Human Motion Analysis for Video Annotation

Rui Rodrigues[1,2], Rui Neves Madeira[1,2], and Nuno Correia[1]

[1] NOVA LINCS, DI, FCT, NOVA University of Lisboa, Caparica, Portugal
rfis.rodrigues@campus.fct.unl.pt, nmc@fct.unl.pt
[2] Sustain.RD, ESTSetúbal, Polytechnic Inst. of Setúbal, Setúbal, Portugal
rui.madeira@estsetubal.ips.pt

Abstract. This paper aims to present ideas and solutions, fostering discussion on how to get advantage from annotating video with the support of automatic human motion recognition. This work arises in the context of previous activities carried out with dance professionals, where a Web annotator called MotionNotes was created and tested. This tool has a feature which makes possible to work with real-time human pose estimation. Users were able to identify and recommend several scenarios in which they considered that motion analysis could be a very important asset in video annotation. Video background subtraction to focus on motion and specialized motion annotations are two new features presented in this paper. We strongly consider that integrating different perspectives to analyse data and interact with them is the best path to take advantage of automatic motion recognition on video.

Keywords. Motion Tracking, Human Pose Estimation, Video Annotations, Performing Arts, HCI.

1. Introduction

Applications that provide users with multimodal interaction [1] for creative processes are usually agnostic about media content, being unable to recommend tailored elements that may be useful to improve productivity or creativity [2, 3]. In order to create a connection between the software features and multimedia data, it is necessary to design models and develop techniques to implement intelligent interactive tools [4].

MotionNotes [5, 6] is a Web-based multi-platform video annotator to support multimodal annotation that can be applied to several working areas, being designed to assist the creative and exploratory processes of both professional and amateur users, working with a digital device for personal annotations. This tool implemented a motion detection system, which works in real-time, achieved through the application of machine learning techniques [7, 8].

The idea for this paper came from informal discussions carried out in MotionNotes test sessions. The evaluation group was composed of seven elements aged between 19 and 52 years old, all with a professional dance background. Several users have reported an acceptable performance in terms of identifying body parts. However, it was also reported a lack of motivation in using the motion feature while doing the regular annotation work, i.e., an increase of performance or creativity was not obvious, as it was initially expected. Several ideas have been discussed with one goal in mind, to make the best use of this motion capture functionality. Everyone agreed with the necessity to add more features and options on top of motion tracking. It was then proposed the possibility to save all the coordinates identified in the video, i.e., each pose of all body parts from every person, kept for later analysis. By having the information processed and stored, the creation of new analysis and annotation scenarios could be possible, as described in the next chapter.

2. Scenarios for Human Motion Analysis

The first scenario proposed by some volunteers was the background subtraction. This means, for instance, reproducing the body parts motion, on a skeleton format, in the same timeframe as the source video, but with a clean background. On top of that, it was suggested the possibility to select the skeleton colour, background colour and audio activation.

Regarding the second scenario, it was based on the idea of having a special type of annotation associated to the movement. This type of annotation should be optional and could be activated or deactivated depending on motion tracking status. In this scenario, users could concentrate only on movement in a specific annotation iteration, leaving other elements for future work.

These two scenarios would require several changes in terms of software structure. Currently, there is no mechanism or structure to store motion. Therefore, there are no communication channels between the application and the server that allows this type of information to be transported over the Web. The user

interface also needs several improvements to be able to interact with the new features. In this paper, we propose a first conceptual architecture to support these two scenarios.

It is crucial to develop a conceptual model to support human motion data. Consequently, it is possible to recognize the clear necessity for an entity to support the "motion" contained in the video. This motion will consist of several "poses" that were identified during the analysis. Therefore, a support for all coordinates "points" for each body part must be created, and each point also needs to contain a detection certainty score and the information regarding which body part is being saved (Figure 1).

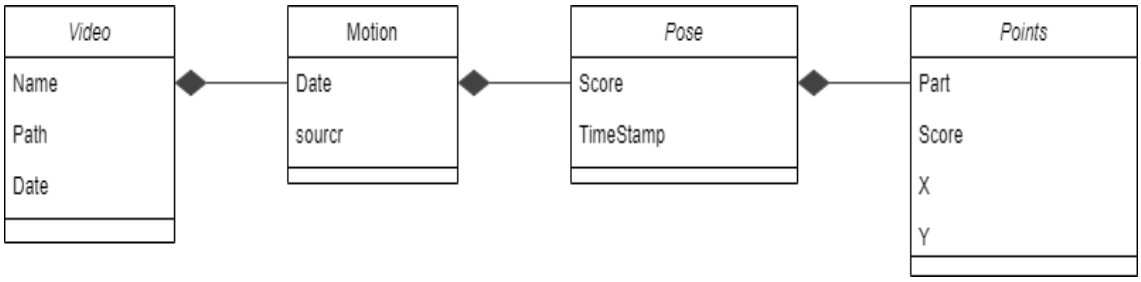

Figure 1- Conceptual model to support motion.

