



**HAL**  
open science

## **Kinetic particles: from human pose estimation to an immersive and interactive piece of art questioning thought-movement relationships.**

Mickael Lafontaine, Julie Cloarec-Michaud, Kevin Riou, Yujie Huang, Kaiwen Dong, Patrick Le Callet

### ► To cite this version:

Mickael Lafontaine, Julie Cloarec-Michaud, Kevin Riou, Yujie Huang, Kaiwen Dong, et al.. Kinetic particles: from human pose estimation to an immersive and interactive piece of art questioning thought-movement relationships.. IMX '23: ACM International Conference on Interactive Media Experiences, Jun 2023, Nantes France, France. pp.382-385, 10.1145/3573381.3597228 . hal-04265659

**HAL Id: hal-04265659**

**<https://hal.science/hal-04265659v1>**

Submitted on 31 Oct 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Kinetic particles : from human pose estimation to an immersive and interactive piece of art questioning thought-movement relationships.

Mickael Lafontaine\*  
mickael.lafontaine@gmail.com  
New Media Artist  
Nantes, France

Julie. Cloarec Michaud\*  
julie.cloarecmichaud@gmail.com  
Intermittent artist  
Nantes, France

Kevin Riou  
Nantes Université, École Centrale  
Nantes, CNRS, LS2N, UMR 6004  
Nantes, France  
kevin.riou@univ-nantes.fr

Yujie Huang  
Nantes Université, CAPACITÉS SAS  
Nantes, France  
yujie.huang@etu.univ-nantes.fr

Kaiwen Dong  
China University of Mining and  
Technology  
Xuzhou, China  
dongkaiwen@cumt.edu.cn

Patrick Le Callet  
Nantes Université, École Centrale  
Nantes, CNRS, LS2N, UMR 6004  
Nantes, France  
patrick.lecallet@univ-nantes.fr



Figure 1: Preview of the proposed demonstration in use. On the right image, the "Particules" scenario illustrates how the human body movement is involved in the piece of art. On the left image, another scenario illustrates the immersive ability of the proposed installation.

## ABSTRACT

Digital tools offer extensive solutions to explore novel interactive-art paradigms, by relying on various sensors to create installations and performances where the human activity can be captured, analysed and used to generate visual and sound universes in real-time. Deep learning approaches, including human detection and human pose estimation, constitute ideal human-art interaction mediums, as they allow automatic human gesture analysis, which can be directly used to produce the interactive piece of art. In this context, this paper presents an interactive work of art that explores

\*Both authors contributed equally to this research.

Permission to make digital or hard copies of part or all 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 third-party components of this work must be honored. For all other uses, contact the owner/author(s).

IMX '23, June 12–15, 2023, Nantes, France

© 2023 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-0028-6/23/06.

<https://doi.org/10.1145/3573381.3597228>

the relationship between thought and movement by combining dance, philosophy, numerical arts, and deep learning. We present a novel system that combines a multi-camera setup to capture human movement, state-of-the-art human pose estimation models to automatically analyze this movement, and an immersive 180° projection system that projects a dynamic textual content that intuitively responds to the users' behaviors. The demonstration being proposed consists of two parts. Firstly, a professional dancer will utilize the proposed setup to deliver a conference-show. Secondly, the audience will be given the opportunity to experiment and discover the potential of the proposed setup, which has been transformed into an interactive installation. This allows multiple spectators to engage simultaneously with clusters of words and letters extracted from the conference text.

## CCS CONCEPTS

• Applied computing → Performing arts.

