This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 899739

CROWD DNA

H2020 EU Fet-Open Project

Technologies for computer-assisted crowd management

www.crowddna.eu

Call: H2020-FETOPEN-2018-2019-2020-01

Type of action: RIA

Grant agreement: 899739

WP N°3:	**Macro-to-micro crowd motion analysis**
Deliverable Nº3.2:	**ML Detector to analyse crowd movements in videos**
Task lead:	**UCL**
WP lead:	**URJC**
Version Nº:	**1.0**
Date:	**22/12/2024**

Document information	
Deliverable N° and title:	D3.2 – ML Detector to analyse crowd movements in videos
Version N°:	0.4
Task lead:	UCL
WP lead:	URJC
Author(s):	He WANG (UCL/UL)
Reviewers:	Julien PETTRE (Inria), Solenne FORTUN (Inria)
Submission date:	22/12/2023
Due date:	31/12/2023
Type:	DEM
Dissemination level:	PU

Document history			
Date	Version	Author(s)	Comments
27/11/2023	0.1	He Wang (UCL/UL)	Template creation
27/11/2023	0.2	He Wang (UCL/UL)	First draft of Executive Summary, Section 1 and Conclusion
15/12/2023	0.3	He Wang (UCL/UL)	Finalising the report and revision based on Solenne's feedback
18/12/2023	0.4	He Wang (UCL/UL)	Revision based on Juliane's feedback
22/12/2023	0.5	He Wang (UCL/UL)	Revision based on Julien's feedback, adding more technical details in 1.3 and some new results in 1.4.1
22/12/2023	0.6	He Wang (UCL/UL)	Adding a video link in Evacuation Drill
22/12/2023	1.0	Solenne Fortun (Inria)	Final layout

Table of Contents

Table of Figures

Acronyms and Abbreviation

CDI	Crowd Dynamics International Limited
EC	European Commission
EMT	Executive Management team
FZJ	Forschungszentrum Julich Gmbh
GA	Grant Agreement
INRIA	Institut National De Recherche En Informatique Et Automatique
KPIs	Key Performance Indicators
ONH	Onhys
PO	Project Officer
UL	University of Leeds
ULM	Universität Ulm
URJC	Universidad Rey Juan Carlos
WP	Work-package

Executive Summary

This deliverable describes the efforts done during Period 1, Period 2 and part of Period 3 in the Work Package 3 of the CrowdDNA project towards developing a new crowd simulator algorithm tailored to model both macro and micro-level crowd characteristics. As a reminder, the overall objective of WP3 is to deliver a new method for crowd motion analysis that is capable of estimating the intensity of physical interactions between individuals from the observation of macroscopic crowd motion features. Our approach is a Machine Learning one. We have two competitive methods in mind to perform such mapping: i) training a generator that is based on a set of features we pre-determined, or ii) train a detector that would identify itself the most relevant features to extract from the crowd motion. Training will be based on the datasets generated in WP1. The second goal, i.e. the detector is the main goal of D3.2, which is to develop a detector to analyse crowd movements in videos.

To this end, two partners UCL/UL have developed relevant new algorithms and models to detect crowd behaviours from videos. The approach focuses on fine-grained high-density crowd behaviours, based on vastly different data types and qualities, under common in-the-wild data collection settings, to best accommodate real-world high-density crowd data. UCL/UL have focused on videos of high-density crowds where each person occupies merely a few pixels and the data is captured by far-distance cameras so the data contains excessive noise.

Figure 1 depicts the different components of the proposed detection methods, next to the CrowdDNA partner who led each of the developments.

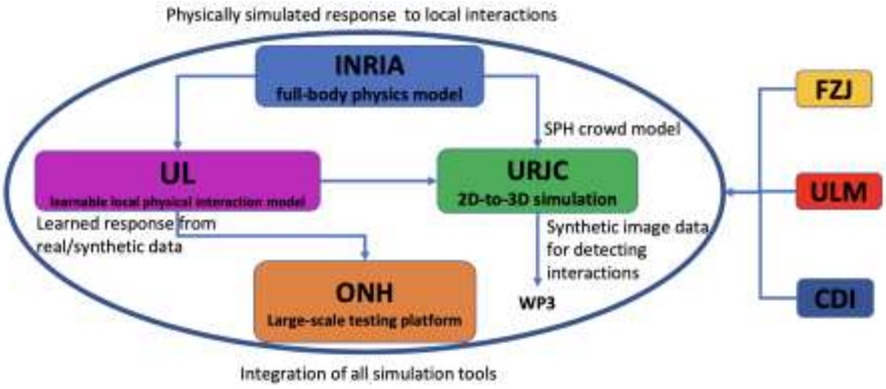

Figure 1. Components of CrowdDNA simulator.

