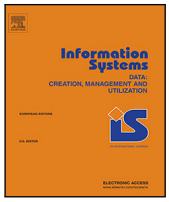
Information Systems 124 (2024) 102409

Contents lists available at ScienceDirect

Information Systems

journal homepage: www.elsevier.com/locate/is



HUM-CARD: A human crowded annotated real dataset

Giovanni Di Gennaro [a,*], Claudia Greco [b,c], Amedeo Buonanno [d], Marialucia Cuciniello [b,c], Terry Amorese [b,c], Maria Santina Ler [b], Gennaro Cordasco [b,c], Francesco A.N. Palmieri [a], Anna Esposito [b,c]

[a] Dipartimento di Ingegneria, Università degli Studi della Campania "Luigi Vanvitelli", via Roma, 29, Aversa (CE), 81031, Italy
[b] Dipartimento di Psicologia, Università degli Studi della Campania "Luigi Vanvitelli", Viale Ellittico, 31, Caserta, 81100, Italy
[c] International Institute for Advanced Scientific Studies (IIASS), via G. Pellegrino, 19, Vietri sul Mare (SA), 84019, Italy
[d] Department of Energy Technologies and Renewable Sources, ENEA, P.le Enrico Fermi, 1, Portici (NA), 80055, Italy

ARTICLE INFO

Recommended by Dennis Shasha

Keywords:
Human motion prediction
Human behavior analysis
Multi-agent systems
Pedestrian dynamics
Behavior categorization

ABSTRACT

The growth of data-driven approaches typical of Machine Learning leads to an ever-increasing need for large quantities of labeled data. Unfortunately, these attributions are often made automatically and/or crudely, thus destroying the very concept of "ground truth" they are supposed to represent. To address this problem, we introduce HUM-CARD, a dataset of human trajectories in crowded contexts manually annotated by nine experts in engineering and psychology, totaling approximately 5000 hours. Our multidisciplinary labeling process has enabled the creation of a well-structured ontology, accounting for both individual and contextual factors influencing human movement dynamics in shared environments. Preliminary and descriptive analyzes are presented, highlighting the potential benefits of this dataset and its methodology in various research challenges.

1. Introduction

The integration of intelligent systems within human environments is a reality that encompasses a wide range of domains and aims to become increasingly pervasive. In this context, one of the most exciting challenges is being able to transfer to such intelligent systems the ability to identify and understand human behaviors within real environments. In fact, humans show a natural ability to anticipate others' movements, avoiding potential dangers and obeying to social rules as greetings, group synchronous movements, avoiding collisions, etc. This ability is almost instinctive and linked to complex and not yet understood learning mechanisms, which inevitably fail to generate adequate mathematical models.

In recent years there has been a growing interest within the scientific community for these issues, linked both to purely research fields and their applicability within social and industrial domains as: abnormal behavior detection [1], emergency responses management [2–4], infrastructure design [5], transportation systems development [6,7], autonomous navigation systems [8], socially aware robots [9], and more. The interest is mainly supported by the incredible results achieved in the field of *Machine Learning*, particularly through *Deep Neural Networks*. Despite great strides, even the most modern systems [7,10,11] currently show a marked inability to capture human behaviors in complex scenario. In particular, crowded environments still prove to be particularly challenging, mainly because the increase in persons' density causes occlusions and overlaps as well as producing additional constraints among the various agents to be modeled. One of the main problems in this area is related to the lack of adequate video datasets on which to train and evaluate intelligent systems. Indeed, many of the existing video datasets do not consider crowded scenes and/or show semi-automatic and incomplete labeling. Furthermore, despite representing fundamental traits for understanding the scene [12–14], no one seems to have ever considered the psychological aspects linked to the heterogeneity of humans' conducts.

This work develops within the Socially-Aware Learning through Interactions in Crowded Environments (SALICE) project, whose objective is precisely to overcome the previous limitations by adopting a multidisciplinary approach. Aiming to combine social and technical expertise, this paper presents the HUM-CARD video dataset, a dataset hand-labeled by experts from various fields in order to gain a better understanding and description of human behavior in crowded environments.

In particular, a team comprising nine experts specializing in psychological and engineering disciplines dedicated approximately eight months to defining and labeling this dataset. The decision to limit the group size was based on considerations of organizational and managerial feasibility, facilitating mutual oversight among experts during the

* Corresponding author.
E-mail address: giovanni.digennaro@unicampania.it (G. Di Gennaro).

https://doi.org/10.1016/j.is.2024.102409
Received 29 March 2024; Received in revised form 17 May 2024; Accepted 18 May 2024
Available online 21 May 2024

labeling process. In total, the labeling process consumed approximately 5000 hours thus far, accounting for the time contributed by each individual expert.

Each pedestrian is identified by a specific ID, assigned to a bounding-box that suitably adapts to the shape and movements of the human it refers in the scene. Through an annotation ontology, any information deemed capable of representing an important feature for behavioral analysis (gender, age, current type of interactions, etc.) was added to the agent. Finally, a preliminary analysis is conducted of the possible benefits that this type of dataset can bring to multiple research fields.

2. Crowded environments

Many macroscopic and microscopic models have been proposed to operationalize the rules underlying collective motion patterns and individuals' interactions in crowded environments. Among these, physics-based ones are currently the most used to examine pedestrians' behaviors. Some examples are fluid-dynamic models, which represent global characteristics such as speed and density by considering the crowd as an entity subject to the same physical rules applicable to fluids, and those based on social forces, which consider the behavior of pedestrians by modeling attractive and repulsive rules to reflect different external and internal motivations [15,16].

More recently, some visual-based models, adopting a cognitive science approach to the collective behaviors modeling, have suggested that the local interactions among agents and their physical surroundings are driven by visual and cognitive heuristics, which allow individuals to make decisions under time pressures or a large information load [17,18]. Although the literature reports a variety of pedestrian models tested through experimental or simulation techniques, existing works on the human behavior in real-world crowded environments still lacks reliable data on pedestrians' features and a comprehensive classification of complex crowd behaviors [19,20]. However, such difficulty is understandable when considering the multiple internal and contextual factors that influence the human behavior in shared environments.

On a psychological level, when humans perform an action there are several internal processes that drive their behavior: the desired future goal (e.g., "I want to arrive as fast as possible at the destination"); the beliefs orienting the decisions (e.g., "I think this is the best path to arrive to the destination"); the way the objective is achieved (e.g., "I do not want to be harmed in reaching the destination") [21]. Moreover, pedestrians usually vary previous choices along the path to reach their goals, making different intermediate decisions that influence both the evolution of trajectories and changes in speed [22]. Obviously, these internal aspects are inextricably related to the surroundings' characteristics [6]. In fact, they concern both static and dynamic entities that compose the environment, such as the physical configuration of the space, the road layout, the lighting, the speed and density of the other agents, as well as the weather conditions.