## KEYWORDS

Interactive Art, Digital Art, Immersive Content, Real-time Multi-Person Pose Estimation

### ACM Reference Format:

Mickael Lafontaine, Julie. Cloarec Michaud, Kevin Riou, Yujie Huang, Kaiwen Dong, and Patrick Le Callet. 2023. Kinetic particles : from human pose estimation to an immersive and interactive piece of art questioning thought-movement relationships.. In *ACM International Conference on Interactive Media Experiences (IMX '23)*, June 12–15, 2023, Nantes, France. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/3573381.3597228>

## 1 INTRODUCTION

An interactive work of art is described by Kluszczyński [10] as an event, where an artist doesn't create a final, complete piece of art, but instead produces an area of activity, where the participative behavior of the audience will bring to life the artwork-event. Kluszczyński [10] categorizes the audience as participants, performers, executors, or (co)creators of an artwork-event, depending on the interactive art strategy considered, and further details eight interactive art strategies : the strategy of instrument, game, archives, labyrinth, rhizome, system, network, and spectacle. In this work, we focus on the strategy of instrument, where an interface suggests the audience to create a performance by generating visual [9] or audio-visual events [12, 13], based on its behaviors. Digital tools offer extensive ways to explore strategies of instrument, by relying on various sensors to create installations and performances where the activity of the body generates visual and sound universes in real-time [2, 7, 11]. Deep learning approaches, including human detection[3] and human pose estimation[4, 6, 8], constitute ideal human-art interaction medium, as they allow automatic human gesture analysis, which can be directly used to produce the interactive piece of art. The interactivity offered by such solutions allows for a sensory and poetic augmentation of the gestures of the performer or spectator. Rhizomatiks [7] especially showed that coupling such solutions with video projection systems has considerable and innovative interest in the context of a creative conference during which the speaker literally manipulates their concepts and research universe. The video projection system is an important factor that allows the performer or spectators to be immersed, in the literal sense of the term, in these visual universes generated in real-time. With a similar state of mind, we propose a demonstration that combines dance, philosophy, numerical arts and deep learning, and which explores the relationship between thought and movement. Concretely, we propose a multi-camera setup coupled with deep learning approaches that allows to map the movements of several users in the form of a words ballet, that are thereafter projected on the walls surrounding the users. The demonstration is decomposed in two parts, 1) a professional dancer uses the proposed setup to give a conference-show, and 2), the audience is invited to try and explore possibilities of the proposed setup, which is converted to an interactive installation, where several spectators can simultaneously interact with clouds of words and letters taken from the conference text. Note that the number of cameras in our setup can scale with the number of participants, and/or with the experiment area size, as our software is agnostic to the number of cameras used. Visitors can also add their own text to demonstrate how other

researchers can easily re-appropriate this device. Our goal is to present this installation in the 180-degree video projection setup of the XP Lab in order to create a completely immersive experience.

## 2 APPROACH

### 2.1 The artistic project

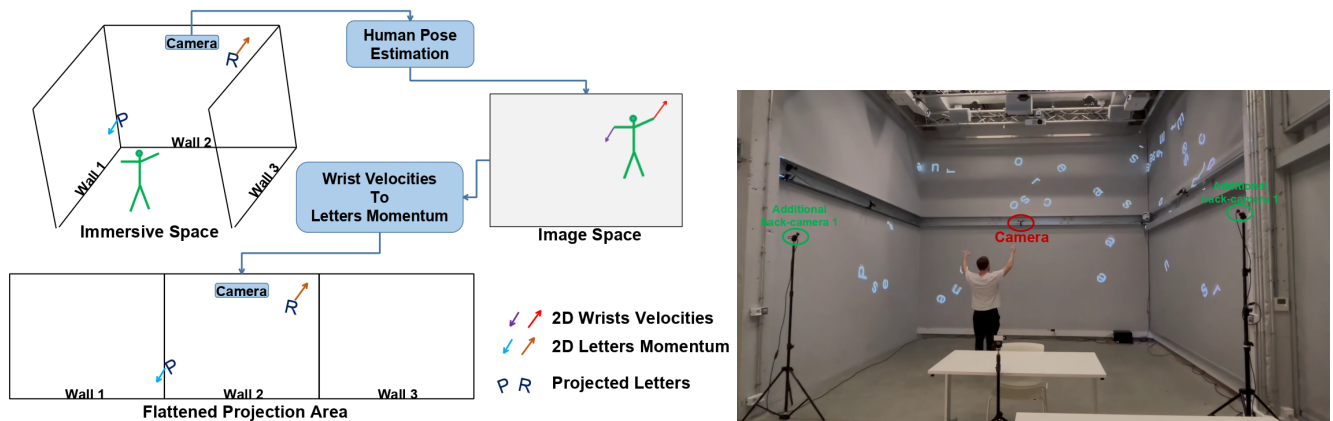
This work aims at finding connections between thought and movement. It consists of a conference-show combined with an interactive installation that allows the spectator to actively participate in the process and personally and physically experience this linking between speech and action. The user is involved in an immersive universe made up of texts that comes to life according to their movements. The authors of the work give the words a real materiality and autonomy of movement, so that the interaction between them and the users is both intuitive, organic, lively, and choreographic. This artistic performance, both hybrid and transdisciplinary, in the very scope of the lecture and the inherent artistic creation, represents the immutable link between thought and movement, philosophy and dance, words and gestures.

