Using Motion Expressiveness and Human Pose Estimation for Collaborative Surveillance Art

Jonas Aksel Billeskov, Tobias Nordvig Møller, Georgios Triantafyllidis, and George Palamas(✉)

Aalborg University Copenhagen, Copenhagen, Denmark
{tnmal4, jbille14}@student.aau.dk,
{gt, gpa}@create.aau.dk

Abstract. Surveillance art is a contemporary art practice that deals with the notion of human expressiveness in public spaces and how monitoring data can be transformed into more poetic forms, unleashing their creative potential. Surveillance, in a sociopolitical context, is a participatory activity that has changed radically in recent years and could be argued to produce, not only social control but also to contribute to the formation of a collective image of feelings and affects expressed in modern societies. The paper explores a multidisciplinary approach based on tracking human motion from surveillance cameras on New York Time Square. The performed human trajectories were tracked with a real-time machine vision framework and the outcomes were used to feed a generative design algorithm in order to transform the data into emotionally expressive 3D visualizations. Finally, a study was conducted to assess the expressive power of this approach so as to better understand the relationships among perceived affective qualities and human behaviors.

Keywords: Generative art · Surveillance data · Motion expressiveness
Data transformation · Human pose estimation · 3D visualization

1 Introduction

In an age where a large part of an everyday life happens online and mostly everything done online will leave a digital footprint, people are getting increasingly concerned with digital privacy. Some go towards big data and endorse of collecting it for the use in advertisement targeting and scientific research, while others are against, hiding their digital traces by using VPN servers, proxies or boycotting sites such as Facebook. It is a difficult area to maneuver, because it is not certain exactly what data is being saved or how it is being used.

The proposed paper describes a method of how data can serve artistic purposes, in the context of generative art, and how using personalized public data does not have to be intrusive of personal privacy. Data can be used to create art pieces, where the content represented by the data may or may not be apparent. There have been examples of weather data being turned into complex sculptures such as the work of Nathalie Miebach [1] or Laurie Frik's sleep patterns [2] displayed in a non-traditional matter. Recent work in pose recognition allows us to capture pose data by inputting video data

A. L. Brooks et al. (Eds.): ArtsIT 2018/DLI 2018, LNICST 265, pp. 111–120, 2019.
https://doi.org/10.1007/978-3-030-06134-0_12

using the software library OpenPose [3, 4]. This project utilise surveillance camera data in order to extract pose information about the people in the footage and interpreting this to create graphical 3D representations. The art pieces should encapsulate the motion expressiveness of the surveillance data and retain this through a transformation into generative art. These art pieces can be viewed as a way of mediating reality to make people ponder about the use of data and surveillance and showing that even ordinary movement can be turned into something aesthetically pleasant.

2 Background

The background section will discuss previous work done in this field and how it relates to this project. It will also discuss other art installations, which inspired this work.

2.1 Motion in Art

Obradović and Marković [5] explore how time can be used to create 3D shapes. They do this by using time as a way of sculpting their shapes. They extrude shapes over time by changing relative scaling and rotation as the object moves through space. The models were visualized in blender and depicted the variety of different shapes that can be obtained with their proposed method. They also suggest that their shapes could be used for decoration and architecture. Using time as a way of sculpting shapes is an interesting idea and this paper will in some ways try to extend this concept.

In traditional art, the Italian painter Giacomo Balla 1871–1958, who is one of the founders of futurism, spent much of his artistic career investigating how art can convey motion. One of his most famous paintings shows a dog where the legs move so fast that they merge together (Fig. 1). Balla was also inspired by technology as seen in paintings such as "*Swifts: Paths of Movement + Dynamic Sequences*". The shapes in these paintings could seem similar to the shapes procedurally generated in the project by

Fig. 1. Dynamism of a Dog on a Leash (1912) by Giacomo Balla

Obradović and Marković [5] and explores the same concept about extending or duplicating the shape in motion to convey this motion to the viewer.