All in all, due to the complex nature of high-density crowd behaviours, to detect crowd behaviours and accommodate different settings of the crowd observatories, we have explored data of various qualities, from where it is still possible to observe full/partial bodies of each individual, to where it is impossible to identify detailed individual body motions.

1. Crowd Behaviour Detection on Noisy Videos (UCL/UL)

1.1. Background

UL and UCL (Feixiang He, He Wang) have led the work in developing a new method that can estimate the low-level individual interactions from high-level video data, T3.2 in WP3. Feixiang He is a PhD student of He Wang who is funded by other sources instead of CrowdDNA, so no expense has been claimed for the PhD student. This is because He Wang leveraged an external PhD funding for CrowdDNA and formed the PhD project to be jointly conducted with T3.2 for mutual benefits.

1.2. Motivation and Challenges

The goal of this work is to be able to analyse, detect, predict and even simulate high-density crowds in a relatively confined space, such as festivals, religious venues, concerts, etc. This is exactly the type of crowds with which CrowdDNA is particularly concerned. Analysing/simulating/predicting crowd movements in videos have been a long-standing problem in computer vision. Deviating from existing research, we aim to analyse/simulate/predict high-density crowd movements where we are particularly interested in density that is at least 5p/m^2 (people per square meter) where life threatening crushes can start to develop.

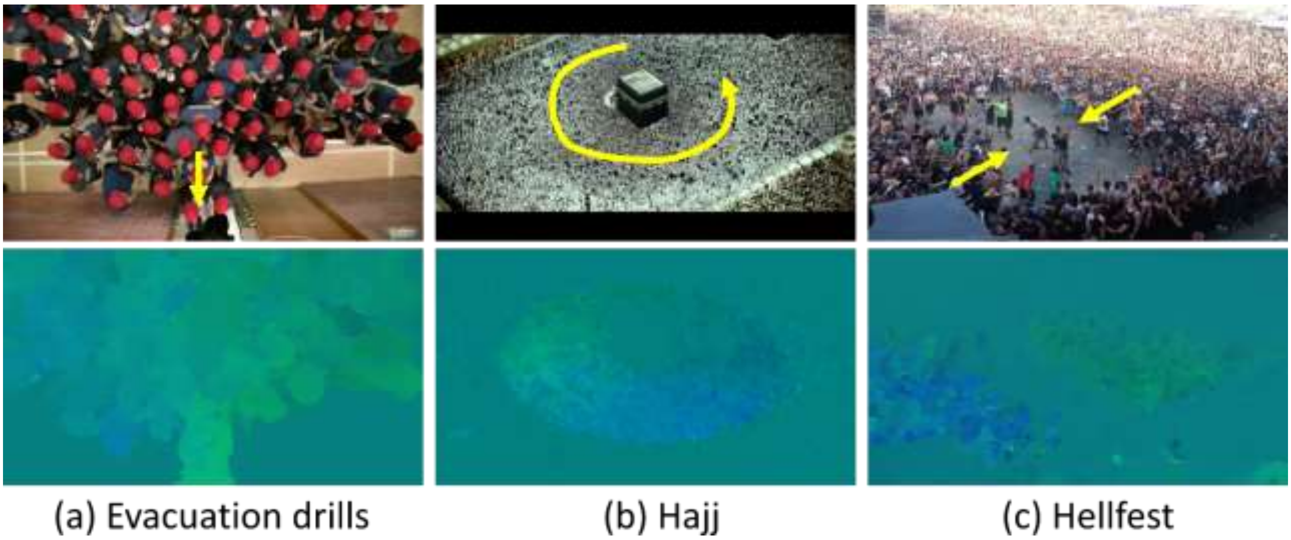

Figure 2. Figure 1. Components of CrowdDNA simulator.High-density crowd videos. Left: Evacuation Drill (Angel Garcimartin, 2016), Middle: Hajj[1], Right: Hellfest[2]. The first was an experiment in lab recorded by a top-down view camera, while the other two are CCTV videos. The bottom row is the optical flows extracted from the videos.

