Spatio-Temporal Action Localization and Detection for Human Action Recognition in Big Dataset

Sameh Megrhi, Marwa Jmal, Wided Souidene, Azeddine Beghdadi

PII: S1047-3203(16)30221-8
DOI: http://dx.doi.org/10.1016/j.jvcir.2016.10.016
Reference: YJVCI 1884

To appear in: *J. Vis. Commun. Image R.*

Received Date: 8 December 2015
Accepted Date: 28 October 2016

Spatio-Temporal Action Localization and Detection for Human Action Recognition in Big Dataset

Sameh MEGRHI[a], Marwa JMAL[b,c,1], Wided SOUIDENE[a,b], Azeddine BEGHDADI[a]

[a] L2TI, Institut Galilée, Université Paris 13, 99, Avenue Jean-Baptiste Clement, 93430 Villetaneuse, France
[b] SERCom Laboratory, Ecole Polytechnique de Tunisie, Université de Carthage, B.P.743, 2078. La Marsa, Tunisie
[c] Telnet Innovation Labs, Telnet Holding, Ariana, Tunis

Abstract

Human action recognition is still attracting the computer vision research community due to its various applications. However, despite the variety of methods proposed to solve this problem, some issues still need to be addressed. In this paper, we present a human action detection and recognition process on large datasets based on Interest Points trajectories. In order to detect moving humans in moving field of views, a spatio-temporal action detection is performed basing on optical flow and dense speed-up-robust-features (SURF). Then, a video description based on a fusion process that combines motion, trajectory and visual descriptors is proposed. Features within each bounding box are extracted by exploiting the bag-of-words approach. Finally, a support-vector-machine is employed to classify the detected actions. Experimental results on the complex benchmark UCF101, KTH and HMDB51 datasets reveal that the proposed technique achieves better performances compared to some of the existing state-of-the-art action recognition approaches.

Keywords: Spatio-temporal action detection, Dense SURF, Optical flow, Action recognition, Selective temporal segmentation, Interest points trajectory.

1. Introduction

Action recognition is an active research field in computer vision. It represents a wide range of applications such as video surveillance, gesture interpretation, robotic vision, video search/retrieval and human-machine interaction. Recognizing human actions in videos is a challenging task due to the large intra-class variations of complex actions, poor quality and camera motion. In order to overcome these issues, a relevant video description based on dividing it into small sequences is required. In the following, we introduce our approach and provide a brief critical literature survey of the different methods developed for each component of the common video description and analysis system dedicated to human action detection and recognition.

Temporal segmentation of videos may be performed in different ways. Some methods are based on the trajectory of interest points (IP)[1, 2]. Various trajectory based descriptors has been proposed in the last decades, [3, 4]. These descriptors are extracted either from optical flow [5, 6], or by matching IP in different frames [4, 5]. For such, the number of frames involved in setting the trajectory length depends on the used approach. In [4], the trajectory length is within a fixed interval while in [7] it is based on a fixed frame number in order to extract a displacement vector.

In some scenarios, action recognition is pre-processed by a motion segmentation step [8]. Thus, its performance is highly related to the segmentation algorithm. Pixel-wise techniques, namely background subtraction and temporal differencing [9], are the most straightforward methods. However, they are only effective under the consideration of static cameras. When dealing with moving cameras, these models are likely to fail as the background is continuously varying in addition to the target's motion. A recent study [10] revealed that optical flow (OF) based methods [11] are one of the most effective techniques in motion segmentation. Horn and Schunck [12] and Lucas and Kanade (L&K) [13]

Email addresses: sameh.megrhi@univ-paris13.fr (Sameh MEGRHI), marwa.jmal@sercom-lab.com (Marwa JMAL), wided.mseddi@univ-paris13.fr (Wided SOUIDENE), azeddine.beghdadi@univ-paris13.fr (Azeddine BEGHDADI)

[1] Corresponding author