Takahashi [6] proposed a project called *Voxelman*, where he uses human motion data and presents them using voxels through Unity game engine. There is no apparent transformation to the data, the main purpose was to represent human motion expressiveness but with an altered meaning and significance.

2.2 Surveillance and Data Art

Surveillance art is a commentary on the surveillance happening in society. This could be art both made with surveillance, portraying the effects of surveillance or just exploring the concept. Data art is an art form that generates artistic content based on large data sets with an aim to create an explicit interpretation available to a larger audience [7]. Belgian contemporary artist Dries Depoorter's installation "*Jaywalking"* [8] allow viewers to look at a surveillance camera placed on a street and report jaywalkers to the police at the press of a button (Fig. 2). It is up to the viewer to consider whether reporting the jaywalker is the right thing to do. This is supposed to give the viewer a sense of power and perfectly shows how surveillance can alter social conditions. This approach creates another perspective on surveillance and puts the viewer very close to the new ethical decisions created by surveillance and data gathering.

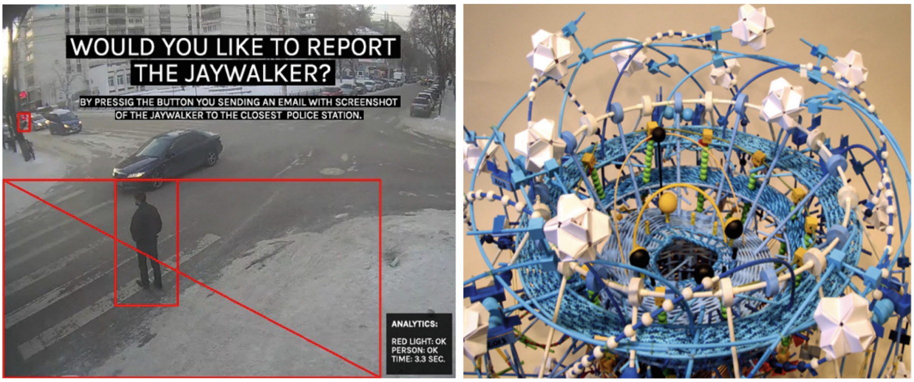

Fig. 2. (Left) Dries Depoorter's installation "*Jaywalking"* (Right) Nathalie Miebach's artwork *"The Perfect Storm"*.

Laurie Frick is a contemporary data artist working with data visualization [2]. Her art pieces use datasets and visualize them mainly with analog media. Her work "*Sleep Patterns*" are data from sleep schedules made analog with watercolors and pieces of wooden bricks. The data in these artworks retain their qualities and were not transformed into any type of metadata.

Nathalie Miebach makes art based on weather data [1] as for example "*The Perfect Storm*" (Fig. 2). These artworks were created in analog ways and are complex and multilayered, making the original data challenging to interpret.

Seb Lee-Delisle created the project "*Lunar Trails*" as an interactive art installation featured in the Dublin Science Gallery in 2013 [9]. It consists of an arcade machine with the 1979 video game "*Lunar Lander*" and a wall where the game data was rendered onto. Each game played will be rendered as a line on the wall and the final rendering will be a collective art piece of all the player's paths in the game. There is almost no transformation in the data visualization and by simply looking at the final piece the viewer can easily get an idea of how all the different players played the game. Collaborative art such as this is a great way to show trends in data in a visually pleasing way.

2.3 Generative Art

McCormack, et al. [10] writes about open questions regarding generative art. In question 5 he asks "*In what sense is generative art representational, and what is it representing?*". In this question the representation of generative art based on real life data is discussed. In generative art there is a transformation between the obtained data and the representation of data. This transformation can be of varying size and often depends on the kind of data the artist is working with and the vision of the artist. Another related question proposed is the difference between data visualization and generative art, and how this difference is defined. Edmonds [11] discusses the difference between interaction and influence in generative art. Influence is different from interaction as the influence doesn't have to be direct and can be a prolonged effect. This is relevant to this project as the people influencing the artwork will not know their data is being used and their influence is not a direct interaction with the art piece.