At this density level, analysis/simulation/prediction becomes extremely challenging as the dynamics can become chaotic. At the same time, as shown in Figure 2 with three typical high-density crowds, the data that is normally available is highly noisy, non-informative (e.g. impossible to track individuals in general, let along full-body individual motions). The data are from (e.g. the Hajj and Hellfest) far distant CCTV cameras, which is a typical situation for data collection in the wild. So the only information we can rely on is the noisy optical flows. Other information such as lighting, texture, individual trajectories, etc. cannot be reliably obtained easily. Also there is no extrinsic calibration and the tilted view makes the situation even worse. Without any extrinsic calibration one could only derive a density in the 2D pixel map which is not based on any real world coordinates.

[1] *https://www.youtube.com/watch?v=KGukAoiGhZU&t=5s*

[2] *https://www.youtube.com/watch?v=ySPlanMCmM4*

Given the data, unfortunately, existing methods cannot be used in our task. On the high level, existing approaches can be categorized into empirical modelling and data-driven methods. Empirical models abstract crowd dynamics into rule-based and deterministic systems. They can simulate high-density crowds via prescribed behaviours, but cannot learn the behaviours of a specific crowd. Meanwhile, data-driven simulation/prediction can learn crowd-specific behaviours, but are not interpretable as they are black boxes (hence no good for analysis) and require large amounts of labelled trajectories (which is absent).

1.3. Our solution

We propose a new machine learning method that can take as input the highly noisy optical flows then learn to predict the crowd movements. The assumption is, although the optical flows are noisy, they do somewhat reflect the underlying high-density crowd movements. So it is possible to infer the individual interactions from the optical flows, at least locally. During learning, we aim to learn the intrinsic behaviours of people within densely packed crowds so that we can infer the intensity of interactions between people, estimate global as well as local movements, predict their motions and even simulate similar behaviours in new environments.

1.3.1. Seeing a High-density Crowd as a (Complex) Continumm

The novel model we propose is a new differentiable physics model that learns a `crowd material', which is a new continuum model that is tailored to capture extremely high-density crowds (> 5p/m^2). It is a model based on Material Point Method (MPM) so we call the model CrowdMPM. In this model, each person is treated as a mass particle in a heterogeneous continumm. However, we might not actually observe the particle movement directly. So we design the model to take either raw low-quality optical flow or labelled/tracked trajectories as input. The model training aims to learn key material parameters associated with the resulting movements of the particles, observed in optical flows or trajectories. After training, it can be used as a predictor for crowd motions for a few seconds, or a simulator to mimic crowd behaviours in different environments. In the following sections, we show the results of prediction and simulation using our learned model. In prediction, we first feed into our model a few seconds data, then let the model predict the next period of time. In simulation, we only set up the environment, the initial positions and the destinations of agents and let the model simulate onward in time. The overall framework is shown in Figure 3.

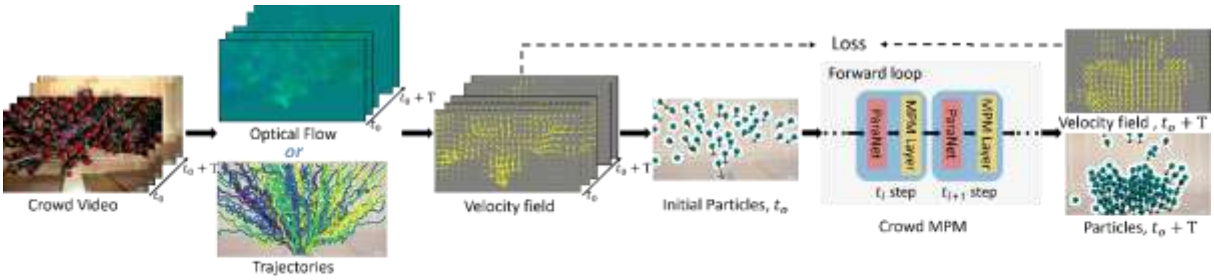

Figure 3. Crowd MPM overview. From left to right: optical flow extraction, velocity field generation, initial particle sampling, crowd simulation and loss calculation

Specifically, we see a high-density crowd as a complex continumm under purely mechanical loading (no heat exchange). Its motion is governed by the conservation of mass and the conservation of momentum, so the system state evolvement is represented by its velocity field evolvement in time (Figure 3).

