# Lights! Dance! Freeze! - Exploring the Dance-musical Filmic Space Using Embodied Search in an Interactive Installation

THEODOROS PAPATHEODOROU, Computational Media and Arts, Hong Kong University of Science and Technology (Guangzhou)
JESSICA WOLPERT, Computing Department, Goldsmiths, University of London

Conventionally, spectators enjoy films passively. This paper describes an interactive film installation that invites participants to use their entire body as a query to search and explore a large corpus of musical films in a novel way. Using an RGB camera, ML-based skeleton tracking technology and a unique pose and film indexing system, this installation tracks a participant's movements and mirrors them in real-time by finding matching poses among hundreds of thousands from well-known musicals. When the participant freezes in a pose, the installation instantly plays back the short film clip that begins with that pose, immersing them in the music and dance from musicals of different eras. This approach explores themes of tangible interfaces and the new possibilities that emerge from employing embodied interaction to traverse the dance pose space, which is traditionally difficult to index and interact with in real time. The pose indexing system and whole-body interaction we propose in this paper open new pathways for cultural participation, as they lend themselves to different datasets and require no technical skills from participants.

CCS Concepts: • **Applied computing**; • **Arts and Humanities**; • **Media arts**; • **Human-centered computing**; • **Human computer interaction (HCI)**;

Additional Key Words and Phrases: interactive art, dance-musical film installation, skeleton tracking, embodied search

**ACM Reference Format:**
Theodoros Papatheodorou and Jessica Wolpert. 2023. Lights! Dance! Freeze! - Exploring the Dance-musical Filmic Space Using Embodied Search in an Interactive Installation. *Proc. ACM Comput. Graph. Interact. Tech.* 6, 2, Article 21 (August 2023), 8 pages. https://doi.org/10.1145/3597620



## 1 INTRODUCTION

*Lights! Dance! Freeze!* uses an experimental, arts-based research-creation research creation approach to explore tangible, whole-body interaction between humans and machines and the new possibilities in traversing a large corpus of data that is difficult to index, such as movement, dance and pose. We propose a circuit involving a pose tracking system, a film clips and pose database and a pose search engine that form an interactive public installation.

An early theoretical influence was the concept of expanded cinema, coined by Gene Youngblood [Sheridan and Bryan-Kinns 2008], by which he refers to the integration of different media, including film, performance and technology, to create immersive and innovative artistic experiences to expand the traditional boundaries of cinema and performance. Aesthetically, we were inspired by *The Clock* [Antle et al. 2010] by Christian Marclay shown at the Tate in London, where a montage of 12,000

Authors' addresses: Theodoros Papatheodorou, Computational Media and Arts, Hong Kong University of Science and Technology (Guangzhou), theodoros@ust.hk; Jessica Wolpert, Computing Department, Goldsmiths, University of London, j.wolpert@gold.ac.uk.

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 the author(s) 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.
© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.
2577-6193/2023/8-ART21 $15.00
https://doi.org/10.1145/3597620

movie clips with references to time were sewn together to form a looped 24-hour video that itself functions as a clock. Similarly, we wanted to allow participants to form a collage of images and movie clips from different musical eras and styles during their experience. Another inspiration was, *Land Lines* [Bausch n.d.] by Zach Lieberman where the trail a user's mouse left on the screen is matched to underlying patterns in satellite images. The idea of being able to visually match things in one domain (mouse gestures) to things in a different domain (shapes in satellite images) was appealing to us. Similarly inspiring was the children's book *Alphabet From The Sky* [Davies 2008] where alphabet-like patterns were identified in satellite images. Memo Akten's interactive installation "Body Paint" [Tsuchida et al. 2019] which allowed participants to paint on a virtual canvas using their body movements functioned as a guide on how technology could be utilised to create an immersive environment that is accessible to a wide range of participants. Inspiration for this work also came from *Move Mirror* [Exceptions to copyright n.d.] where the authors matched participants' live poses to similar poses in a random collection of static images of people performing daily activities.

Building on this idea, we linked participants' poses to videos of dance moves as well as images. To link poses to specific moments in videos we developed custom pre-processing layers to create a pose indexing system and built bespoke software to seek and play back the matching video clips and images in response to the action in front of the camera with very low latency. We also improved the skeleton tracking to deal with even the most challenging dance poses. We employed a large projection surface in a cinema and our source material was dance musicals, a genre that naturally lends itself to this project.

