

Movement Description and Gesture Recognition for Live Media Arts

Prashant Aparajeya
Goldsmiths
University of London
London, UK
p.aparajeya@gold.ac.uk

Vesna Petresin
Goldsmiths
University of London
London, UK
v.petresin@gold.ac.uk

Frederic Fol Leymarie
Goldsmiths
University of London
London, UK
ffl@gold.ac.uk

Stefan Rueger
Knowledge Media Institute
The Open University
Milton Keynes, UK
stefan.rueger@open.ac.uk

Categories and Subject Descriptors

I.2.10 [Vision and Scene Understanding]: Motion, Shape, Video analysis; I.4.7 [Feature Measurement]: Size and shape; I.4.9 [Image Processing and Computer Vision]: Applications

The research aims to develop novel techniques able to recognise different sequential gestures, to the level where they will describe and compute articulated movements in real time. In the context of live media arts, the research outcomes would change the paradigm of creating, learning, performing, designing for live media arts, by giving feedback on performance after analysing, in real time, the streaming video of the performance.

Understanding how we recognise and respond to rhythm patterns kinaesthetically, may support the development of movement description and gesture recognition in real time [5].

One of the main problems in computer vision is defining and describing movement [6]. Traditional approaches mainly rely either on different body-mounted sensors or marker-oriented approaches [2]. Here, the performers have to adapt their natural performance or require substantial training to adopt the tools in their performance. Our principal focus is the development of a system that will free performers from constraints of wearable technology, and allow them to perform in a natural way, without movement and coordination restrictions. Secondly, this will also allow greater precision in analysing the movements while tracking the performance, thus obtaining higher levels of details of movement, allowing in turn the performers to track their performance with greater ease.

Key problems in computer vision for the live media arts include the lack of realtime feedback on performance, the inability to seamlessly control body movement in response to audio and visual output, and the unavailability of consistent feedback and evaluation to the performers on their practice in real time. The main goals of our work are:

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CVMP '15 London, UK

© 2015 ACM. ISBN 123-4567-24-567/08/06...\$15.00

DOI: 10.475/123_4

1. To make an analytical tool for performers that will act as a real-time evaluator and assessor by providing generative feedback on their performance. Expert performers are involved in the training of the evaluation system in order to generate valid ground truth-data. Our research in computer vision and information retrieval algorithms provides an algorithmic chain capable of [1]: (i) analysing shapes in sequential video-frames, (ii) extracting vision based features that will be then used to detect the performers' gestures and movements. Here, the input and feedback of expert performers will be used in order to build a vocabulary of gestures and to validate the results. This will create a feedback loop between performer and the system that will also be useful in the implementation of state-of-the-art machine learning techniques to improve the performance and accuracy of the system over time. In the context of live media arts practice, it will stimulate new approaches to creating, rehearsing, performing and evaluating work.
2. To make a creative tool for multi-media augmentation of live performance. Once the shape has been described and analysed, the system can be used as a tool to augment the surrounding and the stage with multi-media techniques (e.g. projections, interactive audio, responsive lighting, robotics) tightly synced with the movements of the performers and without affecting their freedom of movement.
3. To make a learning tool that will act as a virtual personal tutor to performers with different levels of experience. The system will be able to produce detailed descriptions of the performer's shape and movements. This information, coupled with machine learning techniques [3], will allow the training of the system with the input of a number of expert performers. Once the system will reach a level judged acceptable by the experts, it will be possible to use it as a virtual tutor for performers with different levels of experience. The trainee will also be able to choose the preferred learning style, depending on the styles of the experts that initially trained the system. Moreover, colours and textures (e.g. costumes, stage materials, etc.) can be added to provide more information for improving accuracy.

We present our most recent results in applying a novel representation of dynamic shapes when undergoing articulated movement.