1.3.2. Special Material Properties of Crowd Material

Treating crowds as a continuum, the key is to model how `crowd material' behaves, i.e. the strain-stress relation. We notice some major differences between crowds and common materials. First, crowd can easily scatter but cannot be compressed severely, due to people cannot be superpositioned, i.e. easily stretched but not as easily compressed, which we refer to as *elastic asymmetry*. Further, during compression, there is a multi-phasic resistance change based on the inter-personal distance: little resistance when the distance is larger than a comfortable threshold, but exponentially increased resistance within the threshold until incompressibility (e.g. people in full contact). We refer to this property as *exponential resistance*. Finally, although people resist compression, they are more tolerant to relative motions such as passing each other at a close distance. This suggests `crowd material' has less resistance in deformation associated with shearing and rotations, which we refer to as *compression dominance*. To model the above material features, we treat each individual as a Lagrangian particle. Unlike particles in standard MPM as collocation points and not corresponding to any entities, in Crowd MPM, they are entities, so the strain-stress between them needs to be modelled between particles directly, not via the grid nodes as in standard MPM explained. This is our key technical novelty involving a new stress modelling.

Standard MPM discretizes the domain by a regular Eulerian grid where Lagrangian particles are used to trace information such as mass and velocity. MPM takes three steps: (1) Particle-to-grid transfer (P2G) (2) Grid Operation (GO) and (3) Grid-to-particle transfer (G2P), then finally (4) update the particles. As shown in Figure 4 (a-d), first, P2G transfers mass, velocity and pressure from the particles onto the grid nodes, then GO solves the momentum equation under boundary conditions on the grid nodes, and finally the updated velocity and deformation are transferred back to particles from the grid nodes.

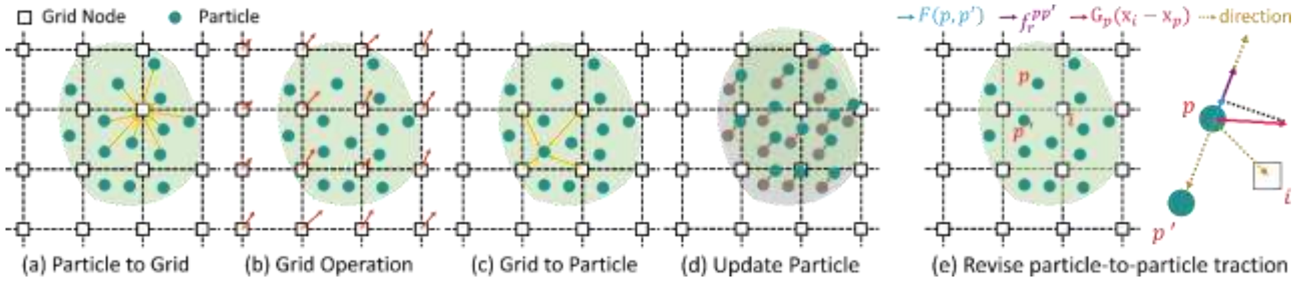

Figure 4. Left: MPM simulation. Right: particle-to-particle traction modeling

To model the elastic asymmetry, we use a Cauchy stress for weakly compressible fluid. Next, to model the exponential resistance, we need to explicitly model the repulsive forces between two close agents. However, In the MPM framework, this force interaction is solved on the grid, not directly on particles. Therefore, to explicitly introduce the interactions between individuals, we first isolate the traction between a pair of particles out (Figure 4 e), then model it separately via a neural network (Figure 3 Crowd MPM). The input of the network is a parameter indicates the size of the comfort zone of a person, and its nearby neighbour position and velocities. The output of this network is a scalar that controls the magnitude of the repulsive forces between these a pair of people. This way, via the training, Crowd MPM learns a heterogeneous material where the main heterogeneity comes from the low-level interactions between people.

1.4. Results

1.4.1. Learned Crowd Material vs Commonly-seen material

We first show the differences between our crowd material and standard material in Figure 5. We give particles an initial velocity to the right. Figure 5 Left shows the stress map. No matter if the stiffness is low or high in standard MPM (Figure 5 Left 1 and 2) where the young's modulus is 1 and 100, it does not model the exponential resistance as our model (Figure 5 Left 3). The overall stress increases steeply when approaching to the right end where the inter-person distance becomes smaller. This indicates even exhaustive hand-tuning MPM will not mimic the crowd material as it does not exist in its parameter space. Next, Figure 5 Right shows the compression dominance. Figure 5 Right 1 and Figure 5 Right 2 are either too soft to resist compression, or resist compression and also shear and rotations. Comparatively, our model (Figure 5 Right 3) resists compression but still allows shearing/rotation as highlighted in the red box. Again, this is evidence that the crowd material does not exist in the parameter space of standard MPM.