This structure will make possible the storage and processing of data to better inform users about human motion. With the implementation of this feature, it will be possible to hide what is happening in the video background and continue reproducing the human motion contained in it. A first test version of the system can be seen in Figures 2 and 3, where on the first one is the standard version. The second one illustrates the same pose with background subtraction.

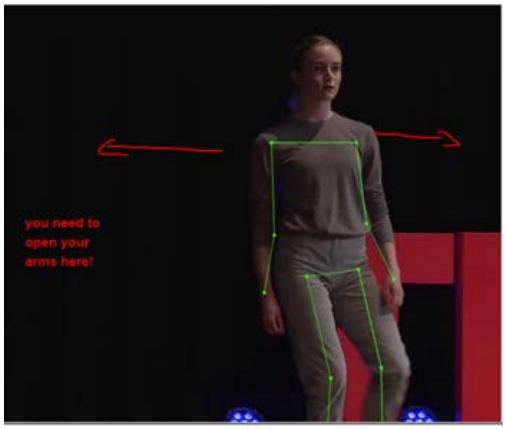

Figure 2- Pose estimation with background

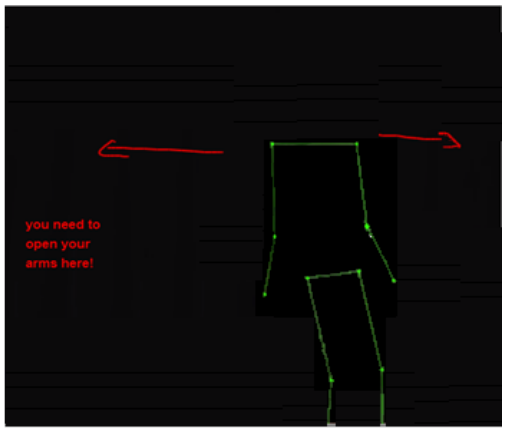

Figure 3 – Pose estimation with background subtraction

3. Conclusion

This paper describes new scenarios to improve the motion tracking feature, which is part of a tool called MotionNotes. These innovations result from several discussions and suggestions collected in meetings with dance professionals. Technical developments are being carried out with the objective of implementing the required interaction. It was necessary to study a new structure for transport, storage, and processing of motion tracking, which was described above. This new platform opens horizons and creates bases for future proposals, such as the motion correction by the application user while doing the annotation work, for instance. The evolution of this tool will soon be available to a group of volunteers to measure the impact of these new features.

Apart the specific case of identifying and using the motion to enrich user experience, we have the strong conviction that applying other Artificial Intelligence techniques will have a significative impact on the Human Computer Interaction area for creativity solutions in the near future. We are studying and working on new approaches to make the software interface and interactions more intelligent to improve productivity and even foster users' creativity.

4. References

[1] M. Turk, “Multimodal interaction: A review,” *Pattern Recognition Letters*. 2014.

[2] A. Liapis an.liapis@gmail.com, G. N. . Yannakakis, C. Alexopoulos, and P. Lopes, “Can Computers Foster Human Users’ Creativity? Theory and Praxis of Mixed-Initiative Co-Creativity.,” *Digit. Cult. Educ.*, vol. 8, no. 2, pp. 136–152, 2016.

[3] J. Frich, L. MacDonald Vermeulen, C. Remy, M. M. Biskjaer, and P. Dalsgaard, “Mapping the landscape of creativity support tools in HCI,” *Conf. Hum. Factors Comput. Syst. - Proc.*, pp. 1–18, 2019.

[4] S. Amershi *et al.*, “Guidelines for human-AI interaction,” in *Conference on Human Factors in Computing Systems - Proceedings*, 2019.

[5] S. Ribeiro, R. Rodrigues, N. Correia, and C. Fernandes, “MotionNotes,” 2019. [Online]. Available: https://motion-notes.di.fct.unl.pt/. [Accessed: 16-May-2020].

[6] R. Rodrigues, R. N. Madeira, N. Correia, C. Fernandes, and S. Ribeiro, “Multimodal Web Based Video Annotator with Real-Time Human Pose Estimation,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11872 LNCS, pp. 23–30, 2019.

[7] Y. Chen, C. Shen, X. S. Wei, L. Liu, and J. Yang, “Adversarial PoseNet: A Structure-Aware Convolutional Network for Human Pose Estimation,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2017.

[8] Z. Cao, T. Simon, S. E. Wei, and Y. Sheikh, “Realtime multi-person 2D pose estimation using part affinity fields,” in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017.