Techniques for automated analysis of saliency in human full-body movement

Serafino Gabriele
Lo Luca

2024

Abstract

In contemporary research, the topic of salience has become a focal point, with a growing number of studies dedicated to its exploration. Various fields, including Medical, Financial, Military, Psychology, Kinesiology, have approached the problem in diverse ways.
Our thesis aims to develop effective heuristics and algorithmic approaches for detecting salience in the context of movement. Specifically, we focused on the domain of dance movements, defining a concept of salience within this context.
In our purpose the salience is meant as a change in the behaviour of the dancer: static movement with respect to dynamic movement, repetitive movement with respect to a chaotic one, using an arm and then stopping it for using the other one, ans so on...
The study involves a thorough analysis of Motion Capture (MOCAP) data, where we navigate dataset constraints and boundaries to select significant features that aptly represent our chosen domain.
To address the salience detection problem, we employ several approaches: among them, a Supervised Machine Learning (ML) Algorithm.
For the ML approach we defined a specific prototype of Sliding Windows (SW) able to convey information through several frames of the videos, with the goal of giving to the specific model the most significant piece of information in order to have the best results.
Then we tested the algorithm at different levels of generalization.
The goal of this work is to develop a system able to recognize the salience in human full-body context in real time.

Contents

List of Figures

List of Tables

Acronyms

3D Three dimensional. 10

EST Event Segmentation Theory. 11

ML Machine Learning. 2

MOCAP Motion Capture. 2, 9, 10

QTM Qualisys Track Manager. 10

SW Sliding Windows. 2

TSV Tab-Separated Values. 10, 12

1 Introduction

1.1 The problem of automatic salience

1.2 Motivation and goals

1.3 Thesis structure

2 Methodologies

2.1 CPD algorithms

2.2 Related Work

The material we started from was provided by researchers at
Casa Paganini - InfoMus [3]. By exploiting the existing results, our goal was to find algorithms and heuristics for dealing with the problem of salience.
The given scenario is the following: three dancers from Casa Paganini - InfoMus [3], Cora Gasparotti (Figure 1), Marianne Gubri (Figure 2) and Muriel Romero (Figure 3), performed dancing movements by emphasizing some specific tasks, in order to highlight a combination of graceful and fluid movements, as well as angular and hurried gestures.

Those performances has been recorded with two professional cameras (frontal and lateral view, 1280×720, 50fps) at Casa Paganini, and the result of this procedure was a set of *takes* (each *take* has a duration between 30 seconds to 180 seconds).
So, for each dancer, we have between 5 to 15 *takes*.

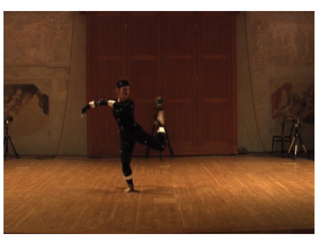

Figure 1:
Cora Gasparotti

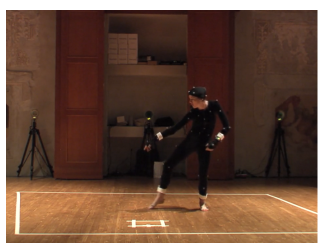

Figure 2:
Marianne Gubri

Figure 3:
Muriel Romero

Raw data has been collected by using a Qualisys [4] Motion Capture (MOCAP) system:

- Qualisys Cameras: these are specialized for capturing and tracking movements in three-dimensional space.
  These cameras operate based on the principle of detecting and analyzing markers placed on objects or individuals within their field of view. The cameras emit light, which is then reflected by small, passive markers, such as reflective spheres or retro-reflective devices. This process allows the cameras to discern and meticulously track the position and movement of each marker within their field of view.
  In our context, a Qualisys system endowed of 16-cameras was used (fs = 100Hz). (Figure 4)

Figure 4: Camera Qualisys

- Markers: are integral components of motion capture systems, they are identifiable points placed on objects or the human body. These markers can take various forms, such as reflective spheres, strips, or discs, each designed for specific tracking purposes.
  In our cases we used 64 spherical markers positioned on the whole body. (Figure 5)

Figure 5: Markers Qualisys

The MOCAP data was collected in a Tab-Separated Values (TSV) file by using Qualisys Track Manager (QTM) [4] which is a software used for files handling MOCAP data, hence for each frame the position of each marker in the 3D space (x,y,z) is available.

2.2.1 State of the art

One of the most important work related to our context was the one done by E. Ceccaldi in her PhD Thesis [2], in that discussion, the problem of salience detection, has been defined with some psychological approach related to the body behaviour, for the definition of the salience, *takes* have been segmented (in the time).
For the sake of dividing the movements in events, the segmentation was applied by using Event Segmentation Theory (EST) concepts as they were proposed in Segmentation Theory [5].
Thus, salience is defined to a higher level as follows:

- C1. Time: which is the timing and the rhythm of the movement, for instance if the dancer accelerates her actions.
- C2. Space: if the attention and the direction of the movement starts to pointing something different.
- C4. Character Location: if the character moves on the stage, or, in our cases, if the dancer moves from a point to another.
- C6. Causes: Which are the causes and appraisal, if a new state of affairs leads to a subsequent event.

Hence the ground truth was taken by considering these 4 different aspects by psychologists which are trained on EST [5] principles by using a software for segmentation: ELAN [1] (Figure 6).
This software allows to make the segmentation of videos (or an audio track) by defining some classes of event segment and a hierarchy among them.
The *features* which have been considered in that case are the following:

- For the *Time* it was used the General Quantity of Movement and the Chest Quantity of Movement.
- For the Space the author employed the Directness of Head Movement.
- The Character Location has been detected by looking the Density of Chest Trajectory.