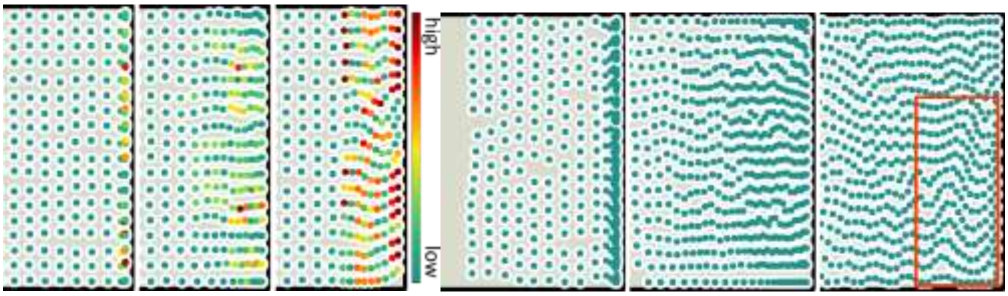

Figure 5. Crowd moving to the right with an initial velocity. Left: stress colormap at t = 0.5s. Right: Deformation at t = 1.8s. In each group, 1: small stiffness in standard MPM. 2:large stiffness in standard MPM. 3: Ours.

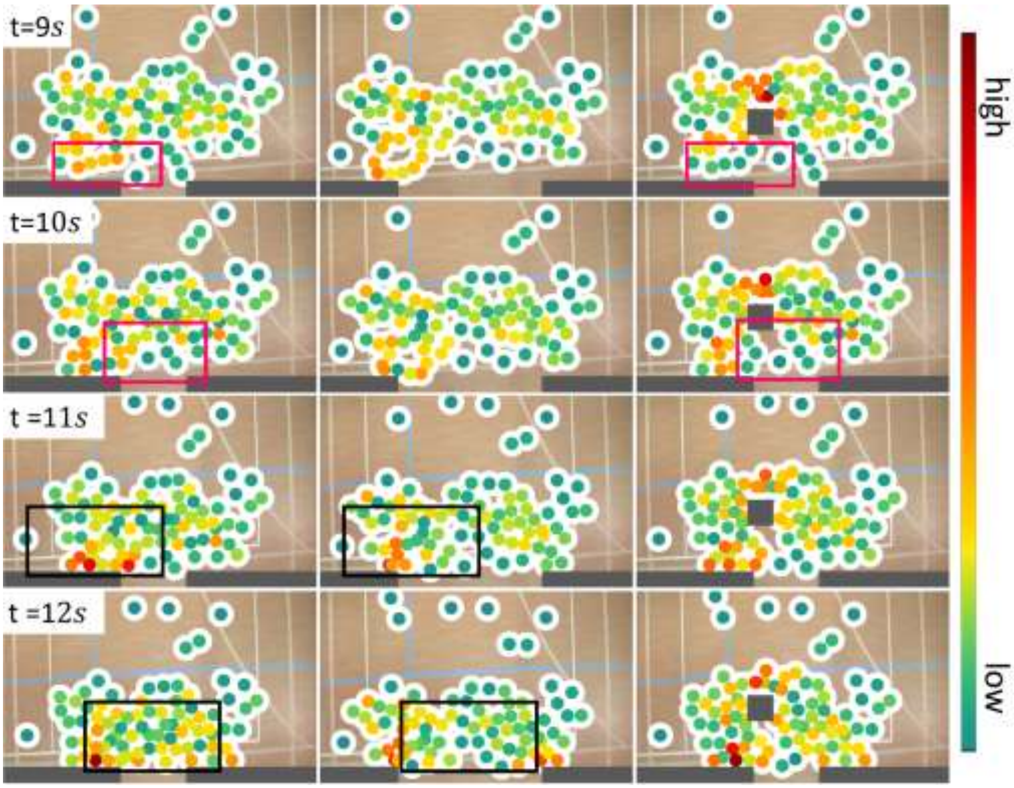

Figure 6. Prediction/Simulation. Left column: prediction in the same setting as the dataset; Middle column: simulation with a wider exit; Right column: simulation with an obstacle in front of the exit.