3 Design and Implementation

The design of the proposed generative art workflow (Fig. 3) was influenced by Obradović and Marković [5] and Balla [6] in the way they use time as a way to shape their models and the way complex shapes could be created from a continued shape. OpenPose [3] was used to read data from a surveillance camera looking at Times Square in New York with an aim to extract a sequence of human pose estimations. The pose data was used to create a 3D shape from the motion of the person [4] under

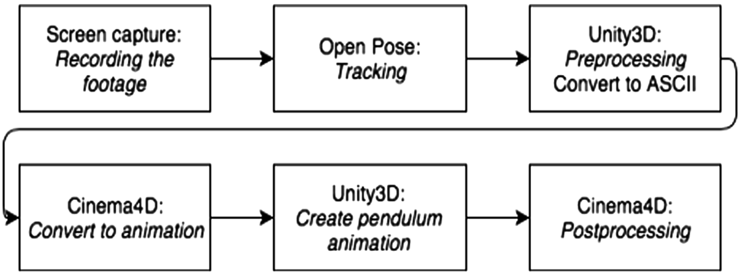

Fig. 3. Flowchart of the implementation process

surveillance. The data then was being processed with a pre-rendering software. Using pre-rendering instead of real-time graphics allowed for better lighting and for a wide variety of materials and textures.

To improve the feeling of having the data in surveillance being used for a different purpose, a simple physics-based transformation were added to the system. This transformation had the form of a double joint pendulum controlled from the human motion under surveillance [12]. The aim of this transformation is to create an artistic representation, rather than a pure data visualization, which retains the human expressiveness of the original data.

3.1 Implementation

This section describes the steps of data collection and the combination of physical pre-processing and conditioning of the motion captured signals.

Data Collection. This step comprises two aspects which are of importance for displaying the human motion expressiveness. First, the method of surveillance camera data collection and second, the data type that was collected. The surveillance camera, which was chosen, is a typical web based camera at Times Square in New York (Fig. 4). The selection of the camera was mainly due to the angle and position provided good data to be analyzed later on. A total of 45 videos captured and saved as individual clips with each of them depicting a single person passing by.

Fig. 4. Screenshot from the surveillance camera from Times Square.

Analyzing Data with OpenPose. The individual clips were fed to the OpenPose software to extract pose information about the person's motion. The software output JSON-files with information regarding 17 two-dimensional key points per person. OpenPose analyzes the data one frame at a time, which means that a 10 s clip at 30 frames per second would output 300 files. JSON-files were used because the data structure makes it easier to transfer the data between different software.