*2.1.1 The conference-show time.* During the conference, a performer or a duo of performers interact with an immersive textual universe in which they can bring to life the textual content projected on the walls surrounding them. For single performer shows, the performer's movement is typically captured by one 3D camera placed in front of the performer. The human movement is analysed by deep learning approaches, detailed in section 2.2, and propagated to a collection of letters projected around the performer. Fig.1 illustrates how the letters can be projected in the immersive space. The letters can be structured in different ways, and several behaviors can be attached to them. For instance, in the scenario that will be displayed during the demonstration, the letters can be considered as independent particles controlled by the performer, which interact with each other, with the wall boundaries and are subject to a virtual gravity. These behaviors allow intuitive and choreographic interaction with the performer. In the duo-performer setup, the movement of the two performers is captured by one or more 3D cameras and each user controls different sets of letters, which provides an additional interactive dimension: the performers can interact with each other by the mean of the artistic projection.

*2.1.2 Interactive and collaborative installation time.* The idea to combine a whole installation offered to the audience with the show-lecture appeared immediately during the creation of the latter with a view of permitting the public to share the live experience. In this installation, several spectators are involved in an immersive video projection where they can simultaneously interact with the textual universes presented during the conference. This installation invites spectators to collaborate both intellectually and choreographically: searching for words, exchanging words, synchronizing movements, etc. This collective participation is undoubtedly innovative given the classical structures offered both in the field of human sciences and the performing arts.

### 2.2 Technical details

The section describes the technical details for the "Particules" scenario, which will run during the conference demonstration. The



**Figure 2: Illustration of the proposed setup.** On the left, the scheme details the pipeline associated to the "Particules" scenario described in section 2.2.1, simplified with 2 projected letters. On the right is a visualization of the real setup, where the circled "camera" corresponds to the camera on the left scheme. Additional cameras, here denoted as "back-cameras" can be added to the setup to tackle occlusions that may occur in multi-user scenarios.

single user setup, associated to the conference-show time is first described. The multi-user setup, associated to the interactive and collaborative installation time is described afterwards.

**2.2.1 Single User Scenario.** The proposed interactive setup that will be displayed during the conference is illustrated in Fig.2. It is composed by (1) a Monocular Human Pose Estimation System, which is a solution to analyse human movement and (2) the "Wrist Velocities to Letter Momentum" module, which is a software that animates textual projections, based on the human movement analysis. Monocular Human Pose Estimation is a long-standing problem [4–6] in computer-vision community, which aims to reconstruct the 2D locations of human body joints in an input image or video. Human Pose Estimation models provide valuable information on the position of the different joints of the user's body, which can be further used to recover the user's joints velocities from adjacent frames. The 2D velocities of chosen target joints can further be propagated to the textual content projected around the user. For instance, in the "Particules" scenario, which will be presented during the conference demonstration, the letters of the projected text are considered as independent particles, subject to gravity. By default, the particles are motionless and grounded by this artificial gravity. When a non-zero velocity is detected on one of the user's wrists, a force is applied to the letters, in the direction of the detected velocity. In this "Particules" scenario, both wrists control a respective half of the projected letters. For the demonstration, the ZED M cameras from stereolabs are used, and the "BODY\_18" integrated solution for 2D body tracking is used as Monocular Human Pose Estimation System.

