ISSN No. 0976-5697

Volume 5, No. 1, Jan-Feb 2014

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

Human Action Recognition Based On Multiview

P. Kalaivani
Ph.D. Scholar, Department of Computer Science
Mother Teresa Women's University
Kodaikanal, India
vanivijay2012@gmail.com

Dr. S. Vimala
H.O.D. In-charge, Department of Computer Science
Mother Teresa Women's University
Kodaikanal, India
vimalaharini@gmail.com

Abstract: This paper presents the different approaches and datasets for recognition of human actions under view changes. Visual analysis of human action is currently one of the most active research topics. This strong interest is driven by a wide spectrum of promising applications in many areas such as virtual reality, smart surveillance, perceptual interface, etc. Human action analysis concerns the detection, tracking and recognition of people, and more generally, the understanding of human behaviors, from image sequences involving humans. We consider the task of labeling videos containing human motion with action classes. The interest in the topic is motivated by the promise of many applications, both offline and online. In this paper, we specifically addressed multi-view front and top independent video analysis, with human action recognition for training and detection of different actions.

Keywords: human action recognition; datasets; detection; tracking; human behaviors;

I. INTRODUCTION

Visual recognition and understanding of human actions have attracted much attention over the past three decades and remain an active research area of computer vision. A good solution to the problem holds a yet unexplored potential for many applications, such as the search for and the structuring of large video archives, video surveillance, human-computer interaction, gesture recognition, and video editing. Recent work has demonstrated the difficulty of the problem associated with the large variation of human action data due to the individual variations of people in expression, posture, motion, and clothing, perspective effects and camera motions, illumination variations, occlusion, and distracting effects of scenes surroundings. Also, actions frequently involve and depend on manipulated objects, which add another layer of variability. Most of the current methods for action recognition are designed for limited view variations.

A reliable and a generic action recognition system, however, have to be robust to camera parameters and different viewpoints while observing an action sequence. The multi-view action recognition from a different perspective and avoids many assumptions of previous methods. Differently from the previous view-based methods, this does not assume multi view action samples either for training or for testing. In this paper, we first discuss related works and present the scope of this overview. Also, we outline the main characteristics and challenges of the field as these motivate the various approaches that are reported in literature. Finally, we briefly describe the most common datasets used for human action recognition.

II. RELATED WORK

A silhouette is the image of a person, an object or scene consisting of the outline and a featureless interior, with the silhouetted object usually being black. From its original graphic meaning, the term "silhouette" has been extended to describe the sight or representation of a person, object or scene that is backlit, and appears dark against a lighter background. Anything that appears this way, for example, a figure standing backlit in a doorway, may be described as "in silhouette". Silhouette used in the fields of fashion and fitness to describe the shape of a person's body or the shape created by wearing clothing of a particular style or period.

Parameswaran and Chellappa[1]propose a quasi-view-invariant approach, requiring at least five body points lying on a 3D plane or that the limbs trace a planar area during the course of an action. However, obtaining automatic and reliable point correspondences for daily video with natural human action is a very challenging and currently unsolved problem, which limits the application of the above mentioned methods in practice.

One alternative to the geometric approach is to represent the actions by samples recorded for the different views. A database of poses seen from multiple viewpoints has been created in Ahmadand and Lee[2]. Extracted silhouettes from a test action are matched to this database to recognize the action being performed. The drawback of these methods is that each action needs to be represented by many training samples recorded for a large and representative set of views. Other methods perform a full 3D reconstruction from silhouettes seen from multiple deployed cameras. This approach requires a setup of multiple views, which again restricts the applicability of methods in practice.

One approach has a close relation to the notion of video self-similarity used by Benabdelkader, Cutler and Davis [3]. In the domain of periodic motion detection, Cutler and Davis [4] track moving objects and extract silhouettes (or their bounding boxes). This is followed by building a 2D matrix for the given video sequence, where each entry of the matrix contains the absolute correlation between the two frames i and j. Their observation is that for a periodic motion, this similarity matrix will also be periodic. To detect and characterize the periodic motion, they resort to Time-Frequency analysis. None of the methods above explores the notion of self-similarity for multi-view unauthorized action recognition. We limit our focus to vision-based human action recognition to address the characteristics that are typical for the domain. We discuss image representation and action classification separately as these are the two parts that are present in every action recognition approach. Due to the large variation in datasets and evaluation practice, we discuss action recognition approaches conceptually, without presenting detailed results. We focus on recent work, which has not been discussed in previous works.