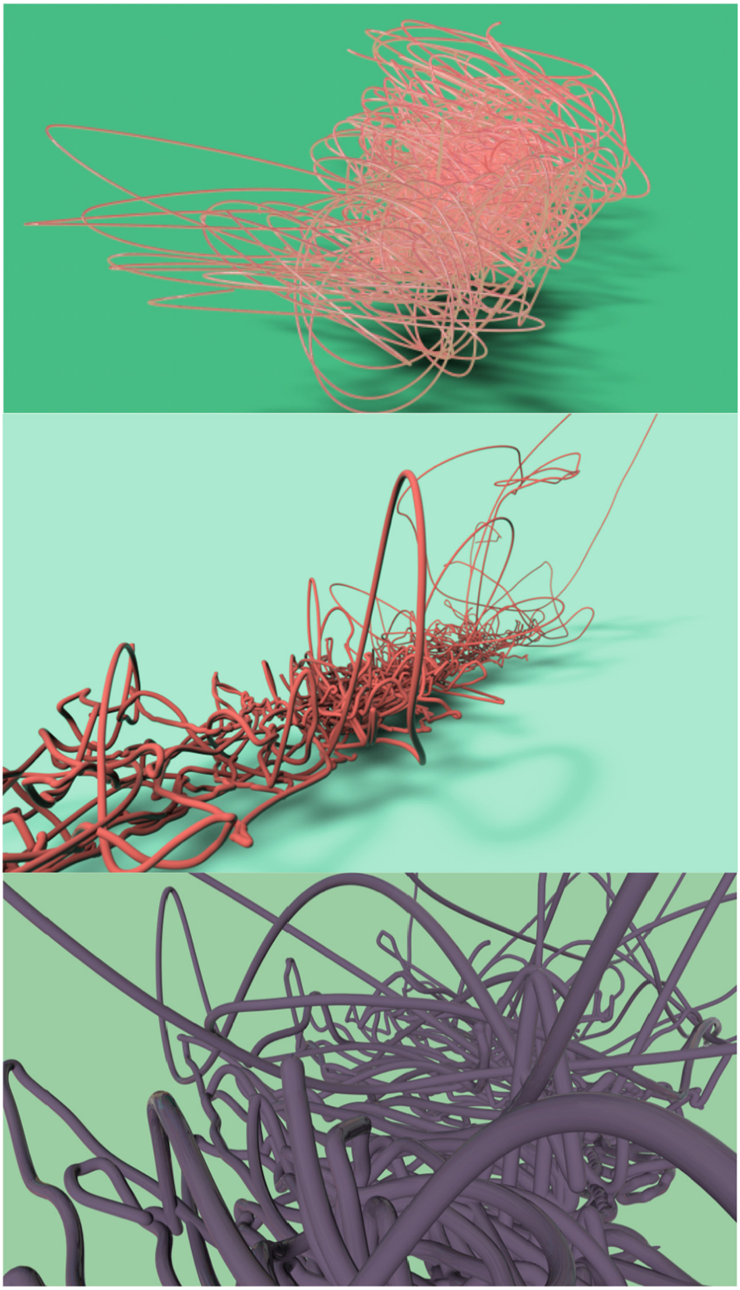

Fig. 5. The performed human trajectories (45 individuals) were used to feed a generative design algorithm in order to transform the data into emotionally expressive 3D visualizations.

Converting the Data. The files exported from OpenPose were then imported into Unity3D using a custom C# script. Each person's trajectory visualized by a single line based on a single point that was calculated as the average of all the key points generated from the human pose estimator. The z-value was then calculated by taking the average length from all key points to the average x- and y- value. This means that the more spread out a person's pose is, the greater the z-value and vice versa. Finally, the data converted into ASCII format and written to a file in order to be used in Cinema4D, using another custom C# script.

Data Transformation. The double-joint pendulum was implemented in Unity3D with 2 spheres connected in a row with hinge joints. Both spheres were equal in size and weight. The data from Cinema4D was then imported into Unity3D as an animation track for the top sphere to follow. The trajectories of the bottom sphere were tracked and then exported back into Cinema4D as ASCII data for final processing and visualization. The pendulum used to transform the data in a way, where the output would be naturally looking with a physically valid motion.

Final Processing and Visualization. In the data visualization stage a number of 3D trajectories produced (Fig. 5) based on the transformed data of the associated human motion. These trajectories were again turned into splines and visualized with a sweep tube following the path generated from the pendulum physical model. Different tubes were rendered with a variety of materials and lighting conditions against a composite ground plane.

4 Evaluation

In order to evaluate the human expressiveness of the visual outcomes, a questionnaire based on the Russel's circumplex model [13] used to assess affective and cognitive properties. The questionnaire based on 5 video clips from a surveillance camera on Times Square, New York with one person passing by at a time. The same 5 video clips were processed in the manner explained in the implementation chapter, and the 5 shapes were presented in the questionnaire. All video clips and shapes were presented in a random order in the questionnaire. The respondents were asked to describe the shape and video clips using a list of 15 different affective states. The Russel's affective model [13] suggests that emotions can be arranged around a circle in a 2D space, with arousal and valence as the dimensions. Based on this model, affective states such as relaxed and calm are related and similar to each other, as they are positioned close to each other in the circle. Affective states such as sad and happy are opposites as seen by their position on each side of the circle. The respondents were asked to choose only one affective state that better describes their feelings in order to evaluate the expressive power of the visuals.