**2.2.2 Multi-User Scenario.** The Multi-User Scenario works similarly to the Single User Scenario, except that independent sets of letters are assigned to the different users. To ensure a temporal smoothness of the movements, the different users are tracked with a unique id to ensure that a given set of letters is always assigned to the same user. In the case of massive amount of users, additional

cameras can be added to our setup, to tackle inter-users occlusions. If additional cameras are needed, they are calibrated together using EasyMocap[1] calibration tools. The skeletons obtained from the different views are aligned by projecting them in a shared 3D world space using extrinsic camera parameters obtained during the calibration. For each user in the interaction area, the skeleton with the highest prediction confidence is kept, projected to the front-camera view, and sent to the "Wrist Velocities to Letter Momentum" module.

### 3 CONCLUSION AND PERSPECTIVES

This project, through its multidisciplinary, ranging from philosophy to digital research, including arts and dance, makes it an excellent field for both theoretical and practical experiences. With this project, we hope to bring reflection on our relationship with thought and movement, with the goal of not separating the two in order to better analyze them, but on the contrary, embracing them in their common movement in order to better grasp their own momentum. As future works on the technical part, we wish (1), to migrate the human-pose estimation models used to a multi-view 3D pose estimation approach, to avoid prediction inconsistencies between views and (2), to invest in another area of research that could greatly improve both the immersive and the interactive abilities of the setup. Our goal is create a smart video mapping system that can calculate and adapt the position of projected animations in the 3D space, which would allow projecting content on target objects positioned in the space, including the users' bodies.

## REFERENCES

- [1] 2021. EasyMoCap - Make human motion capture easier. Github. <https://github.com/zju3dv/EasyMocap>
- [2] M Adrien and B Claire. 2013. Hakanai.
- [3] Hong-Yuan Mark Liao Alexey Bochkovskiy, Chien-Yao Wang. 2020. YOLOv4: YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* (2020).
- [4] Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. 2017. Realtime multi-person 2d pose estimation using part affinity fields. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 7291–7299.
- [5] Yucheng Chen, Yingli Tian, and Mingyi He. 2020. Monocular human pose estimation: A survey of deep learning-based methods. *Computer Vision and Image Understanding* 192 (2020), 102897.
- [6] Bowen Cheng, Bin Xiao, Jingdong Wang, Honghui Shi, Thomas S. Huang, and Lei Zhang. 2020. HigherHRNet: Scale-Aware Representation Learning for Bottom-Up Human Pose Estimation. In *CVPR*.
- [7] Yuya Hanai, Kyle McDonald, Satoshi Horii, Futa Kera, Kisaku Tanaka, Motoi Ishibashi, and Daito Manabe. 2021. Fencing tracking and visualization system. In *SIGGRAPH Asia 2021 Real-Time Live!* 1–1.
- [8] Tao Jiang, Peng Lu, Li Zhang, Ningsheng Ma, Rui Han, Chengqi Lyu, Yining Li, and Kai Chen. 2023. RTMPose: Real-Time Multi-Person Pose Estimation based on MMPose. *arXiv preprint arXiv:2303.07399* (2023).
- [9] Jusub Kim. 2014. Choreographing Digital Water for Interactive Water Performance. In *Intelligent Technologies for Interactive Entertainment: 6th International Conference, INTETAIN 2014, Chicago, IL, USA, July 9-11, 2014. Proceedings* 6. Springer, 36–45.
- [10] Ryszard W Kluszczyński. 2010. Strategies of interactive art. *Journal of Aesthetics & Culture* 2, 1 (2010), 5525.
- [11] Daito Manabe. 2015. Behind the scenes. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 2–2.
- [12] David Rokeby. 1995. Very Nervous System (VNS).
- [13] Paul Sermon. 2000. Telematic dreaming. *Leonardo* 33, 2 (2000), 90–90.