## 2 MOTIVATION

Since the first movie musical, *The Jazz Singer* in 1927, many dance scenes have been etched into our collective mind. Iconic scenes such as Liza Minelli's *Cabaret* performance with a chair, Gene Kelly dancing in the rain and the lift scene in *Dirty Dancing* are searched for with text queries on video sites such as YouTube. If available, they are consumed passively on a screen. In contemporary dance archives [Friedhoff and Alvarado 2018] efforts have been made to document dance in ways other than film, however, there is no facility to search using pose as input. This research makes thousands of dance poses accessible to participants in an interactive installation, by using as a query the same language they are encoded in, the human pose. In this installation, participants do not simply see the dance moves, but fully experience them in a deep, embodied way. Pina Bausch's innovative choreography, which merged dance and theatre [Youngblood 1970], emphasises the importance of the body as a medium for communication and expression. This idea resonates with our goal of allowing participants to experience embodied interaction with dance and musical films.

Finally, installations that need no technical skills to interact with nor use external devices are particularly appealing to museums as they are found to encourage participation and play [Lugaresi et al. 2019] [Lyons et al. 2012]. At the time of writing, we were in negotiations with film institutions to display the installation. Intuitive, embodied user interfaces like the one proposed in this work promote playful engagement because they capitalise on everyday experience [Marclay 2010]. In addition, exertion interfaces increase user affect, engagement and immersion into an installation [Gross 2016] [Lieberman 2016].

## 3 METHODS

Our first task was to build a big enough dataset to create a rich and diverse experience. We used musicals from different eras, from 1936's *Wizard of Oz* all the way to 2016's *La La Land*. We captured diverse cinematography and setting, but also clothing from different eras, musical styles, as well as diverse cast groups in terms of gender and race. This resulted in a selection of 50 films.

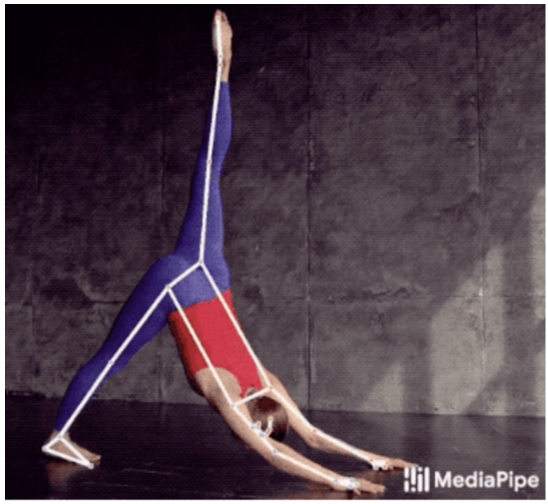


Fig. 1. Pose tracking in MediaPipe can pick out even the most unusual poses contained in dance or yoga images (© Google 2023)

The technical challenge was to be able to make a responsive installation that can jump to any moment in a dance and song scene in the films in an instant. To do that we had to index all the film information in a search-efficient manner. Given we only wanted to use the song and dance numbers from musicals, we only kept those sequences from each film. A typical musical had about 12 musical numbers averaging a total of 50 minutes of song and dance. Each musical number was stored in a separate movie file and the total was approximately 600 musical numbers (2,500 minutes). To optimise seeking in an H.264 format, we processed these 600 video files to generate two I-frames per second.

Having generated separate movie files for each musical number we extracted skeletal information from each scene. To do that, we used the state-of-the-art MediaPipe Pose [Memo Akten 2009] library which uses machine learning to achieve high-fidelity skeletal tracking, extracting 33 landmarks from images. MediaPipe uses Google Research's BlazePose which allows real-time performance on a standard laptop. The main reason we chose this technology compared to PoseNet used in other projects is because it is particularly good in unusual poses such as those found in yoga, dance or fitness videos as seen in Figure 1. Its robustness and ability to detect unusual poses and bodies makes also for more inclusive technology, as it allows for people with disabilities to participate in new media installations.

After processing the video files, we generated a json file for each musical number which contained all the poses found. Each pose entry in the file contained the 33 landmarks describing the parts of the body of an actor in a scene. It also contained the frame number where the pose was detected, the filename of the image containing the extracted frame and the filename of the musical film it belonged to.