The main idea behind this evaluation was to compare the affective states expressed by the generated shapes to the corresponding video clip, in order to see if a correlation existed or if the generated shapes evoked different affections compared to the original video footage. The reason for researching this correlation was to evaluate whether a transformation of motion data, using a simple physical model, can still contain the same expressiveness of affective states as the actual video data.

5 Results

The questionnaire was completed by 45 respondents. The following section describes the results of this comparative study, where video clips containing single persons walking directly compared to the corresponding 3D representations (shapes).

The first person and the derived shape show almost no similar results. The person was mostly perceived as calm and relaxed while the shape was perceived as nervous and stressed. The reason for this could be the person looking relaxed but walking in a high tempo resulting in a jagged shape with sudden turns formed from his movements. In general, the answers were scattered for both the person and the generated shape. The second person and the derived shape show some similarities as both were perceived as

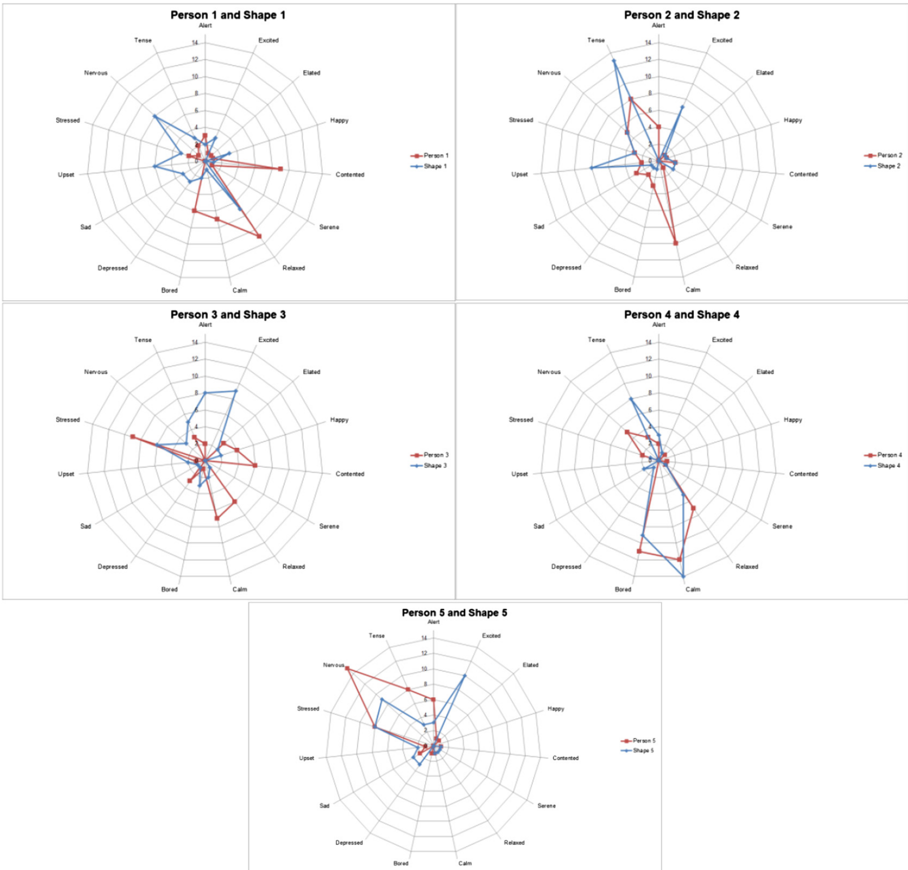

Fig. 6. Respondents' answers for different sets of human motion and generated visuals

tense and nervous. The person was also seen as alert while no one perceived the shape as alert. Despite some perceived the shape as tense, nervous and alert, the video clip is seen as calm, the complete opposite state of affection to the first three affective states.