III. LABELING THE TRAINING DATA

A. *Training Data:*

Many works described in this paper use publicly available datasets that are specifically recorded for training and evaluation. This provides a sound mechanism for comparison but the sets often lack some of the earlier mentioned variations. Recently, more realistic datasets have been introduced. These contain labeled sequences gathered from movies or web videos. While these sets address common variations, they are still limited in the number of training and test sequences.

Adrien Gaidon, Marcin Marszałek, Cordelia Schmid,[5] present an approach to re-rank automatically extracted and aligned movie samples but manual verification is usually necessary. Also, performance of an action might be perceived differently. A small-scale experiment showed significant disagreement between human labeling and the assumed ground-truth on a common dataset[6]. When no labels are available, an unsupervised approach needs to be pursued but there is no guarantee that the discovered classes are semantically meaningful.

B. *Action Detection:*

Imran N Juneo, Emilie Dexter,Ivan Laptep and Patrick Perez [7] observe that the temporal self-similarity matrix of an action seen from different viewpoints is very similar (see Fig.1). They describe a sequence as a histogram of local descriptors, calculated from the self-similarity matrix. Boiman and Irani[8] take a different approach by describing a sequence as an ensemble of local spatial or spatio-temporal patches.

A similarity score is based on the composition of a query sequence from these patches. Similar sequences require less but larger patches.

IV. COMMON DATASETS FOR HUMAN ACTION RECOGNITION

The use of publicly available datasets allows for the comparison of different approaches and gives insight into the respective methods. Here we discuss about some of the most widely used datasets.

A. *Kth Human Motion Dataset:*

The KTH human motion dataset (Fig. 2a [9]) contains six actions (walking, jogging, running, boxing, hand waving and hand clapping), performed by 25 different actors. Four different scenarios are used: outdoors, outdoors with zooming, outdoors with different clothing and indoors. There is considerable variation in the performance and duration, and somewhat in the viewpoint. The backgrounds are relatively static. Apart from the zooming scenario, there is only slight camera movement.

B. *Weizmann human action dataset:*

The human action dataset (Fig. 2b[10]) recorded at the Weizmann institute contains 10 actions (walk, run, jump, gallop sideways, bend, one-hand wave, two-hands wave, jump in place, jumping jack and skip), each performed by 10 persons. The backgrounds are static and foreground silhouettes are included in the dataset. The view-point is static. In addition to this dataset, two separate sets of sequences were recorded for robustness evaluation. One set shows walking movement viewed from different angles. The second set shows front to parallel walking actions with slight variations (carrying objects, different clothing, and different styles).

