Pattern Recognition 47 (2014) 3343–3361

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

A survey on still image based human action recognition

Guodong Guo*, Alice Lai

Lane Department of Computer Science and Electrical Engineering, West Virginia University, Morgantown, WV 26506, United States

ARTICLE INFO

Article history:
Received 23 July 2013
Received in revised form
16 March 2014
Accepted 13 April 2014
Available online 9 May 2014

Keywords:
Action recognition
Still image based
Various cues
Databases
Survey
Evaluation

ABSTRACT

Recently still image-based human action recognition has become an active research topic in computer vision and pattern recognition. It focuses on identifying a person's action or behavior from a single image. Unlike the traditional action recognition approaches where videos or image sequences are used, a still image contains no temporal information for action characterization. Thus the prevailing spatio-temporal features for video-based action analysis are not appropriate for still image-based action recognition. It is more challenging to perform still image-based action recognition than the video-based one, given the limited source of information as well as the cluttered background for images collected from the Internet. On the other hand, a large number of still images exist over the Internet. Therefore it is demanding to develop robust and efficient methods for still image-based action recognition to understand the web images better for image retrieval or search. Based on the emerging research in recent years, it is time to review the existing approaches to still image-based action recognition and inspire more efforts to advance the field of research. We present a detailed overview of the state-of-the-art methods for still image-based action recognition, and categorize and describe various high-level cues and low-level features for action analysis in still images. All related databases are introduced with details. Finally, we give our views and thoughts for future research.

1. Introduction

Recognizing human motion and action has been an active research topic in Computer Vision for more than two decades. This can also be indicated by a series of survey papers in the literature. Earlier review papers focused on human motion analysis and discussed human action recognition as a part, such as the surveys by Cedras and Shah [1], Aggarwal and Cai [2], and Gavrila [3]. Later on, the survey paper by Kruger et al. [4] classified human action recognition approaches based on the complexity of features to represent human actions and considered potential applications to robotics. The survey paper by Turaga et al. [5] covered human activity recognition with a categorization based on the complexity of activities and recognition methodologies. In Poppe's survey [6], various challenges in action recognition were addressed and novelties of different approaches were discussed. In Ji and Liu's survey [7], the concentration was on view-invariant representation for action recognition. They discussed related issues such as human detection, view-invariant pose representation and estimation, and behavior understanding. Finally, the most recent survey was given by Aggarwal and Ryoo in [8], who performed a comprehensive review of recognizing action, activity, gesture, human–object interaction, and group activities. It discussed the limitations of many existing approaches and listed various databases for evaluations. The real-time applications were also mentioned.

Although motion-based/video-based human action recognition is still an active research topic in computer vision and pattern recognition, recent studies have started to explore action recognition in *still images*. As shown in Fig. 1, many action categories can be depicted unambiguously in single images (without motion or video signal), and these actions can be understood well based on human perception. This evidence supports the development of computational algorithms for automated action analysis and recognition in still images. Considering the large number of single images distributed over the Internet, it is valuable to analyze human behaviors in those images. Actually, it has become an active research topic very recently [9].

An analogy to human (body-based) action recognition is facial expression recognition [10,11], sometimes called the facial behavior understanding. In facial expression analysis, either single face images or face videos can be used. Different from action recognition, the studies of facial expressions using single images or videos are almost in parallel. The number of publications using either single images or videos is probably comparable in facial expression recognition. However, in action recognition, a large number of publications are video-based. Only very recently, researchers have begun to focus on still image-based action understanding.

* Corresponding author.
E-mail address: guodong.guo@mail.wvu.edu (G. Guo).