- The causes, instead, has been used only has an additional ground truth.

All these features has been extracted by using an application (patch) developed for EyesWeb XMI (Figure 7), a software platform able to build over the TSV, which contains our raw data, some useful characteristics such as: kinetic Energy, Point density, Directness, and so on...
For instance, kinetic energy (of all the markers) was taken as a measure of the overall amount of movement.
The Chest Quantity of Movement was evaluated by considering the kinetic energy of the marker related to the chest. The Directness of Head is a measurement of how the dancer's head moves in curvilinear trajectory, the higher is the value, the more the movement following a straight line.
For finishing, the Density of chest trajectory is an indicator of whether movement is localized in a small region in the space rather than spanning the whole space, i.e., higher density indicates that the actor has moved in a smaller region.

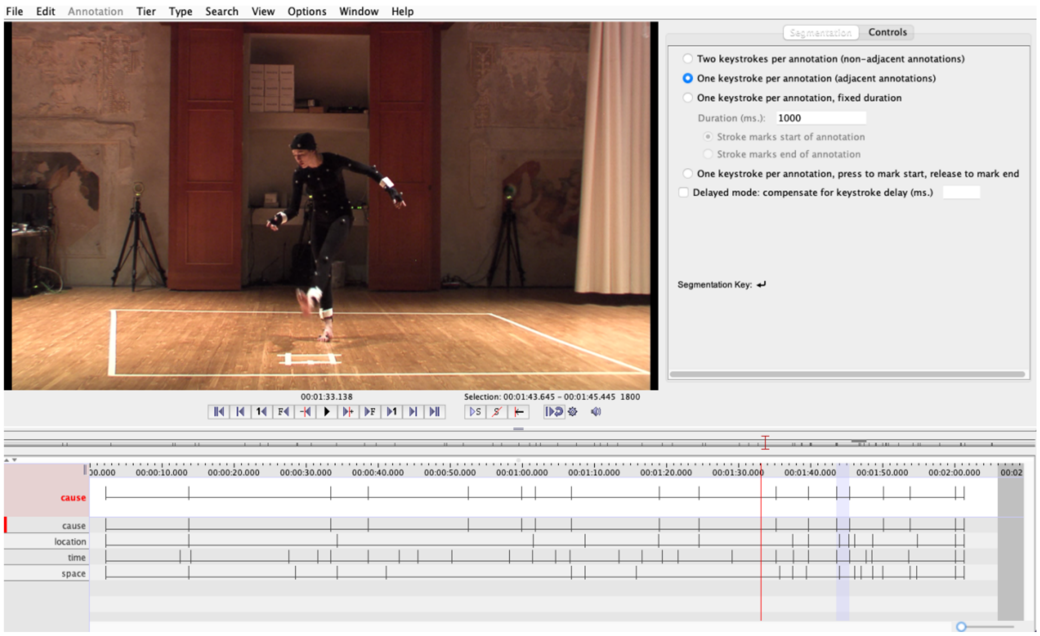

Figure 6: Annotation of the different salience with ELAN.

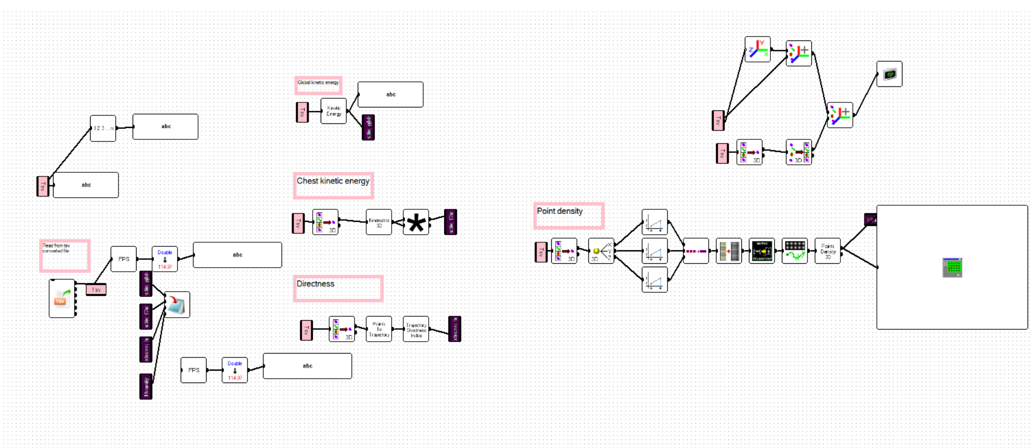

Figure 7: One of the patches
implemented on EyesWeb
for extracting the features.

2.3 Structure of workflow

References to next chapters and workflow schema.

3 Salience definition

3.1 Ground Truth

4 Low level features

5 Supervised ML approach

5.1 Data

5.1.1 Data Cleaning

5.1.2 Sliding Windows

5.1.3 Mid level features

5.1.4 Normalization

5.2 Samples Creation

5.3 Model Selection

5.4 Results

5.4.1 LOSO

5.4.2 LOTO

5.4.3 LODO

6 Statistic approach

6.1 Sliding windows

6.2 Features Selection

6.3 model

6.4 Results

7 Conclusions

7.1 Future Researches

References

[1] The Language Archive. URL: https://archive.mpi.nl/tla/elan.

[2] Ceccaldi E. “CEST: a Cognitive Event based Semi-automatic Technique for behavior segmentation”. PhD thesis. Università di Genova, 2022.

[3] Casa Paganini. URL: http://www.casapaganini.org/.

[4] Qualisys. URL: https://www.qualisys.com/.

[5] J.M. Zacks and K.M. Swallow. “Event segmentation”. In: *Current directions in psychological science* 16.2 (2007), pp. 80–84.