To effectively search the pose-space we developed in python a similar pipeline to the one proposed in [Exceptions to copyright n.d.]. We first turned the 2D coordinates of the 33 landmarks into a 1D vector. To account for differences in image and skeletal size we normalised this pose vector using L2 normalisation. Having normalised the vectors, we calculated the similarity between them using the absolute difference between each element of the vectors. However, not all landmarks were treated equally. In the movie clips, joints were often occluded by other body parts or objects. This

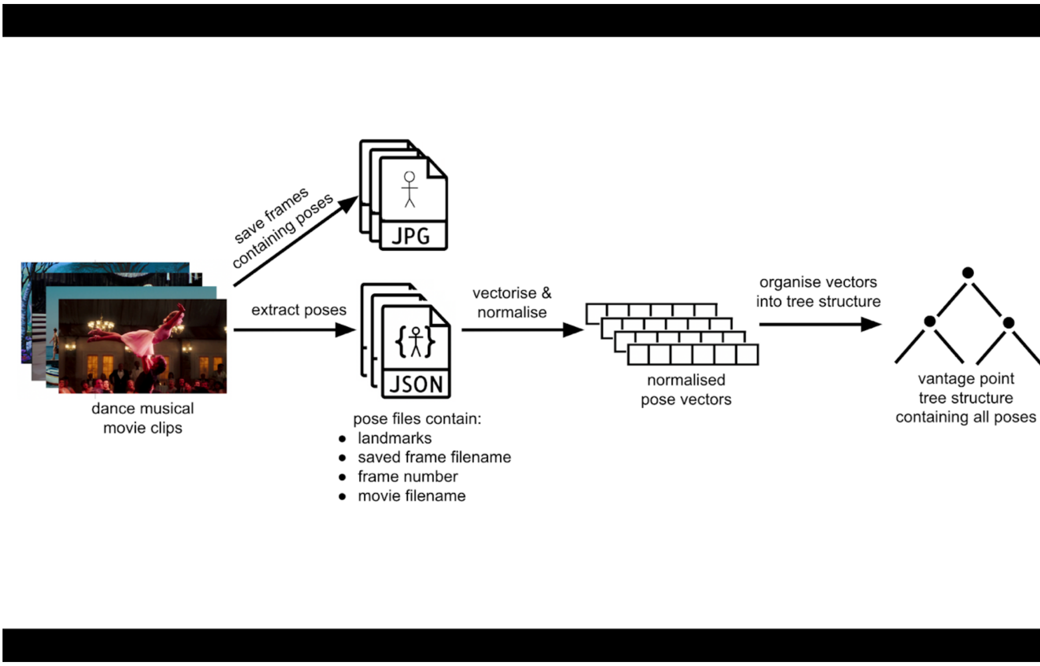


Fig. 2. A simplified outline of the pipeline to build the dataset of poses contained in musical movie files (© Papatheodorou 2023)

is where the confidence score returned by the MediaPipe pose algorithm was vital. We multiplied each absolute difference between landmarks by the confidence score, giving less importance to landmarks where the confidence score was low and higher where the score was high.

If we performed brute force search on the entire corpus of tens of thousands of poses, it would have been too slow to be appropriate for a real-time installation. To find the closest matching pose efficiently we used a vantage point tree search algorithm. This made each search take ~5ms. A simplified pipeline is shown in Figure 2.

## 4 PARTICIPANT INTERACTION

At an early stage of development, we decided there would be no instructions for use. This forced us to make design decisions that encourage an embodied interaction that is intuitive. The prototype installation was set up in a small movie theatre. We used two cinema lights to light the participants (GVM 800D-RGB LED Light Panel), a colour camera (Logitech C920 HD Pro Webcam) and MacBook Pro laptop from 2020 (2.3GHz quad-core Intel i7, 16GB RAM, Intel Iris Plus GPU). The theatre had a 4K projector and a large 5m x 3.5m projection surface. A square on the floor indicated to participants where they should stand.

As the participants move, their skeleton data is extracted, vectorised and L2-normalised. It is then submitted to the pose search engine to find the closest match. The pose search engine uses the vantage point tree generated in the previous step to quickly traverse the pose space to find the closest match. The result is sent as an OSC message to an image and video player written by the authors in openFrameworks. The OSC message contains the file name of the image containing the matching pose, the filename of the musical video from which it was extracted and the frame

Fig. 3. A participant moves her body, and her pose is matched to actors in musicals in real-time. The installation can match even challenging poses (bottom images) of the participant to similar poses in the database in real-time (© Papatheodorou 2023).

number where the matching frame was found in the video stream. These three data points are all we need for the installation's front-end.

As the participants move their body, the front-end of the installation displays frames that were extracted from the musicals, where the pose of the actors matches the participant's. This happens in quick succession, creating a flip-book-like effect. For continuity and a smoother viewing experience the matching film character is cropped and translated to the centre of the screen. The overall setup and the pose-matching is shown in action in Figure 3.

