robotics, and human-centered computing. This chapter presents a thorough review of the field, concentrating the recent action representation methods that use spatio-temporal information. In addition, the authors propose a stationary wavelet-based representation of natural human actions in realistic videos. The proposed representation utilizes the 3D Stationary Wavelet Transform to encode the directional multi-scale spatio-temporal characteristics of the motion available in a frame sequence. It was tested using the Weizmann, and KTH datasets, and produced good preliminary results while having reasonable computational complexity when compared to existing state–of–the–art methods.

INTRODUCTION

Recently, intelligent cognitive systems began to appear with a vision that ambient intelligence in the near future will be a part of our daily life (Pantic, Nijholt, Pentland, & Huanag, 2008). This opened the challenge that computers should be able to understand actions performed by humans and respond according to this understanding.

DOI: 10.4018/978-1-5225-2229-4.ch014

Many applications depend on human action and activity recognition. These applications can be classified into surveillance, control, and analysis applications (Moeslund, Hilton, & Kruger, 2006). Intelligent surveillance is the monitoring process that analyses the scene, interprets object behaviors, and involves as well event detection, object detection, recognition, and tracking. This includes security systems that detect abnormal behavior (Huang & Tan, 2010; Roshtkhari & Levine, 2013) in security sensitive areas like airports (Aggarwal & Cai, 1999), surveillance of crowd behavior (Chen & Huang, 2011; Sharif, Uyaver, & Djeraba, 2010), group activity recognition (Cheng, Qin, Huang, Yan, & Tian, 2014), and person identification using behavioral biometrics (Turaga, Chellappa, Subrahamanian, & Udrea, 2008; Sarkar, Phillips, Liu, Vega, Grother, & Bowyer, 2005).

Control applications are the category of applications that depend on interaction between human and computer (Pantic, Nijholt, Pentland, & Huanag, 2008; Poppe, 2010; Pantic M., Pentland, Nijholt, & Huanag, 2007; Rautaray & Agrawal, 2012). These applications recognize the human gestures to control something such as smart houses (Brdiczka, Langet, Maisonnasse, & Crowley, 2009; Fatima, Fahim, Lee, & Lee, 2013), and intelligent vehicles (Wu & Trivedi, 2006). Analysis applications include content-based image and video retrieval (Laptev, Marszalek, Schmid, & Rozenfeld, 2008), driver sleeping detection, robotics (Freedman, Jung, Grupen, & Zilberstein, 2014), and athletic performance analysis.

The field of action and activity recognition is still an open research area because there are various types of challenges that face it. For action recognition, challenges arise from variations in the rate execution of actions (Cristani, Raghavendra, Del Bue, & Murino, 2013) (Thi, Cheng, Zhang, Wang, & Satoh, 2012) (Ashraf, Sun, & Foroosh, 2014). As the number of individuals and interactions increase, the complexity of the task increases. Therefore, higher behavior understanding faces some more difficult challenges including the number of modalities to be used, how to fuse them, and how to make use of the context in the process of learning and recognition (Vishwakarma & Agrawal, 2013).

Poppe (Poppe, 2010), defined vision-based human action recognition as: "The process of labeling image sequences with action labels". Following Weinland et al. (Weinland, Ranford, & Boyer, 2011), an action is a sequence of movements generated by a performer during the performance of a task, and an action label is a name, such that an average human agent can understand and perform the named action.

Different methods have been proposed for segmenting, representing, and classifying actions. These methods can be classified into different taxonomies (Weinland, Ranford, & Boyer, 2011), (Pantic, Pentland, Nijholt, & Huanag, 2006), (Turaga, Chellappa, Subrahamanian, & Udrea, 2008). One of the famous methods that have been used for holistic motion representation is the Motion History Image (MHI) (Davis, 2001) (Babu & Ramakrishnan, 2004) (Ahad, Tan, Kim, & Ishikawa, 2012). Motion History Images are temporal templates that are simple, but robust in motion representation, and they are used for action recognition by several research groups (Ahad, Tan, Kim, & Ishikawa, 2012).

This chapter aims at providing a review on the recent advances in the field, with a focus on spatio-temporal action representation. In addition, the chapter proposes a multi-scale spatio-temporal action representation based on 3D Stationary Wavelet Analysis. In this paper, a stationary wavelet-based directional action representation is proposed. The proposed representation is based on the 3D Stationary Wavelet Transform (SWT) that has been proposed and used in (Al-Berry, Salem, Hussein, & Tolba, 2014) for spatio-temporal motion detection. The 3D SWT succeeded in motion detection in the presence of illumination variations in both indoor and outdoor scenarios while having reasonable complexity. In the proposed action representation, the 3D SWT is used to encode the action into 3 directional wavelet-based templates. Hu invariant moments (Hu, 1962) have been used for describing the templates, and concatenated into a combined feature vector. The preliminary results obtained using these simple