Other contextual factors are those related to the cultural, demographic, and physical features of the pedestrians [13]. Cultural differences are particularly evident precisely in crowded contexts, suggesting that the speed of Indian pedestrians is less influenced by the density of people than that of German pedestrians [23], and that continental Europeans prefer to avoid collisions by moving to the right side while in Japan and Korea the left side is more frequently chosen [24]. Regarding physical characteristics, gender differences are commonly found in pedestrian behaviors [25]. Many studies report that male and female pedestrians differ both due to microscopic movements characteristics, including step-length, walking speed, pace or gait [26–28] and to rule compliance degree and risk-taking behavior [29]. Findings suggest that males walk faster, tend to wait for shorter amounts of time and infringe more rules when crossing a roadway compared to females, who generally engage in safer behavior. Furthermore, age also appears to be associated with differences in walking behavior. More specifically, there is a general agreement on the hypothesis that walking speed declines with age [30,31]. This may be ascribed to cognitive and motor changes that slow down movements in the elderly. In this context, factors like exhaustion, perceived fatigue, and stress may also influence walking behavior [32,33].

2.1. Interactions

In a multi-agent setting, special attention should be paid to interactions between pedestrians. Interactions play a pivotal role in shaping the actions that agents undertake on the scene, thus defining the relations among them. These relations represent the foundation on which social groups are built.

Although the most accurate description of social groups identifies them as three or more people perceived (by themselves or others) as a group [34], this definition is usually also extended to dyads, typically considered social groups in the field of human motion dynamics [35]. Despite the lack of a univocal definition of social groups in pedestrian crowds, there is a general agreement on the effects that they may exert on the collective dynamics on different observational levels.

From a macroscopic point of view, the presence of social groups influences the influx and density of the crowd [36]. Indeed, various studies report that social groups move more slowly than individual pedestrians, slow down the average speed of crowd flow, and are associated with larger empty space around group members [37,38]. At the microscopic level, however, social groups exhibit specific configurations and patterns that influence movements and trajectories [39]. This level of observation concerns the typical arrangements that social groups adopt when they move in space, mainly motivated by strategies that favor communication and interaction between group members and which at the same time aim to avoid collisions with other pedestrians present on the street. scene [39–41]. For instance, it has been found that large groups (more than 6 people) tend to split into smaller groups in order to preserve the members' relative positions [40], while members of dyads generally walk side by side, with a distance of approximately 80 cm between their centers of gravity. Conversely, triads walk by forming a V-shape, with the pedestrian in the middle walking a little bit behind the others, a pattern that typically turns into a U-shape when the number of members increases to 4 or 5 people [39,42–44].

However, just as the presence of a group affects the crowd dynamics, crowd features can also force the group structure. For example, as crowd density increases, the spatial arrangement on the group tends to narrow laterally, forming a river-like configuration [45]. In conclusion, there is a considerable corpus of literature corroborating the mutual relationship between pedestrian interactions within social groups and collective dynamics. Given their importance in modeling the evolution of human trajectories in a shared space, these aspects cannot be excluded or superficially evaluated within the datasets needed to train and test modern intelligent systems, since the reliability of such methods is strictly linked to the quality of the information provided as ground truth.

2.2. Pedestrian datasets: state-of-art

A bibliographic search of the existing datasets for human trajectory forecasting and/or pedestrian detection tasks has been performed to examine the unexplored aspects in the literature and identify alternative solutions for data collections. Two of the most used datasets for this purpose are the *UCY* [46] and the *ETH BIWI Walking Pedestrian* [47] datasets. Both of them contain real videos with rich multi-human interaction scenarios collected in public spaces with a top-view camera (captured at 2.5 Hz).

The UCY dataset contains four sequences (two relating to an urban street and two collected on the university campus), resulting in a total of 17′13″ of video in which the trajectories of over 700 pedestrians

were semi-automatically detected. The ETH BIWI dataset, on the other hand, consists of two pedestrian scenes (an urban street and the university campus of ETH Zurich) in which the positions and speeds of approximately 650 pedestrians were manually annotated over a total of 21′32″ of video.

Another well-known dataset is the *Stanford Drone Dataset* [48], which includes different types of targets (pedestrians, bicycles, cars, skateboarders) moving around the Stanford university campus. Videos were recorded through a 4k camera mounted on a drone (1400 × 1904) and data consists of more than 100 different top-view scenes with a total of 20 000 semi-automatically annotated targets and trajectories with bounding boxes. In addition to pedestrian trajectories, annotations also specify two types of auto-detected interactions: the target-target interaction (e.g., a cyclist avoiding a pedestrian) and the target-space (e.g., a cyclist going around a roundabout).

Similarly, the *Edinburgh Informatics Forum Pedestrian* dataset [49] was also collected on the university campus from a bird-eye perspective, and consists of more than 92 000 automatically captured trajectories of pedestrians in an open space at the University of Edinburgh. Although the specifications mention the use of a 640×480 px camera at 9 fps (excluding program crashes), the dataset lacks raw images; it only provides bounding boxes for the targets and an RGB histogram of the associated pixels. The *PETS 2009* dataset [50] is instead noteworthy precisely for the quantity of images made available, consisting of three video sequences recorded through 8 monocular cameras (768 × 576 or 720 × 576) from different perspectives of a public space in the Whiteknights Campus, University of Reading (UK).

The *DUT* dataset [51] was collected on the campus of Dalian University of Technology (DUT) in China, with a DJI Mavic Pro drone (1920 × 1080 at 23.98 fps), considering areas accessible to both pedestrians and vehicles. The dataset consists of 17 pedestrian crossing scenarios and 11 shared space scenarios near a roundabout, for a total of 1793 automatically annotated trajectories. The *ATC* dataset [52] features videos captured by 3D range sensors (10–30 Hz) in a shopping center spanning 900 m^2 for 92 days. The tagging was performed manually over just two days in November, extrapolating the position, speed, height and body angle of the 407 pedestrians.

A similar dataset collected from publicly available webcams is the *Mall* dataset [53], which includes head position annotations of 60 000 pedestrians, automatically labeled in 2000 frames in JPEG format (640× 480 captured at <2 Hz). Another indoor dataset is the *Grand Central Station* (GCS) one [54], which consists of a black and white video whose duration is 33′20″ (720 × 480 at 25 fps). In the original work only a small number of the total paths were automatically annotated, but in a subsequent work [55] the same video was used to manually annotate the paths of all the 12 684 pedestrians.

A different class of datasets includes video sequences from various urban environments, such as those provided by the *Multi-Object Tracking Challenge* (MOT).[1] The MOTs datasets [56–58] contain video sequences in unconstrained environments, classified according to type of camera (static and moving), viewpoint (high, medium, low) and weather conditions (sunny and cloudy). Annotations were made following a protocol which was not the same across the sequences. They include three possible categories: moving or standing pedestrians; people that are not in upright position; vehicles and obstacles. Also, the *VIRAT* video dataset [59] is a collection of 16 different video sequences recorded in outdoor scenes. Data consists of 25 h of recordings made with stationary ground HD cameras (1080p or 720p, at a 25–30 Hz) with varying viewpoints. Tracks of bounding boxes for moving objects were annotated through Mechanical Turk, whereas events annotation were identified by experts. Event types were categorized into three different classes: single person events (e.g., walking, running, standing); person and vehicle events (e.g., getting in or out of a car, bicycling); person and facility events (e.g., entering or exiting a facility). Within the human trajectory prediction context, the TrajNet meta-dataset [60] represents a large-multi scenario forecasting benchmark. TrajNet is defined as meta-dataset since it combines annotations from four datasets: ETH BIWI Walking pedestrian, UCY, the Stanford Drone and the PETS, for a total of 11 448 trajectories.

