Eurographics/ ACM SIGGRAPH Symposium on Computer Animation (2012)
M. Lau (Posters Editor)

Data-Driven Full Body Action Recognition

Jan Baumann[1], Susanne Eichert[2], Raoul Wessel[2], Björn Krüger[2] and Andreas Weber[2]

[1] Fraunhofer FKIE, Unmanned Systems Group, Wachtberg
[2] Bonn University, Institute of Computer Science II, Computer Graphics

Abstract

This poster presents a novel data-driven method for recognizing human full body actions. The approach detects actions in an online manner, meaning the action recognition takes place as the motion traverses from beginning to end, signalling if an action has ended at the currently processed frame.
The method is evaluated using various motion capture databases and query motions ranging from optical motion capture data to skeleton data obtained from a depth camera.

Categories and Subject Descriptors (according to ACM CCS): I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Animation

Keywords: depth camera; gesture detection; action recognition; computer animation

1. Introduction

Consumer motion capture systems (like Kinect, WiiMote or EyeToys) have received a lot of attention in recent years, primarily because they enable the user to interact with an application in a very natural way using low cost consumer hardware. The field of usage exceeds replacing the classic game controller in computer games, e.g. detecting referee signals at competitions, as sketched in this work.

We come up with a fully data driven action recognition scheme, where motion sequences can be classified in real time. In order to evaluate our novel method we apply it to several motions typically found in a motion capture database as generic examples and to a Judo referee signal movement database as a second class of specific example.

For the purpose of evaluating different aspects of our method we apply our technique on the first hand to prerecorded high quality motion capture data, and on the other hand to live captured low quality motion data obtained by a Microsoft Kinect sensor. All these motions can then be compared in realtime with previously recorded sample motions in the database. A detector then detects if the performed motion is similar to an annotated motion contained in the database.

Our approach uses a framework for data driven motion synthesis and motion reconstruction using k-nearest neighbor searches. Devising techniques for fast and robust k-nearest-neighbor searches for motion segments and nearest-neighbor-classifications of motion data we show that these techniques can be adapted and are then also very suitable for the task of action recognition for data obtained by a Microsoft Kinect sensor.

Our techniques for action recognition require very little preprocessing—only sample motions for each action to be recognized have to be labeled by the name of the action. No further explicit learning phases are required.

2. Overview of the Action Recognition Method

The workflow of the proposed action recognition method can be divided into three distinct steps. First, in an offline step, the database is created from motion data and annotated by hand with the actions of interest. Then, in a preprocessing step, a kD-Tree is created using an application specific featureset, allowing fast k-nearest neighborhood searches on the poses in the database.

The online phase consists of recognizing actions from an input motion sequence by feeding new frames of the input motion into the annotation module, which uses similar poses retrieved from the kD-tree in a neighborhood graph called *action graph* to output all recognized actions as soon as they are detected.

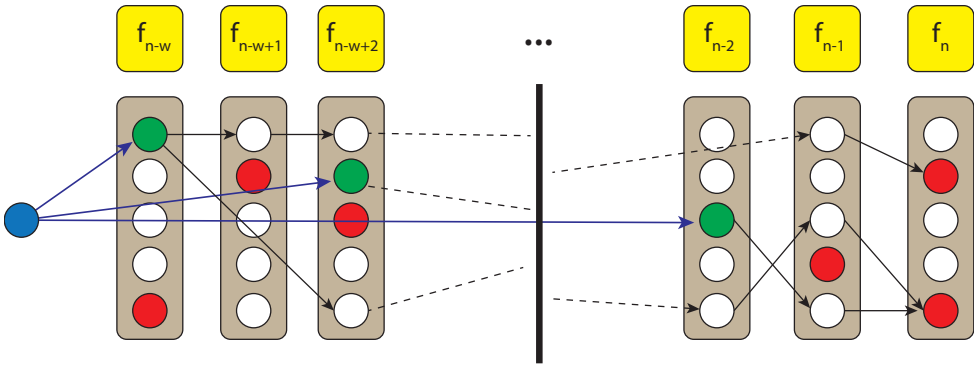

Figure 1: *Finding action path candidates for frame f_n using the action graph with window size w. First, all poses annotated with starting poses of actions (green) are connected to the single source vertex required for the single-source shortest path algorithm, regarding all past neighborhoods up to a certain window size w, which is set to 1024 in our experiments. Now, for every neighborhood, poses are connected with edges according to the allowed time and pose steps. Then, for every candidate path terminating at an action ending pose (red), the path's nodes are checked if they are consistently annotated with the same action, in which case this action is reported as found.*

2.1. Action Graph based Recognition

Detecting actions on the basis of individual poses does not take into account the temporal continuity of the underlying motion. Rather, our action recognition method detects if an action ends at the current frame and then tries to find possibly time-warped motion segments spanning the action in its entirety.

To bridge the gap from locally matching poses from the retrieved pose neighborhoods of the query motion to globally matching similar motions annotated with specific actions in the database, the proposed method uses a modified version of the *Lazy Neighborhood Graph* proposed by Krüger et al. [KTWZ10]. In their work, a directed acyclic graph is constructed by regarding the retrieved neighboring poses of the normalized query motion as vertices. An edge connects a pair of neighbors, if certain stepsize conditions are satisfied, similar to *Dynamic Time Warping*. Then, a single source vertex is connected to all the neighbors in the first neighborhood, which reduces the problem for finding a motion contained in the database which is most similar to the query motion to solving a single-source shortest paths problem. The entire global matching can be solved in $O(km\log(n))$, where k is the number of retrieved nearest neighbors, m the number of frames contained in the query motion and n the number of frames in the motion capture database.

In contrast to the LNG, the developed action recognition framework tries to find motion segments which start close to the beginning of an annotated action having the currently processed frame close to the terminating frame of an annotation (see Fig. 1). This is accomplished by first inspecting the current pose neighborhood for annotated ending poses. Now, every annotated action starting pose containing the same annotation as the found ending pose in the queued neighborhood is connected with the single source vertex mentioned above. This results in paths from the beginning of an action to the end of the action, containing only the specified annotation and which are possibly time-warped according to the allowed time-steps.

3. Conclusion and Future Work

This poster presented a method to automatically detect human full body motions using the skeletons obtained from optical motion capture systems as well as from a Microsoft Kinect camera. Since our approach is not restricted at all to streams of depth image data such as obtained from the Kinect, it could also be used with other motion sensor devices. We presume that action recognition works well with 4 or even less accelerometers attached to arms and legs—as even the more complex task of motion reconstruction from sparse accelerometer data has been shown to work with these settings [TZK*11] for many motions.

References

[KTWZ10] KRÜGER B., TAUTGES J., WEBER A., ZINKE A.: Fast local and global similarity searches in large motion capture databases. In *Proceedings of the 2010 ACM SIGGRAPH/Eurographics Symposium on Computer Animation* (Madrid, Spain, July 2010), SCA '10, Eurographics Association, pp. 1–10. 2

[TZK*11] TAUTGES J., ZINKE A., KRÜGER B., BAUMANN J., WEBER A., HELTEN T., MÜLLER M., SEIDEL H.-P., EBERHARDT B.: Motion reconstruction using sparse accelerometer data. *ACM Trans. Graph. 30* (May 2011), 18:1–18:12. 2