1.4.2. Drill Behaviour Learning and Simulation

We show one example where our model learns from an Evacuation Drill dataset (Figure 2 a). This is a dataset collected in a lab environment where the subjects and tasks are under control. The area is well intrinsically and extrinsically calibrated. We can obtain the head trajectories, not only in pixel coordinates, but in x-y-z coordinates in a world coordinate system. So, the density measured here is reliable. We start with this dataset due to it is much easier to interpret the ground-truth behaviours of people, which is more difficult for data in the wild. It will also make it easier to apply the model to some of the data collected WP1[3] later due to that they have similar settings in data collection. On this data, both the optical flows are steady and the trajectories of individuals are available. We learn our model on both data separately. The left column is a prediction for 12 seconds, the middle column is a simulation with a wider exit and the right column is with a pillar placed in front of the exit. Colours here indicate internal stress of the material, which is similar to how strongly people are pushing each other. The prediction (left column) shows high pressure starts to build up at the exit, while a

[3] *https://ped.fz-juelich.de/da/doku.php?id=evacuation_narrow_door_obstacle*

wider exist (middle column) would reduce that pressure and a pillar in front the exist (right column) would shift the high pressure from the exit to the pillar. These results are consistent with what was observed in the experiments and widely agrees with the qualitative results from existing research, which shows the model we propose can learn the behaviours under evacuation drills.

To give a more intuitive illustration, we show a video of simulation of the Evacuation Drill using Crowd MPM. We also compare with results with another the-state-of-the-art method Neural Social Physics (NSP). The video can be found on Youtube[4].

1.4.3. Hajj Behaviour Learning and Simulation

We also learn from the Hajj data (Figure 2 b). Note the data is from Youtube and was recorded by a distant camera in the wild. It was not possible to reliably get trajectories at all. So we trained our model on the extracted noisy optical flows.

After learning, we simulate Hajj which is a circular movement around the Kaaba (the central object). We first initialise the positions and velocities of all agents for a circular motions, then let the model simulate the movements in time. In the simulation, we apply an attraction force to all agents towards the centre of the Kaaba. Kaaba is a square object (Figure 7a) with a semi-circular wall on one side called Hateem (Figure 7c). The Hateem is considered to affect the nearby crowd flow. To verify whether our simulator can reflect this influence, we simulate the crowd with/without the Hateem. Figure 7 shows the results, wherein (a) and (c) are the snapshots at t=80s, (b) and (d) present the averaged density map of the central area (in the black box) over 40s duration.

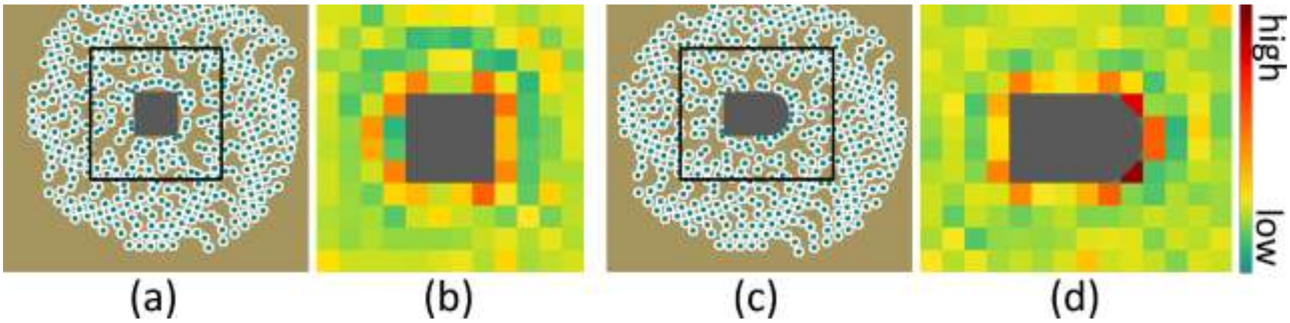

Figure 7. Left (a-b): learned behaviour in Hajj, Right (c-d): changed environment (the changed shape of the Kaaba). Colours indicate pressure. The dots represent the mass distribution of the crowd.