http://dx.doi.org/10.1016/j.patcog.2014.04.018
0031-3203/

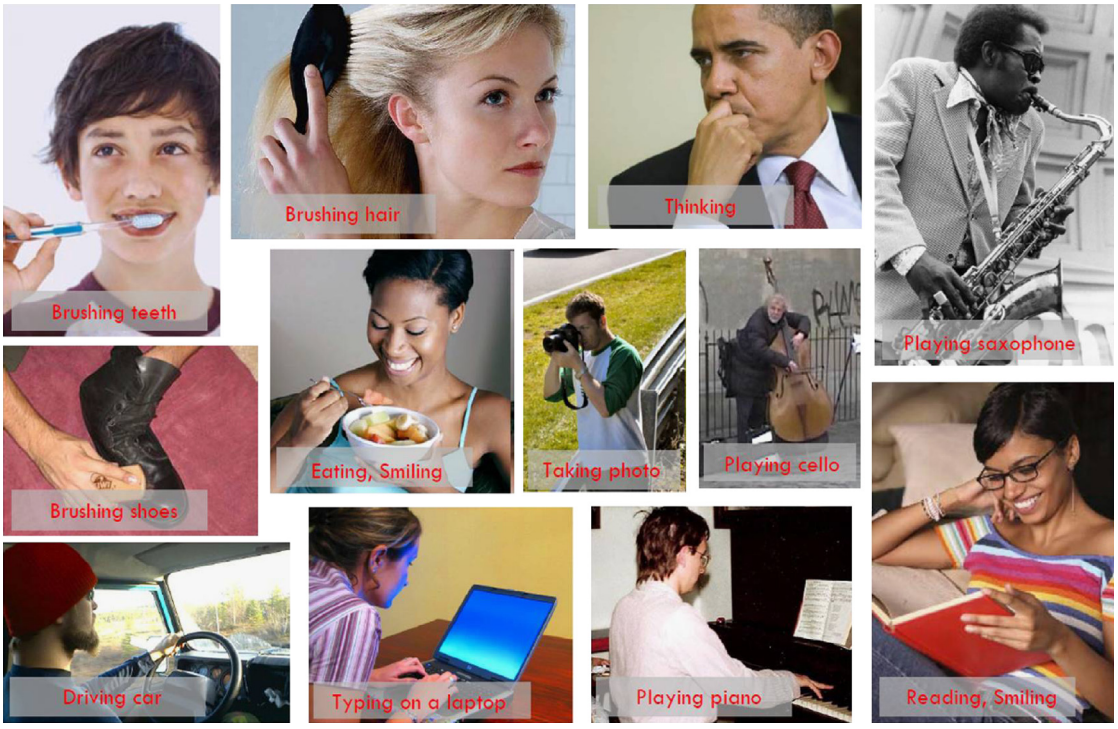

Fig. 1. Some examples of still image-based action recognition. Only single images are sufficient to tell the corresponding actions, originally shown in [9].

Compared to the traditional video-based action recognition, still image-based action recognition has some special properties. For example, there is no motion in a still image, and thus many spatiotemporal features and methods that were developed for traditional video-based action recognition are not applicable to still images. And also, it is not trivial to segment the humans from the background in still images [12–14], since there is no motion cue to utilize and the scene can be very cluttered. Thus there are new challenges in solving the problem of still image-based action recognition.

Still image-based action recognition has quite a few useful applications: (1) image annotation. A huge amount of still images are distributed over the Internet, and new images are being acquired more and more. Automated action recognition in still images can help to annotate "verbs" (for actions) on Internet images (such as the examples shown in Fig. 1). (2) Action or behavior based image retrieval. Similar to off-line image annotation, the automated action recognition can also help search and retrieve online images based on action queries. (3) Video frame reduction in video-based action recognition. When still image based action recognition can achieve a high accuracy for some categories of actions, the long video sequences for those actions can be reduced to a small number of single frames for action representation, and thus significantly lower the redundant information without satisfying the action recognition accuracy. (4) Human computer interaction (HCI). Similar to the traditional video-based action recognition, still image based action recognition can also be useful for HCI, especially for actions that do not require a long time period to execute the whole process, e.g., touching, thinking, and smiling.

Although there are useful applications in practice, the study of still image-based action recognition has a very short history, compared to the video-based action recognition research. Starting at about the year of 2006, it appears to have some research papers on still image-based action recognition. Following 2006, only a very small number of papers related to action recognition based on single images appeared, since not many researchers have realized that it is an interesting topic. More papers were published recently. To show the trend of publications on this topic, the annual and the accumulative number of published papers are drawn in a yearly basis and shown in Fig. 2. The criterion of selecting papers on still image based action recognition is based on whether still images are used as the test data for action recognition. The training examples can be purely still images or single image frames extracted from video sequences. The time period to collect the related publications for this survey starts from the year that the earliest paper was published in 2006 until 2013. From Fig. 2, we can see that there are more publications in very recent years, such as 2011 and 2012, with about 10 papers each year, while much less numbers before 2011. The research on this topic has become more active since 2011. We also noticed that there were less numbers of publications after 2011. One reason is that there might be a "bottleneck" to improve the recognition accuracies significantly. New ideas and approaches are needed to address this challenging problem.

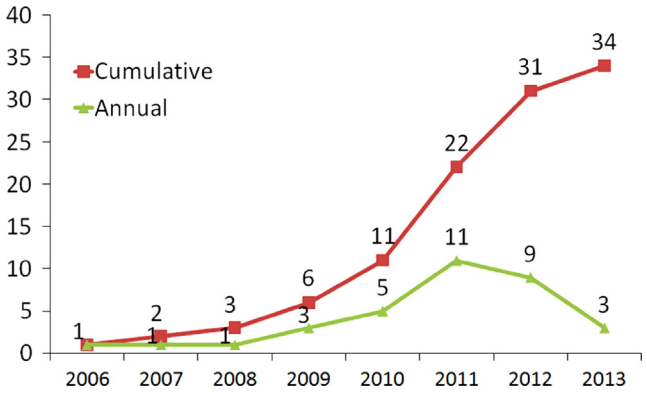

Fig. 2. Graphic display of the number of publications annually and accumulatively on still image-based action recognition. Based on the slope of the piecewise curve, there are more publications in 2011 and 2012.