2.2.1. Datasets in controlled environments

The mentioned datasets were collected in real environments (shopping malls, urban streets, or university campuses), while an entirely different class of human motion datasets are those developed under controlled conditions. An example is represented by the *bottleneck* dataset [61], which focuses unidirectional pedestrian flow through bottlenecks under experimental conditions. The experimental setup was arranged at the Central Institute for Applied Mathematics (ZAM) of the Research Centre Julick. Participants were ZAM students and staff who were asked to walk at normal speed from a waiting area through a narrower corridor without pushing others, to examine crowd flow formation. The setup was filmed by two cameras (720 × 576 at 25 fps) located above the center and on the entrance of the bottleneck. Trajectories were annotated by manually marking the center of the head of each person.

The *CITR* dataset [51] consists of experimentally designed vehicle-crowd interaction scenes. Data was collected in a parking space close to the Control and Intelligent Transportation Lab at the Ohio State University, through a DJI Phantom 3SE drone (1920 × 1080 at 29.97 fps). To test the influence of a vehicle on pedestrian behavior, six different scenarios were carried out: three of them entail only pedestrians, and the remaining ones entails both a vehicle and pedestrians interacting with each other. Pedestrians were members of the lab which were instructed to walk from a starting point to a specific destination. In total, 38 videos were recorded including about 340 pedestrian trajectories. The tracking of pedestrians' and vehicles' locations was automatically initialized upon entering the crosswalk region of interest and stopped upon exiting, with only initial positions manually annotated.

Finally, the recent *THOR* dataset [62] was developed for a human–robot navigation study. It includes motion trajectories of individuals in an indoor experimental setting, featuring interactions between pedestrian groups and a robot navigating through varied obstacles. Over 600 participants and group trajectories have been recorded in 60 minutes. To collect motion trajectories, the Qualisys Oqus 7+ motion capture system (100 Hz) with 10 infrared cameras placed on the room's perimeter was used. The experimental procedure required to assign social role and tasks to the participants in order to mimic typical activities observed in crowded spaces. Three social roles were considered: visitors (i.e., participants walking alone or in group in the space), workers (i.e., participants carrying boxes around the space) and inspector (i.e., a lone participant moving between multiple targets in the space). The dataset includes 13 scenes in 3 different conditions: one obstacle (i.e., no robot and one obstacle in the space), moving robot (i.e., the robot is moving, and one obstacle is included in the space), three obstacles (i.e., three obstacles and no robot in the space).

2.3. Behavior classification

The description just made on existing datasets obviously has the primary purpose of determining the state of the art on the topic and identifying unexplored aspects of the annotated datasets, which require further considerations. In this regard, the reviewed literature suggests that most datasets are automatically or semi-automatically annotated, which can cause a biased and noisy evaluation of the underlying information. Other problems are related to the length of the considered trajectories, which are often not long enough to achieve the prediction tasks, or to the relatively moderate number of annotated agents. However, more important seems to be the fact that both pedestrian and group features are usually discarded during the annotation process,

[1] https://motchallenge.net/.

Table 1
Description of the behavioral labels used in human trajectory prediction and detection fields.

Behavior	Description	Reported in:
Side-by-side	Agents walking side by side in a line perpendicular to the direction.	[39]
Collision avoidance or Steering	An agent adjusting trajectory to avoid another agent coming from the opposite direction.	[64,65]
Forming lane	Agent in unidirectional pedestrian flow, moving in river-like configuration.	[66]
Group	More agents moving together by keeping a close and consistent distance with at least one neighbor on his/her side during the entire scene.	[46,64]
Halting	A single agent (not part of a pedestrian flow) that stops to let another agent pass ahead of him/her, to avoid a collision.	[67]
Holding hands	An agent (stationary or in movement) holding hands with another one.	[13]
Imitation	An agent modifying his/her trajectory or destination to move in the same direction of the other agents nearby (emergency situations).	[68]
Passing Pause to look at something	An agent increasing his/her walking speed to pass another agent(s). An agent interrupting his/her walk to visibly direct the attention towards something.	[67] [46]
Person following or Leader following	An agent following the trajectory of the other agent in front of him/her (leader) towards a common destination.	[64,69]
Physical interaction	An agent physically touching another or other agents (high crowd-density situation).	[68]
River-like	A configuration characterized by the presence of a leader who guides the other member(s) in crossing the space in a riverlike pattern.	[41]
Smooth collision avoidance	An agent avoiding another agent by just moving his/her body without changing the original trajectory.	[46]
Splitting	An agent which temporarily splits from the rest of his/her group to re-join it after the obstacle has been passed.	[70]
Stationary crowd behavior	More than two agents being part of a stationary group of people.	[71]
Stop-And-Go	An agent in a pedestrian flow that stop to let another agent pass ahead of him/her (applies to bottleneck; bidirectional flows).	[66]
U-shape	A group configuration in which an agent is part of a group with more than 3 agents, in "U" like walking pattern.	[39]
V-shape	A group configuration in which an agent is part of triad in "V" like walking pattern.	[39]

focusing exclusively on trajectories and omitting classification in terms of age, gender, group relations or their configurations in space (for a review, see [63]). To introduce such characteristics it is therefore necessary to first create a taxonomy of pedestrian behaviors, individually or in groups, which can be used in the fields of human trajectory prediction and/or detection. Searching for these peculiarities in the literature, Table 1 shows the list of the main behaviors, their short description and references.

3. HUM-CARD

With the aim of introducing the behaviors just listed into the pedestrian annotations of real crowded scenarios, the *Human Crowded Annotated Real Dataset* (HUM-CARD) was therefore created. For this purpose, an *Annotation Ontology* (AO) was defined, consisting of a systematic nomenclature of the collective and individual characteristics of human movement chosen to manually annotate the selected videos. The latter represents the most important part of the annotation process, which will ultimately consist in the application of this ontology to assign the various behaviors both to individual pedestrians and to their configurations as a group.

3.1. Dataset creation

Before introducing the ontology, and independently of the previous bibliographic research, a further research phase was conducted to identify potential videos that include the crowding characteristics underlying the entire project. Further refinement was then carried out to exclude the videos:

- not public;
- based on video recordings from cameras mounted on moving vehicles;
- related to the top-view and/or first perspective of the agents.

While the ultimate goal is to continue annotating the set of videos exhibiting the above-mentioned characteristics, currently only three videos have been selected for the time-consuming manual annotation process.

The first video is black and white surveillance footage recorded in a shopping mall (Fig. 1(a)) and available for free on the YouTube website.[2] It has a duration of 13 seconds, contains 341 frames (720 × 1280 at 25 fps) and is in MP4 format (2.99 MB). The second video was a sequence extracted from the UCY dataset [46] (Fig. 2(a)). This is one of the sequences recorded on the university campus, the duration of which is 3′36″ for a total of 5405 frames (576 × 720 at 25 fps). This video is also in MP4 format (64 MB). The last video is represented by the black and white surveillance footage of Grand Central Station in New York City extracted from the GCS [54] dataset (Fig. 3(a)), which, as mentioned, has a duration of 33'20" for a total of 50 010 frames (captured at 25 fps). The original format of the downloaded video was AVI (1.1 GB), but it was converted to MP4 to facilitate the subsequent labeling process. For the same purpose, the video was split into three segments (each about 11 minutes long) using Avidemux editing software An example featuring images cropped from the center of 10-second segments of the three videos used is shown in Fig. 4.