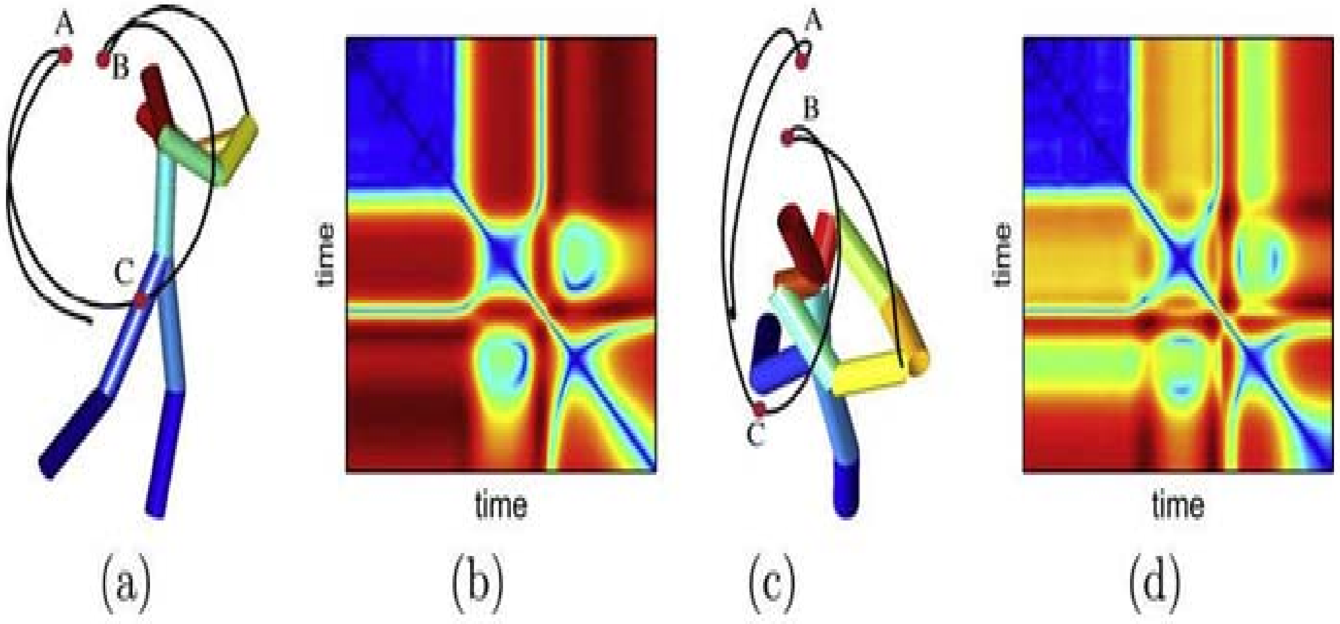

Figure1. Example of cross-correlation between viewpoints, (a and c) a golf swing seen from two different viewpoints, (b and d) the corresponding self-similarity matrices

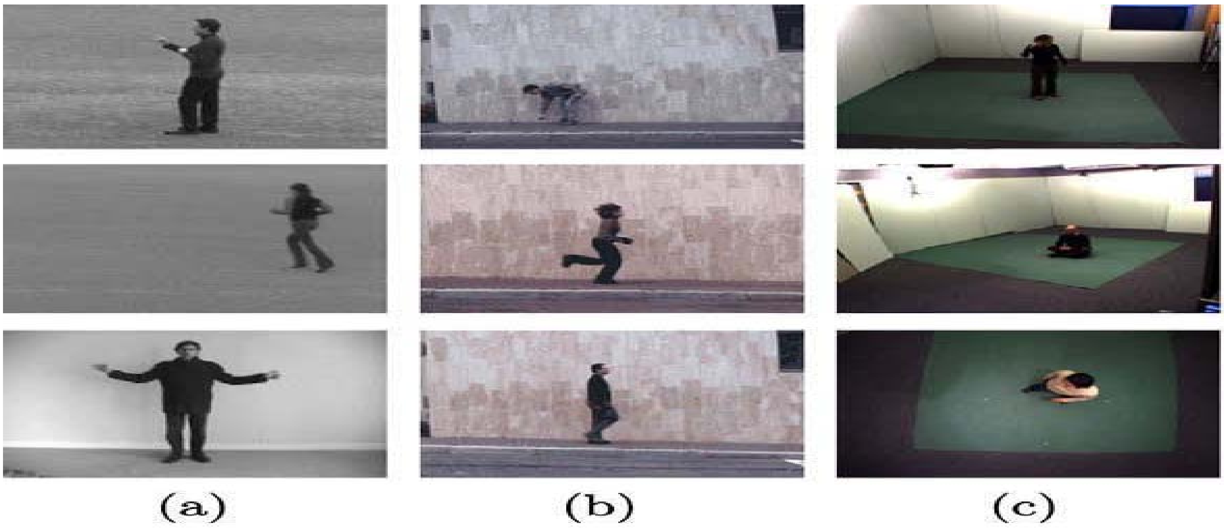

Figure 2. Example frames of (a) KTH dataset, (b) Weizmann dataset, (c) Inria XMAS dataset

C. INRIA XMAS multi-view dataset:

The IXMAS dataset (Fig. 2c[11]) that contains actions captured from five viewpoints. A total of 11 persons perform 14 actions (check watch, cross arms, scratch head, sit down, get up, turn around, walk, wave, punch, kick, point, pick up, throw over head and throw from bottom up). The actions are performed in an arbitrary direction with regard to the camera setup. The camera views are fixed, with a static background and illumination settings. Silhouettes and volumetric voxel representations are part of the dataset.

V. RESULTS AND DISCUSSION

In this section, we discuss the features that are extracted from the image sequences. Ideally, these should generalize over small variations in person appearance, background, viewpoint and action execution. At the same time, the representations must be sufficiently rich to allow for robust classification of the action.

The temporal aspect is important in action performance. Some of the image representations explicitly take into account the temporal dimension; others extract image features for each frame in the sequence individually. In this case, the temporal variations need to be dealt with in the classification step.