This flip-book-like experience continues while the participant is moving. As soon as the participant stops, the matched, cropped image showing the actor moves into the slot it occupied in the original film frame and its boundaries grow to become the full frame. As soon as it reaches the full frame size, a ten second "phrase" from the corresponding scene in the musical plays in a loop. If the participant moves again, the film excerpt is instantly interrupted and new images of actors in matching poses are shown on the screen.

To increase the diversity of the images shown, we created a *frame diversifier* filter, which keeps a log of recently shown matching frames and if one has been shown in the last 20 seconds, the next best match is chosen. This resulted in many more unique frames being shown during the running

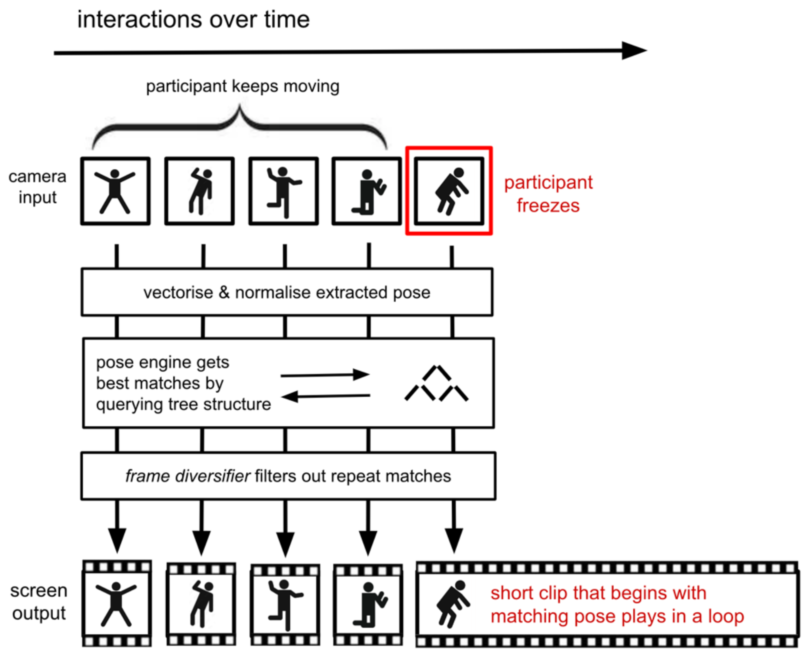


Fig. 4. A simplified overview of the parts involved in the running of the installation (© Papatheodorou 2023).

of the installation, maintaining the novelty and interest. A simplified diagram of the interaction is shown in Figure 4.

## 5 EARLY USER FEEDBACK

The prototype installation was visited by a test cohort of 30 participants to get early feedback before sending it to its final destination. In response to this we received a variety of responses that helped us reassess the scope and possibilities of the work.

Video artist and curator Phoenix Fry said, "Seeing the crowd around the installation I was intrigued about how it activates embodied memory." He was also interested in the potential for collaborative play between participants. We too noticed that often people stayed around the installation, giving instructions to the participants.

Two museum curators raised issues about copyright law and how it could affect such work. We consulted a copyright lawyer who advised us that this installation is covered by the *parody, caricature and pastiche* clause of UK and EU copyright law [Zuckerman and Gal-Oz 2013].

Dancer, choreographer and university lecturer Irini Kalaitzidi was interested in indexing her dance poses and use the output of a modified pipeline as a teaching tool. She was also interested in comparing dance moves quantitatively across the ages. How has musical dance evolved over the years?

Seven people told us they actively chose difficult, uncomfortable poses, because they wanted to challenge the pose tracker. Six participants told us they were shy at first, but after playing for a minute they forgot people were watching. Another six told us they enjoyed being represented on the screen by dancers of the opposite gender.

People used the installation for an average of five minutes.

Fig. 5. Skeleton tracking of statues to match them in real time to poses of participants (© V&A Museum, London 2023)

## 6 DISCUSSION AND FUTURE WORK

The matches that our pipeline returned were always of very high similarity to the pose the participants had, even when these poses were challenging, such as when the participants were kneeling, lying on the floor etc., as shown in the bottom two images in Figure 3, something which delighted the audience.

For an intuitive user experience, low latency was key. The architecture allowed us to search in real-time more than 600 video files, 2,500 minutes of film without perceptible delay. Moreover, the corresponding video file played back immediately when the actor froze in a specific pose. Our architecture allows the system to take many more films without any noticeable drop in performance.