To facilitate the annotation process, all videos were initially uploaded to the Labelbox platform.[3] Labelbox provides users with an online tool that enables efficient data management, while also providing various data labeling tools. This choice was mainly dictated by two reasons: an intuitive and user-friendly interface, and the ability to develop a well-structured ontology, including all the information of interest and their interrelationships.

The approach used for video labeling involves manually creating a bounding box for each agent present in the scene (Figs. 1(b), 2(b)

[2] https://www.youtube.com/watch?v=WvhYuDvH17I.
[3] https://labelbox.com/.

(a) (b)

Fig. 1. Example frame from YouTube video (a) and the corresponding annotation (b).

(a) (b)

Fig. 2. Example frame from UCY video (a) and the corresponding annotation (b).

(a) (b)

Fig. 3. Example frame from GCS video (a) and the corresponding annotation (b).

and 3(b)). Having chosen to work directly on the image plane (avoiding transformations that can strongly depend on the precision of the context data), the size and shape of each bounding box is adapted in each frame, based on the movements observed. The annotation of each frame is also necessary to correctly specify the individual and collective characteristics of the various agents, which may be subject to possible state changes (for example, an agent can stop and start walking again at subsequent instants of time). The labeling process included observing the subject at different time intervals and conducting a double-blind check among multiple experts for any doubtful cases.

3.2. Limitations and difficulties in the labeling process

Nonetheless, there are some difficulties related to particular videos, which currently limit the annotation process. These problems have so far mainly concerned the Grand Central Station video, which due to its technical characteristics (lack of colors and limited resolution), has hindered the creation of bounding boxes for some pedestrians due to the high density of them in specific points of the scene. The same reasons have also made difficult to reach a unanimous consensus regarding some characteristics of particular pedestrians. This led, for example, to limiting the age options to the child/adult dichotomy, as any attempt at further precision on the age range (e.g. child, adolescent, middle-aged, elderly) was not possible. Finally, the density of agents and the variety of their behaviors on the scene significantly influence annotation times, putting a strain on both the workload and the computing power of the Labelbox platform. However, having omitted the labeling of some agents represents only a temporary choice which will certainly be compensated for later. In fact, it should be noted that the annotation process is still ongoing and that the dataset will be periodically updated.

While Labelbox has proven to be an effective annotation tool, it too ultimately poses various limitations. The first is related to the creation of the ontology, which allows only a finite number of options to be generated in the subclassifications. Additionally, Labelbox only allows

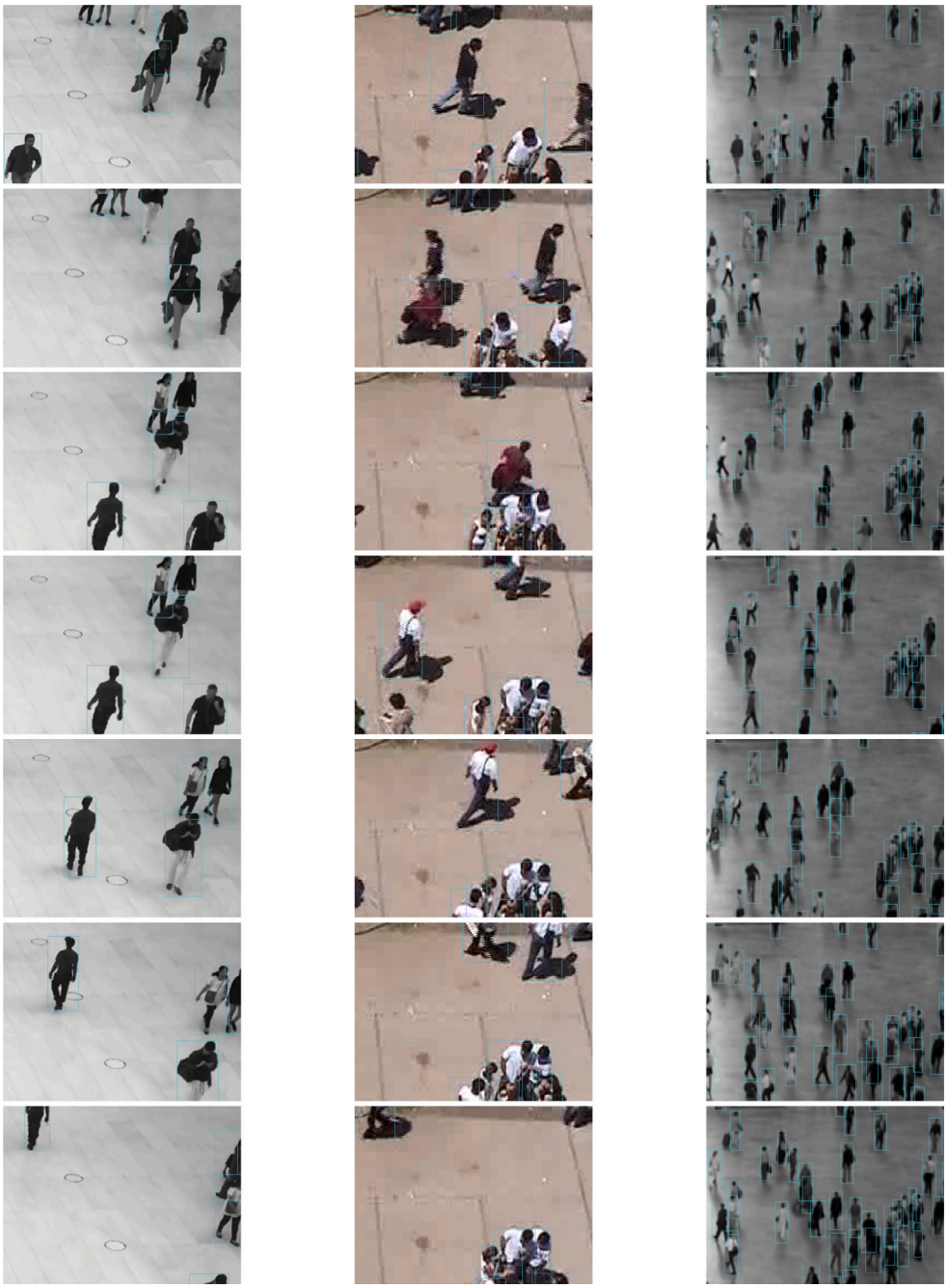


Fig. 4. Example featuring cropped images extracted from a 10-second segment of the three videos, using the same zoom level for all. YouTube's image is on the left, UCY is in the center, and GCS is on the right.

you to work online, which means the annotation process is sometimes affected by bandwidth and connection speed. Thus, a significant portion of the labeling was conducted using custom Python software, allowing for the expansion of options that could not be entered directly.

3.3. Annotation ontology

An ontology can be viewed as an organized collection of terminologies denoting various types of entities found in the real world of interest. Its primary purpose is to comprehensively represent and classify the entities within a specific domain of knowledge. In simpler terms, ontologies serve as well-structured systems of naming entities, wherein the terms are arranged hierarchically in order to improve the systematic understanding and analysis of the domain of interest. To create our AO, we initially examined potential motion behavior categories based on existing literature on human trajectory prediction [19,62,72–74]. Subsequently, these categories were specifically expanded through the creation of further terms that took into consideration elements present in the various videos examined.