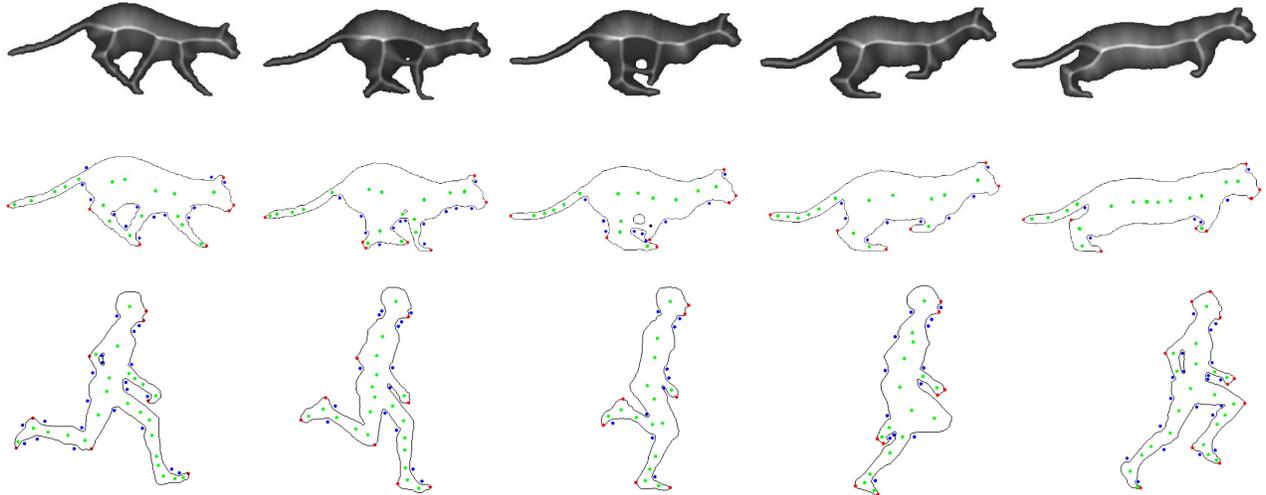


Figure 1: *1st row:* Interior medialness map for a sequential set of frames of the movement of a running cat. The maxima (white spots) of the function are good candidates as primitives for biological motion representation as “hot spots”. *2nd row:* Changes in feature point loci — medial hot spots as green dots, significant convexities and concavities as red and blue dots — for the (same) running cat. *3rd row:* Changes in feature point loci for a few frames of a running athlete captured by E. Muybridge in 1887.

Our method is based on an adaptation of perception-based results [4] including the codon features and medialness hotspots [5, 1]. In every frame of a video sequence (from a single camera) we can reliably obtain a shape representation which expresses in a compact way the highest values of medialness (aka hot spots) and the most descriptive convexities and concavities along contours (Figure 1). An example of our work on a short video sequence is available on YouTube: <https://youtu.be/OVbLvX57vhY> (Figure 2).

The outcomes of the research may lead to applications in areas where a change of shape is indicative of variations in characteristics of the observed object in motion or under transformation: VFX, gesture interface development, enhancing learning for disabilities (e.g. improve coordination through feedback), medical procedures (enhanced feedback on surgical procedures, anomaly detection), scientific measurements and observation of dynamics processes (e.g. in graphonomics).

1. REFERENCES

- [1] P. Aparajeya and F. F. Leymarie. Point-based medialness for 2D shape description and identification. *Multimedia Tools and Applications*, May 2015. DOI: 10.1007/s11042-015-2605-6.
- [2] E. Ceseracciu, Z. Sawacha, and C. Cobelli. Comparison of markerless and marker-based motion capture technologies. *PLoS One*, 9(3), 2014.
- [3] A. Elhayek et al. Efficient ConvNet-based marker-less motion capture in general scenes. In *Proceedings of IEEE CVPR*, pages 3810–8, 2015.
- [4] I. Kovács. "Hot spots" and dynamic coordination in Gestalt perception. In *Dynamic Coordination in the Brain*, chapter 14, pages 215–28. MIT Press, 2010.
- [5] F. F. Leymarie, P. Aparajeya, and C. MacGillivray. Point-based medialness for movement computing. In *ACM Int'l Wksh on Movement & Computing*, pages 31–6, 2014.
- [6] S. Saini et al. Markerless human motion tracking using hierarchical multi-swarm cooperative particle swarm optimization. *PLoS One*, 10(5), 2015.

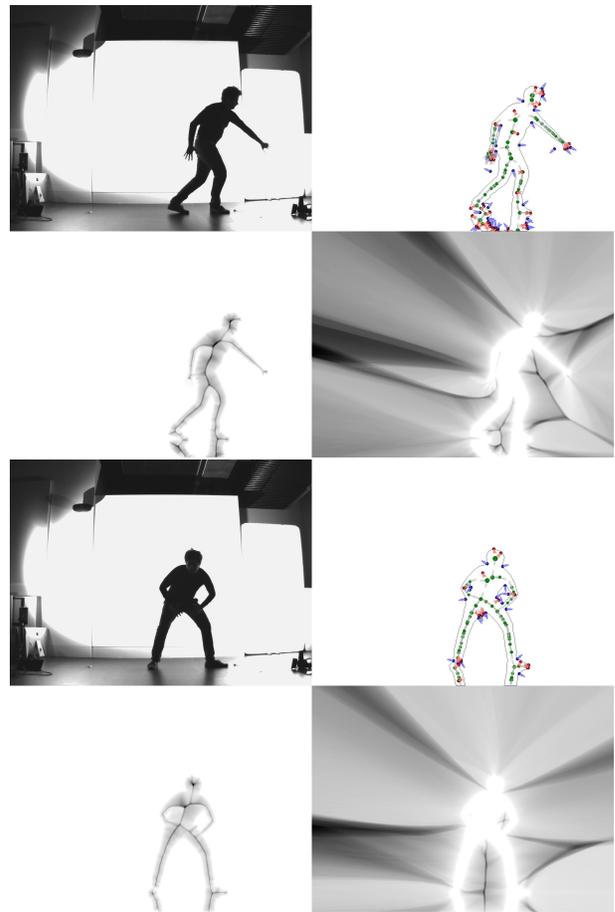


Figure 2: Two video frames processed to extract shape features in terms of oriented hotspots (green dots), and significant convexities and concavities (blue and red arrows). Interior and exterior medialness maps are also shown.