From the test cohort of users, we observed that the vast majority needed no instructions and understood intuitively, how to engage with it. Once it is installed in its final destination with a lot of participants, we would like to document, more formally and multidimensionally, the interactive experience. We would like to analyse how this work challenges traditional gender roles within the context of stage performance and musical films. We would also like to record the choices of poses and how they may reflect the participants' understanding of gender in dance and performance. Can such installations offer opportunities for participants to explore gender expression in a novel manner? Furthermore, we would like to collect data about whether this embodied interaction might enhance the participants' awareness of their own performativity and contribute to their understanding of the performative nature of dance and musical films. How much do these tangible interfaces enhance the understanding of dance and movement as embodied forms of cultural expression?

On a technical level, we would like to research a temporal search, using "gestural phrases" rather than individual frame-poses. Recent work [Price et al. 2003] explores a dance music retrieval system based on body-motion similarity with promising results.

Using the techniques proposed in this paper opens new pathways for cultural participation. In collaboration with a cultural institution in London we are working on matching visitors' poses with poses of sculptures in their collection. Initial tests of skeleton tracking on statue images were promising as seen on Figure 5. This opens exciting new possibilities for an embodied and interactive experience with more traditional forms of art.

## REFERENCES


Alissa N. Antle, Allen Bevans, Theresa Jean Tanenbaum, Katie Seaborn and Sijie Wang. 2010. Futura: design for collaborative learning and game play on a multi-touch digital tabletop. In Proceedings of the fifth international conference on Tangible, embedded and embodied interaction (TEI '11). Association for Computing Machinery, New York, NY, USA, 93–100.

Pina Bausch. n.d.. Official website, Retrieved April 11, 2023 from https://www.pinabausch.org/

Siobhan Davies et. al. 2008. RePlay. Retrieved January 17, 2023 from http://www.siobhandaviesreplay.com/

Exceptions to copyright. n.d.. UK government website, Retrieved April 11, 2023 from https://www.gov.uk/guidance/exceptions-to-copyright

Jane Friedhoff and Irene Alvarado. 2018. Move Mirror. Retrieved January 17, 2023 from https://experiments.withgoogle.com/move-mirror

Benedikt Gross. 2016. Alphabet From The Sky. Penguin Random House

Zach Lieberman. 2016. Land Lines. Retrieved January 17, 2023 from https://experiments.withgoogle.com/land-lines

Camillo Lugaresi, Jiuqiang Tang, Hadon Nash, Chris McClanahan, Esha Uboweja, Michael Hays, Fan Zhang, Chuo-Ling Chang, Ming Yong, Juhyun Lee, Wan-Teh Chang, Wei Hua, Manfred Georg, Matthias Grundmann. 2019. MediaPipe: A Framework for Perceiving and Processing Reality.Third Workshop on Computer Vision for AR/VR at IEEE Computer Vision and Pattern Recognition

Leilah Lyons, Brian Slattery, Priscilla Jimenez, Brenda Lopez and Tom Moher. 2012. Don't forget about the sweat: effortful embodied interaction in support of learning. In Proceedings of the Sixth International Conference on Tangible, Embedded and Embodied Interaction

Christian Marclay. 2010. The Clock. Retrieved January 17, 2023 from https://www.tate.org.uk/art/artworks/marclay-the-clock-t14038

Memo Akten. 2009. Body Paint, Retrieved April 11, 2023 from https://www.memo.tv/works/bodypaint/

S Price, Y Rogers, M Scaife, D Stanton, H Neale. 2003. Using 'tangibles' to promote novel forms of playful learning, Interacting with Computers, Volume 15, Issue 2, Pages 169-185,

Jennifer G. Sheridan, Nick Bryan-Kinns. 2008. Designing for performative tangible interaction. Int. J. Arts and Technology, Vol. 1, Nos. 3/4

Tsuchida, Shuhei & Fukayama, Satoru & Goto, Masataka. 2019. Query-by-Dancing: A Dance Music Retrieval System Based on Body-Motion Similarity. 10.1007/978-3-030-05710-7_21

Gene Youngblood. 1970. Expanded Cinema. P. Dutton & Co., Inc.

Oren Zuckerman, Ayelet Gal-Oz. 2013. To TUI or not to TUI: Evaluating performance and preference in tangible vs. graphical user interfaces. Internat. J. Human-Computer Studies, Vol 71, 803-820