As already mentioned, the Labelbox platform was used to build the AO, which allows the specification of the main object, its subclassifications, related options and the annotation input type. With the exception of the numerical values reserved for IDs, only two types of input were used: the radio control, which requires the choice of a single option from a list of alternatives, and the checklist, which vice versa allows the selection of multiple options. The main classes of our AO

are represented by the following characteristics, linked to the bounding box of each agent.

- ID (Number, mandatory): represents the unique numeric value assigned to each agent in the scene.
- Action (Radio, mandatory): describes the type of movement of the agent in space.
- Age (Radio, mandatory): determines whether the agent is an adult or a child.
- Gender (Radio, optional): defines the gender of the agent, when it is recognizable.
- Detail (Checklist, optional): contains several options that describe some characteristics that could influence the movement behavior of the agents.
- Group (Checklist, optional): establishes whether the agent belongs to a group or not, possibly linking it to the IDs of the other members of the same group.
- Group Type (Radio, optional): highlights, if the agent belongs to a group, the possible physical configuration of that group.
- Group Role (Radio, optional): in particular group configurations, this characteristic defines whether the agent takes on the role of follower (the agent considered follows the other agent/s) or leader (the agent considered precedes the other agent/s).
- Collision (Checklist, optional): it is a list of agent IDs that physically collides with the considered agent.
- Holding hands (Checklist, optional): it is a list of agent IDs that holds hands with the considered agent at a particular moment in the scene.
- Partial view (Checklist, optional): specifies whether or not the agent appears completely in the scene, identifying visible body parts including legs, arms, torso and head. In fact, when an agent enters or exits the scene, his bounding box will only include a part of his body, which will be highlighted through this feature.

The details of the possible options not defined in the list of previous classes, and fundamental for the definition of the labeling protocol followed, are finally reported in Table 2.

To illustrate the completeness and complexity of the scenarios defined by the ontology, Fig. 5 presents a single agent tracked over several steps, with corresponding main labels for each moment.

Fig. 5. Example extracted from GCS, illustrating the tracking of a single agent and highlighting the main labels associated with various moments in time. The agent is highlighted in green, while the other members of the groups that form are shown in blue.

4. Results and strengths

The definition of an extremely varied AO, which uses most of the indications coming from various branches of literature, has allowed us to define a labeling process that could significantly contribute to the field of modeling human behavior in crowded environments. Adding more features to the various agents within video datasets is of fundamental importance to enrich the understanding and improve the performance of analysis and prediction models, especially considering that existing datasets completely lack such additional information. Introducing details on the individual characteristics of the agents (such as age, gender, etc.) allows us to more precisely model human behavior in specific contexts. On the other hand, adding information about social interactions can improve understanding of relationships between people. This is particularly relevant in contexts such as surveillance or analyzing behavior in crowded spaces. Moreover, additional features give models more information on which to base their predictions. For example, understanding whether a person has a backpack could be crucial to predicting their path in certain situations.

4.1. Dataset statistics for additional features

Analysis of the three videos included in our dataset reveals a variety of crowded contexts and human movement dynamics. Each video captures real-world scenarios, providing a broad spectrum of situations that can be explored through manual annotations. Interesting, in this context, is the analysis of the statistics relating to the presence of particular additional characteristics, which alone seem to be able to identify the reference scenario.

In particular, for the first video extrapolated from YouTube, in which all agents have been completely labeled, almost only people considered adults are observed. This characteristic corroborates the idea that the filmed area is far from the play areas typically found in shopping centers, and the greater presence of the male gender (Fig. 6(a)) suggests the proximity of specific shops (e.g., men's clothing). The excessive presence of bags or packages carried (Fig. 6(c)) also reinforces the awareness of being in a shopping center, where most of the groups are actually made up of dyads placed side by side (Fig. 6(b)).

Similarly, in the UCY video the observation of an external environment is also identified by the almost uniform presence of gender (Fig. 7(a)).

Although even in this case the presence of transported objects is high (after all, we are on a university campus, where the presence of bags is inevitable) it is however almost compensated by the total of

Table 2
Sub-classification's options of the annotation ontology.

Class	Option	Description
Action	Walk	The agent is moving through the scene or is wandering around with no specific destination.
	Paused	The agent is not moving on the scene or stops for few seconds. In the frames considered the bounding box's position is not changing.
	Run	The agent is running through the scene.
Detail	Mobile phone	The agent (moving or not) is focused on the phone screen and not paying much attention to the path.
	Waiting for	The agent is on pause, visibly waiting for someone (e.g., another agent) or something (e.g., his/her turn).
	Observing	The agent (moving or not) is observing something other than the phone (e.g. a newspaper, the train timetable, etc.).
	Carrying something	The agent is carrying something (e.g., a bag, a backpack, a trolley, a cup, etc.).
	Sitting	The agent is sitting somewhere on pause.
	Leaned	The agent is leaning against a wall.
	Go up	The agent is walking up a flight of stairs.
	Go down	The agent is walking down a flight of stairs.
	In line	The agent is waiting in a queue (e.g., to buy a ticket, at the info point, etc.)
	On a call	The agent (moving or not) is on a phone call. This event is different from looking at the phone screen, where the agent's attention is divided between the screen and the path ahead.
	Briefly interact	Two or more agents briefly interact with each other (e.g., an agent asking something to another). The interaction is rapid and not prolonged in time. This label is not assigned to the members of the same group.
Group type	V-shape	A group of 3 or 4 moving agents, whose structure shows a central position moved slightly behind or ahead of the others.
	Side-by-side	Two or more members of a group (moving or not) placed side by side.
	River-like	Group members which are walking in a single line.
	Circle	Groups with more than 2 members which are in a circle.
	Face-to-face	Dyads in which the two agents are one in front of the other.

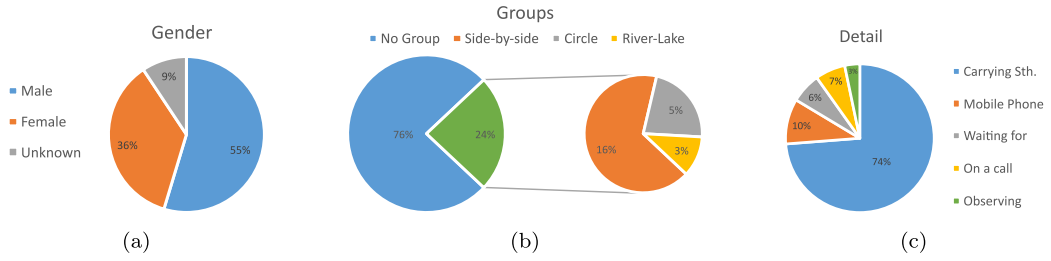


Fig. 6. Label distribution for three different main classes of YouTube video.

all the other possible activities (Fig. 7(c)). Furthermore, the location on a university campus also appears to be supported by the large presence of groups (Fig. 7(b)) compared to the total of adult agents currently labeled. Finally, the greater crowding is also observed from the collisions observed.