Different from (b) where the density distributions on each side of the Kaaba are similar, the densities near the semi-circular Hateem in (d) are higher than the counterpart on the other side. This is because the semicircle area can disrupt the counter-clockwise motion, thereby causing a compressed and mobbed area near the curved surface, which agrees with existing study. Besides, the densities around the corners are relatively high. A sudden turning is required in these areas, which will lead to congestion and high density. The simulation further proves our model can learn reasonable behaviours in alternative environment settings.

[4] *https://youtu.be/ODJwtzocuT0*

1.4.4. Hellfest Behaviour Learning and Analysis

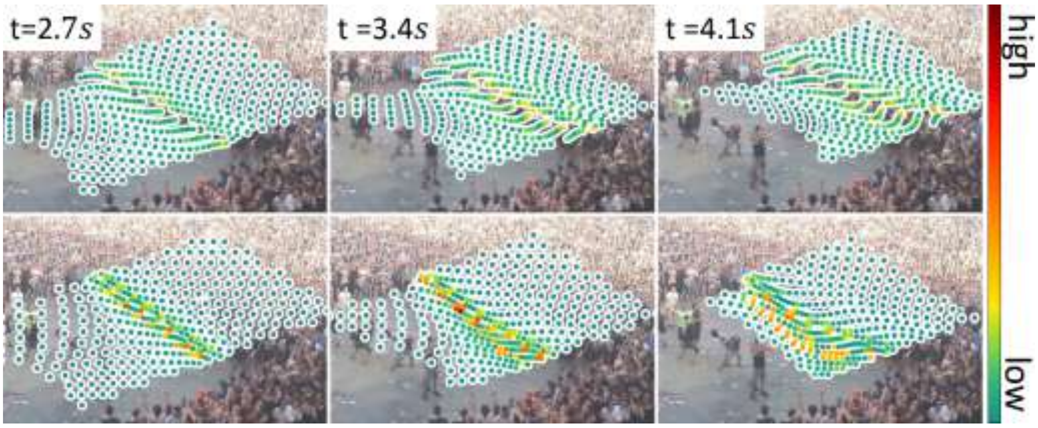

Figure 8. Hellfest Behaviour and Analysis. Top: a simulation of crowd trying to avoid direct crushing into one another. Bottom: a simulation of crowd aiming to directly crush into one another.

In Figure 8, we show that our model, after learning the behaviours, can simulate alternative situations of high-density crowds. The situation is that two groups of people intentionally run into one another in Hellfest (Figure 2 c). Although this video was recorded before our project, we chose it as it contains a special behaviour where the intentional crush in the middle of the crowd forms a 'wall of death', which was not recorded in Hellfest 2022/2023 within our project span. We use two settings of our model to analyse and to learn the behaviours. The top row shows that under the assumption that people might mitigate the crush in the middle by steering to the sides, the learned behaviours shows vorticity in the middle where the people in the centre of crush still try to avoid directly colliding with the other group. The bottom row shows an alternative behaviour where people intentionally crush into each other and the group on the right hand side slightly overwhelm the left group. Both behaviours are possible, but on this specific video, the latter behaviours happened based on our visual inspection. Nevertheless, this shows our model can potentially learn a diversity of behaviours in crushes.

1.5. Future work

The model and algorithm design and implementation have been completed. This source code is ready. This work will be submitted to a conference in the near future. Currently, we are finalising the methodology and conducting more experiments for evaluation. After the paper acceptance, we will also share the code and data. This will provide a novel tool to detect intense physical interactions between individuals from high-level video data. One way to combine this work with the tools developed in WP2 is to infer specific 3D body movements in physical interaction in high stress areas. It can also provide high-level simulation guidance with body-augmented simulation shown in D3.1. From the WP1 data, we will have both video data, the body motions and the exact trajectories from individuals in the video, which will be employed to train the combined model.

Conclusion

In conclusion, we have achieved our goals in D3.2 and delivered new tools for high-density crowd analysis and behavioural detection. Our deliverable contains one new system that can take raw and noisy high-density crowd videos, and analyse, predict and simulate crowd movements when the videos do not contain detail individual motions. This accomplishes the goal of proposing a new ML Detector to analyse crowd movements in videos.

References

Angel Garcimartín, Daniel R Parisi, Jose M Pastor, Cesar Martin-Gomez, and Iker Zuriguel. Flow of pedestrians through narrow doors with different competitiveness. Journal of Statistical Mechanics: Theory and Experiment, 2016(4):043402, 2016.