As the results in the next section show, most generated visuals perceived with almost opposite affective states compared to the corresponding surveillance video (person 1–3). The persons were perceived mostly as calm, relaxed and contented while the corresponding shapes were perceived as mostly alert, exited and tense. The answers were scattered, while the shape was fairly centered on affective states with high arousal. Some pairs of person motion and associated shape (person 4 and 5) were perceived as similar with affection states very close to each other in the model (Fig. 6).

6 Conclusions

The results of the evaluation did not show conclusive results on whether a specific motion expressiveness of affective states could be retained through such a transformation to a 3D shape. This could be due to various conditions and this section will discuss some of these. The video clips used in this evaluation were based on people passing by and given the simple nature of these video clips, the motion of the persons were not very expressive or lacking granularity of expressed affection. For the first three videos, the data pointed in different directions of the circumplex model of affect and no single affective state from a 3D representation was able to describe the person's affective state. However, in the last two clips there was generally more agreement on the expressiveness of the person and this was also the case for the shape. These two clips had both more focused answers and more agreement between respondents, for both the person and the derived shape. From these results can be concluded that a very expressive motion might generate a shape with the same affective state. With the Gestalt law of common fate in mind, it would have been interesting to differentiate the shapes based on their rotations and thus, avoiding participants grouping the shapes because of similar directions. The same principal applies to Gestalt law of good form, because all the shapes in this experiment had the same color.

The aim of this project was to provide new ways to explore human motion expressiveness using generative art as a way to communicate with an audience. Transforming data from surveillance cameras into generative art, while retaining motion expressiveness required a transdisciplinary approach based on state of art human pose estimation technology. Considering the amount of transformation done to the original data, the 3D representations were aesthetically appealing but without conclusive results related to the communication of affectivity.

References

1. Miebach, N.: Weather (2018). http://nathaliemiebach.com/weather.html
2. Frick, L.: Sleep Patterns (2012). http://www.lauriefrick.com/sleep-patterns

3. Cao, Z., Simon, T., Wei, S. E., Sheikh, Y.: Realtime multi-person 2D pose estimation using part affinity fields. In: Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017 (2017). https://doi.org/10.1109/CVPR.2017.143
4. Qiao, S., Wang, Y., Li, J.: Real-time human gesture grading based on OpenPose. In: Proceedings - 2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics, CISP-BMEI 2017 (2018). https://doi.org/10.1109/CISP-BMEI.2017.8301910
5. Obradović, M., Marković, S.: Creating 3D shapes by time extrusion of moving objects. In: 20th Generative Art Conference GA 2017 (2017)
6. Takahashi, K.: Voxelman, 30 May 2018. https://github.com/keijiro/Voxelman
7. Carlucci, D., Schiuma, G., Santarsiero, F.: Toward a data-driven world: challenges and opportunities in arts and humanities. In: Big Data in the Arts and the Humanities: Theory and Practice. CRC Press (2018)
8. Depoorter, D.: Jaywalking (2016). https://driesdepoorter.be/jaywalking/
9. Lee-Delisle, S.: Lunar Trails, 25 May 2012. http://seb.ly/work/lunar-trails/
10. McCormack, J., Bown, O., Dorin, A., McCabe, J., Monro, G., Whitelaw, M.: Ten questions concerning generative computer art. Leonardo **47**(2), 135–141 (2014)
11. Edmonds, E.: From interaction to influence: generating form and space. In: GA 2015 – XVIII Generative Art Conference (2015)
12. Bisig, D., Bogner, F.: Pendulum – Exploiting Simple Physics for Generative Art. In: Generative Art Conference GA 2017 (2017)
13. Posner, J., Russell, J.A., Peterson, B.S.: The circumplex model of affect: an integrative approach to affective neuroscience, cognitive development, and psychopathology. Dev. Psychopathol. **17**(3), 715–734 (2005)