In the last video, corresponding to the Grand Central Station (GCS) dataset, more than 300 agents have been completely labeled. However, the number of children remains very low, which is normal given the location in a train station. Interesting is once again the fact that the luggage (Fig. 8(c)) actually represents a clear element of the context in which this video was made. Furthermore, also in this case, as in the first video, the presence of the groups seems less preponderant (Fig. 8(b)), although the groups present take on more varied facets in their typology.

4.2. Comparison with automatic detection

The importance of focusing on crowded environments (also introducing bounding boxes that take into account the dimensions of the agent on the image plane), is represented by the difficulty with which the various tasks can currently be carried out by modern automated systems. To provide a demonstration of this, Fig. 9 illustrates the results obtained on the same three frames extracted from the scenarios considered (see Figs. 3(b), 2(b) and 1(b) for comparison) of the application of the YOLOv7 model, currently considered state of the art.[4]

As we can observe, several pedestrians are not detected, due probably to very crowded environments, the too small dimension of the pedestrians or the presence of shadows. These results show how an automatic detection, even though highly performant as YOLOv7, is not able to solve the problem of the detection of pedestrians in very crowded scenarios and further demonstrate the need for the type of dataset created.

5. Future research directions

We think that the realm of predicting human trajectories based on their interactions is very promising and, for this reason, we will

[4] https://github.com/WongKinYiu/yolov7.

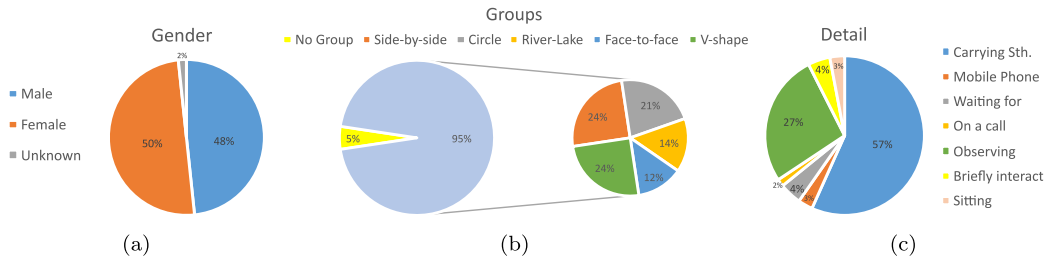


Fig. 7. Label distribution for three different main classes of UCY video.

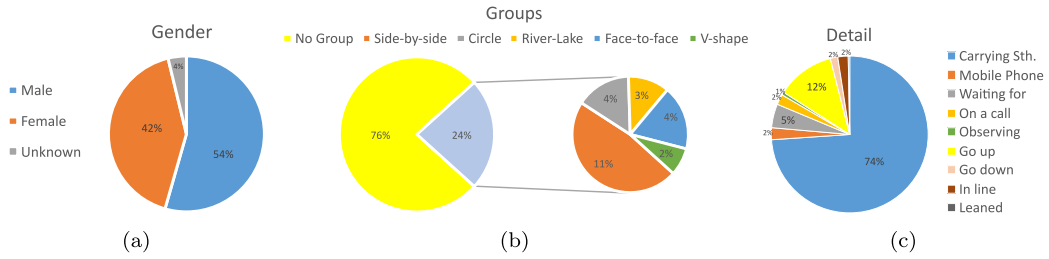


Fig. 8. Label distribution for three different main classes of GCS video.

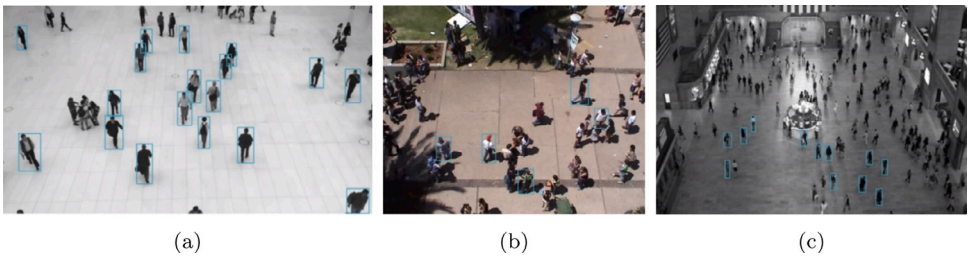


Fig. 9. Application of the YOLOv7 model to frames extracted from the three videos.

continue to enhance the dataset with new annotations and/or different considered scenarios. To facilitate research in this domain, a dedicated web page will be maintained, providing users with access to the different versions of the dataset (https://mlunicampania.gitlab.io/hum-card/). This web page will serve as a central hub for researchers and practitioners that will be able to explore various approaches and create interdisciplinary collaborations combining insight from computer vision, behavioral psychology, and urban planning to develop holistic models that take in account environmental and human-centric variables. The HUM-CARD dataset, composed of labeled videos of pedestrians in various movement scenarios, along with their interaction types, represents a valuable resource for the study of Multi-Agent Systems (MAS) in different contexts. Some research perspectives related to the HUM-CARD dataset can be:

- Behavioral Modeling: considering the labeled interactions, the researchers can identify patterns, norms, and deviations useful in different contexts such as video-surveillance, emergency, big events' management, etc. [75,76].
- Agents' Simulation: the labeled pedestrian interactions allow to develop and validate simulation models of pedestrian's movement. For example, researchers can use this data to design agents that mimic the observed pedestrian behaviors, enabling simulations that closely resemble real-world scenarios [77–80].
- Human-Robot Interaction: In the context of MAS, understanding how pedestrians interact with each other is crucial for developing effective human–robot interaction strategies. This dataset can aid in the design and evaluation of robotic systems that navigate among pedestrians in a socially acceptable and safe manner [81, 82].
- Traffic Flow Optimization: understanding how people move can play a significant role in the traffic flow modeling and control in both indoor and outdoor spaces [83,84].
- Human-Centric Design: Architects and urban planners can benefit from insights gained through the dataset to create human-centric designs for public spaces. Understanding how pedestrians interact with the environment and each other can be used to enhance user experience, safety, and security in public spaces. The dataset can be used to identify potential risks, vulnerabilities, and factors influencing pedestrian safety, aiding in the design of safer urban environments [85].

6. Conclusion

The current work proposes the HUM-CARD annotated dataset, with the aim to contribute knowledge in the field of the modeling of human motion behaviors in shared environments. The dataset may represent an innovative attempt for solving some still debated issues related to human trajectory forecasting and detection. Indeed, starting from a comprehensive bibliographic search, many individual and contextual factors, which have been overlooked in the literature, have been here considered in the definition of the methodology underlying the dataset creation. In this regard, the paper provides an organized ontology of the human behaviors which could be observed in crowded spaces, by considering both individual and relational features. Such categories have been used in the labeling process of three selected videos, which depict different types of outdoor and indoor spaces and differ from each other in terms of crowd density and contexts.

The proposed ontology seems to be applicable to all the considered environments, supporting its suitability and replicability features. Analyses performed on the collected annotations suggest that the dataset entails actions and relationships among pedestrians reflecting a wide range of individual and interactional behaviors in shared environments. Furthermore, the comparative analysis results support that the proposed dataset is able to provide more accurate information compared to one of the most performing models in the field. Overall, the HUM-CARD dataset may represent a valuable benchmark for human detection and trajectory forecasting, with many possible applications both in the academic and practical domains.