A. Image representation:

Identify the foreground and background pixel of a frame. Background model stores the values of a particular pixel which corresponds to the background colors. Pixel Change History (PCH) is represented for a pixel. Similar foreground pixels are grouped to form a blob. A behavior pattern is represented as a sequence of various events. For example, behavior patterns A and B contains events of classes a, b and a, c and e. Behavior patterns A and B are deemed as different since the events and their orders differs. Build training data set and group training behavior patterns upon which a model for normal behavior can be built.

Many approaches assume that the video is readily segmented into sequences that contain one instance of a known set of action labels. Often, it is also assumed that the location and approximate scale of the person in the video is known or can easily be estimated. The action detection task is thus ignored, which limits the applicability to situations where segmentation in space and time is possible. While several works (e.g. [12, 13]) have addressed this topic, it remains a challenge to perform action detection for online applications.

VI. CONCLUSION

In this paper, we specifically addressed for multi-view front and top independent video analysis, with human action recognition as a central application. This would be a big step towards the fulfillment of the longstanding promise to achieve robust automatic recognition and interpretation of human action. Another aspect of human action recognition is the current evaluation practice. Publicly available datasets (see Section IV) have shaped the domain by allowing for objective comparison between approaches on common training and test data. Experimental validation on action recognition, as well as for the different problem of action synchronization, clearly confirms the stability of this type of description with respect to view variations. Results on public multi-view action recognition data sets demonstrate superior performance of this method compared to alternative methods in the literature.

VII. ACKNOWLEDGEMENT

The Authors expresses their sincere thanks to Meenakshi College of Engineering, Chennai for providing necessary facility to conduct the research work.

VIII.REFERENCES

[1]. Parameswaran.V and Chellappa.R, “View Invariance for Human Action Recognition,” IJCV, vol. 66, no. 1, pp. 83-101,2006.

[2]. M. Ahmad and S. Lee, “HMM-based human action recognition using multiview image sequences,” in Proc. ICPR, 2006, pp. I:263–266..

[3]. C. Benabdelkader, R. Cutler, and L.Davis, “Gait Recognition Using Image Self-Similarity,” EURASIP J. Applied Signal Processing, vol. 2004, no. 1, pp.572-585, Jan, 2004.

[4]. R. Cutler and L. Davis, "Robust real-time periodic motion detection, analysis, and applications," PAMI, vol. 22, no. 8, pp. 781–796, 2000..

[5]. Adrien Gaidon, Marcin Marszałek, Cordelia Schmid, "Mining visual actions from movies," Proc. of the British Machine Vision Conference (BMVC'09), London, United Kingdom, in press.

[6]. Alonso Patron-Perez, Ian Reid, "A probabilistic framework for recognizing similar actions using spatio-temporal features," Proc. of the British Machine Vision Conference (BMVC'07), Edinburgh, United Kingdom, September 2007, pp. 1–10.

[7]. Imran N Juneo, Emilie Dexter,Ivan Laptep and Patrick Perez, "View Independent Action Recognition From Temporal Self Similarities," PAMI, vol 33, no1, pp 172-185,2011.

[8]. Oren Boiman, Michal Irani, "Detecting irregularities in images and in video," IJCV, 74 (1) (2007) 17–31.

[9]. Moshe Blank, Lena Gorelick, Eli Shechtman, Michal Irani, Ronen Basri, "Actions as space–time shapes", Proc. of the International Conference on Computer Vision (ICCV'05), vol. 2, Beijing, China, October 2005, pp. 1395–1402

[10]. Christian Schuldt, Ivan Laptev, Barbara Caputo, "Recognizing human actions: a local SVM approach," Proc. of the International Conference on Pattern Recognition (ICPR'04), 2004, vol. 3, Cambridge, United Kingdom, 2004, pp. 32–36.

[11]. Daniel Weinland, Remi Ronfard, Edmond Boyer, "Free viewpoint action recognition using motion history volumes," Computer Vision and Image Understanding (CVIU) 104 (2–3) (2006) 249–257.

[12]. Matteo Bregonzio, Shaogang Gong, Tao Xiang, Recognising action as clouds of space–time interest points, in: Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR'09), Miami,FL, June 2009, pp.1–8.

[13]. Fabrice Caillette, Aphrodite Galata, Toby Howard, "Real-time 3-D human body tracking using learnt models of behaviour," Computer Vision and Image Understanding (CVIU) 109 (2) (2008) 112–125.