CRediT authorship contribution statement

Giovanni Di Gennaro: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Project administration, Methodology, Funding acquisition, Data curation, Conceptualization. **Claudia Greco:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Methodology, Investigation, Data curation, Conceptualization. **Amedeo Buonanno:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Methodology, Investigation, Data curation, Conceptualization. **Marialucia Cuciniello:** Writing – review & editing, Visualization, Validation, Investigation, Conceptualization. **Terry Amorese:** Writing – review & editing, Visualization, Validation, Data curation. **Maria Santina Ler:** Writing – review & editing, Visualization, Validation, Data curation. **Gennaro Cordasco:** Writing – review & editing, Visualization, Validation, Supervision, Methodology, Investigation, Conceptualization. **Francesco A.N. Palmieri:** Writing – review & editing, Visualization, Validation, Supervision, Project administration, Methodology, Investigation, Conceptualization. **Anna Esposito:** Writing – review & editing, Visualization, Validation, Supervision, Project administration, Methodology, Investigation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The dataset will be shared at the link in the paper after the paper is accepted.

Acknowledgments

This work was partly supported by the project SALICE (DR 834/2022), from the program “Giovani Ricercatori” (DR 509/2022) of the Università della Campania “Luigi Vanvitelli”, the EU-H2020 program grant No. 823907 (MENHIR), and the NextGenerationEU, PNRR Mission 4 Component 2 Investment 1.1 – D.D n.1409 del 14-09-2022 PRIN 2022 – Project code “P20222MYKE” - CUP: B53D2302598000 (IRRESPECTIVE). The work of G. Di Gennaro was also supported by the Italian Ministry for University and Research (MUR) - PON Ricerca e Innovazione 2014–2020 (D.M. 1062/2021).

References

[1] R. Mehran, A. Oyama, M. Shah, Abnormal crowd behavior detection using social force model, in: IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2009, pp. 935–942, http://dx.doi.org/10.1109/CVPR.2009.5206641.

[2] H. Dong, M. Zhou, Q. Wang, X. Yang, F.-Y. Wang, State-of-the-art pedestrian and evacuation dynamics, IEEE Trans. Intell. Transp. Syst. 21 (5) (2020) 1849–1866, http://dx.doi.org/10.1109/TITS.2019.2915014.

[3] X. Zheng, T. Zhong, M. Liu, Modeling crowd evacuation of a building based on seven methodological approaches, Build. Environ. 44 (3) (2009) 437–445, http://dx.doi.org/10.1016/j.buildenv.2008.04.002.

[4] D. Helbing, I. Farkas, T. Vicsek, Simulating dynamical features of escape panic, Nature 407 (6803) (2000) 487–490, http://dx.doi.org/10.1016/j.buildenv.2008.04.002.

[5] B. Jiang, Simped: Simulating pedestrian flows in a virtual urban environment, J. Geogr. Inf. Decis. Anal. 3 (1) (1999) 21–30, URL https://publish.uwo.ca/jmalczew/gida_5/Jiang/Jiang.htm.

[6] A. Rasouli, J.K. Tsotsos, Autonomous vehicles that interact with pedestrians: A survey of theory and practice, IEEE Trans. Intell. Transp. Syst. 21 (3) (2020) 900–918, http://dx.doi.org/10.1109/TITS.2019.2901817.

[7] C. Chen, Y. Liu, S. Kreiss, A. Alahi, Crowd-robot interaction: Crowd-aware robot navigation with attention-based deep reinforcement learning, in: International Conference on Robotics and Automation, ICRA, 2019, pp. 6015–6022, http://dx.doi.org/10.1109/ICRA.2019.8794134.

[8] A.V. Savkin, C. Wang, Seeking a path through the crowd: Robot navigation in unknown dynamic environments with moving obstacles based on an integrated environment representation, Robot. Auton. Syst. 62 (10) (2014) 1568–1580, http://dx.doi.org/10.1016/j.robot.2014.05.006.

[9] S. Guillén-Ruiz, J.P. Bandera, A. Hidalgo-Paniagua, A. Bandera, Evolution of socially-aware robot navigation, Electronics 12 (7) (2023) 1–28, http://dx.doi.org/10.3390/electronics12071570.

[10] P. Trautman, A. Krause, Unfreezing the robot: Navigation in dense, interacting crowds, in: IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS, 2010, pp. 797–803, http://dx.doi.org/10.1109/IROS.2010.5654369.

[11] A. Alahi, V. Ramanathan, K. Goel, A. Robicquet, A.A. Sadeghian, L. Fei-Fei, S. Savarese, Learning to predict human behavior in crowded scenes, in: V. Murino, M. Cristani, S. Shah, S. Savarese (Eds.), Group and Crowd Behavior for Computer Vision, Academic Press, 2017, pp. 183–207, http://dx.doi.org/10.1016/B978-0-12-809276-7.00011-4.

[12] P. Kothari, B. Sifringer, A. Alahi, Interpretable social anchors for human trajectory forecasting in crowds, in: IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, 2021, pp. 15551–15561, http://dx.doi.org/10.1109/CVPR46437.2021.01530.

[13] M.H. Dridi, List of parameters influencing the pedestrian movement and pedestrian database, Int. J. Soc. Sci. Stud. 3 (4) (2015) 94–106, http://dx.doi.org/10.11114/ijsss.v3i4.870.

[14] A. Kendon, Conducting Interaction: Patterns of Behavior in Focused Encounters, Cambridge University Press, Cambridge (NY), 1990.

[15] J. Schellinck, T. White, A review of attraction and repulsion models of aggregation: Methods, findings and a discussion of model validation, Ecol. Model. 222 (11) (2011) 1897–1911, http://dx.doi.org/10.1016/j.ecolmodel.2011.03.013.

[16] G. Kouskoulis, C. Antoniou, Systematic review of pedestrian simulation models with a focus on emergency situations, Transp. Res. Rec. 2604 (1) (2017) 111–119, http://dx.doi.org/10.3141/2604-14.

[17] J. Ondřej, J. Pettré, A.-H. Olivier, S. Donikian, A synthetic-vision based steering approach for crowd simulation, in: ACM SIGGRAPH, 2010, pp. 1–9, http://dx.doi.org/10.1145/1833349.1778860.

[18] M. Moussaïd, D. Helbing, G. Theraulaz, How simple rules determine pedestrian behavior and crowd disasters, Proc. Natl. Acad. Sci. 108 (17) (2011) 6884–6888, http://dx.doi.org/10.1073/pnas.1016507108.

[19] X. Shi, Z. Ye, N. Shiwakoti, Z. Li, A review of experimental studies on complex pedestrian movement behaviors, in: COTA International Conference of Transportation Professionals, CICTP, 2015, pp. 1081–1096, http://dx.doi.org/10.1061/9780784479292.101.

[20] A. Rasouli, Pedestrian simulation: A review, 2021, arXiv:2102.03289.

[21] H. Klüpfel, H. Timmermans, Crowd dynamics phenomena, methodology, and simulation, in: H. Timmermans (Ed.), Pedestrian Behavior, Emerald Group Publishing Limited, 2009, pp. 215–244, http://dx.doi.org/10.1108/9781848557512-010.

[22] M. Bierlaire, T. Robin, Pedestrians choices, in: H. Timmermans (Ed.), Pedestrian Behavior, Emerald Group Publishing Limited, 2009, pp. 1–26, http://dx.doi.org/10.1108/9781848557512-001.

[23] U. Chattaraj, A. Seyfried, P. Chakroborty, Comparison of pedestrian fundamental diagram across cultures, Adv. Complex Syst. 12 (03) (2009) 393–405, http://dx.doi.org/10.1142/S0219525909002209.

[24] M. Moussaid, S. Garnier, G. Theraulaz, D. Helbing, Collective information processing and pattern formation in swarms, flocks, and crowds, Top. Cogn. Sci. 1 (3) (2009) 469–497, http://dx.doi.org/10.1111/j.1756-8765.2009.01028.

[25] M.-A. Granié, A. Abou-Dumontier, L. Guého, How gender influences road user behaviors: The bringing-in of developmental social psychology, in: N.A. Stanton (Ed.), Advances in Human Aspects of Road and Rail Transportation, CRC Press, 2012, pp. 754–763, http://dx.doi.org/10.1201/b12320.

[26] W. Inoue, T. Ikezoe, T. Tsuboyama, I. Sato, K.B. Malinowska, T. Kawaguchi, Y. Tabara, T. Nakayama, F. Matsuda, N. Ichihashi, Are there different factors affecting walking speed and gait cycle variability between men and women in community-dwelling older adults? Aging Clin. Exp. Res. 29 (2) (2017) 215–221, http://dx.doi.org/10.1007/s40520-016-0568-8.

[27] D. De Bartolo, M. Iosa, The walking brain: Factors influencing human gait, EC Psychol. Psychiatry 7 (12) (2018) 960–963, URL https://ecronicon.net/assets/ecpp/pdf/ECPP-07-00374.pdf.

[28] J. Wang, M. Boltes, A. Seyfried, A. Tordeux, J. Zhang, W. Weng, Experimental study on age and gender differences in microscopic movement characteristics of students, Chin. Phys. B 30 (9) (2021) 1–16, http://dx.doi.org/10.1088/1674-1056/ac11d4.

[29] A. Tom, M.-A. Granié, Gender differences in pedestrian rule compliance and visual search at signalized and unsignalized crossroads, Accid. Anal. Prev. 43 (5) (2011) 1794–1801, http://dx.doi.org/10.1016/j.aap.2011.04.012.

[30] R.L. Knoblauch, M.T. Pietrucha, M. Nitzburg, Field studies of pedestrian walking speed and start-up time, Transp. Res. Rec. 1538 (1) (1996) 27–38, http://dx.doi.org/10.1177/0361198196153800104.

[31] M. Schimpl, C. Lederer, M. Daumer, Development and validation of a new method to measure walking speed in free-living environments using the actibelt® platform, PLoS One 6 (8) (2011) 1–12, http://dx.doi.org/10.1371/journal.pone.0023080.

[32] W. Daamen, P.H.L. Bovy, S.P. Hoogendoorn, A. van de Reijt, Passenger route choice concerning level changes in railway stations, in: 84th Transportation Research Board Annual Meeting, 2005, pp. 1–18.

[33] T. Zheng, W. Qu, Y. Ge, X. Sun, K. Zhang, The joint effect of personality traits and perceived stress on pedestrian behavior in a chinese sample, PLoS One 12 (11) (2017) 1–18, http://dx.doi.org/10.1371/journal.pone.0188153.

[34] C. Stangor, Social Groups in Action and Interaction, Psychology Press, New York, 2004.

[35] A. Nicolas, F.H. Hassan, Social groups in pedestrian crowds: review of their influence on the dynamics and their modelling, Transportmetrica A: Transp. Sci. 19 (1) (2023) 1970651, http://dx.doi.org/10.1080/23249935.2021.1970651.

[36] G. Vizzari, L. Manenti, L. Crociani, Adaptive pedestrian behaviour for the preservation of group cohesion, Complex Adapt. Syst. Model. 1 (7) (2013) 1–29, http://dx.doi.org/10.1186/2194-3206-1-7.

[37] M. He, Z.-Q. Han, H.-N. Yu, D. Fan, Pedestrian simulation model considering groups dynamic pattern with communication, J. Transp. Syst. Eng. Inf. Technol. 17 (2) (2017) 136–141, http://dx.doi.org/10.16097/j.cnki.1009-6744.2017.02.020.

[38] Y. Hu, J. Zhang, W. Song, N.W.F. Bode, Social groups barely change the speed-density relationship in unidirectional pedestrian flow, but affect operational behaviours, Saf. Sci. 139 (2021) 1–12, http://dx.doi.org/10.1016/j.ssci.2021.105259.

[39] M. Moussaïd, N. Perozo, S. Garnier, D. Helbing, G. Theraulaz, The walking behaviour of pedestrian social groups and its impact on crowd dynamics, PLoS One 5 (4) (2010) 1–7, http://dx.doi.org/10.1371/journal.pone.0010047.

[40] M. Costa, Interpersonal distances in group walking, J. Nonverb. Behav. 34 (1) (2010) 15–26, http://dx.doi.org/10.1007/s10919-009-0077-y.

[41] A. Gorrini, S. Bandini, M. Sarvi, Group dynamics in pedestrian crowds: Estimating proxemic behavior, Transp. Res. Rec. 2421 (1) (2014) 51–56, http://dx.doi.org/10.3141/2421-06.

[42] W. Ge, R.T. Collins, R.B. Ruback, Vision-based analysis of small groups in pedestrian crowds, IEEE Trans. Pattern Anal. Mach. Intell. 34 (5) (2012) 1003–1016, http://dx.doi.org/10.1109/TPAMI.2011.176.

[43] M. Schultz, L. Rößger, H. Fricke, B. Schlag, Group dynamic behavior and psychometric profiles as substantial driver for pedestrian dynamics, in: Pedestrian and Evacuation Dynamics 2012, 2014, pp. 1097–1111, http://dx.doi.org/10.1007/978-3-319-02447-9_90.

[44] F. Zanlungo, T. Ikeda, T. Kanda, Potential for the dynamics of pedestrians in a socially interacting group, Phys. Rev. E 89 (1) (2014) 1–18, http://dx.doi.org/10.1103/PhysRevE.89.012811.

[45] D. Helbing, L. Buzna, A. Johansson, T. Werner, Self-organized pedestrian crowd dynamics: Experiments, simulations, and design solutions, Transp. Sci. 39 (1) (2005) 1–24, http://dx.doi.org/10.1287/trsc.1040.0108.

[46] A. Lerner, Y. Chrysanthou, D. Lischinski, Crowds by example, Comput. Graph. Forum 26 (3) (2007) 655–664, http://dx.doi.org/10.1111/j.1467-8659.2007.01089.x.

[47] S. Pellegrini, A. Ess, K. Schindler, L. van Gool, You'll never walk alone: Modeling social behavior for multi-target tracking, in: IEEE International Conference on Computer Vision, ICCV, 2009, pp. 261–268, http://dx.doi.org/10.1109/ICCV.2009.5459260.

[48] A. Robicquet, A. Sadeghian, A. Alahi, S. Savarese, Learning social etiquette: Human trajectory understanding in crowded scenes, in: European Conference on Computer Vision, ECCV, 2016, pp. 549–565, http://dx.doi.org/10.1007/978-3-319-46484-8_33.

[49] B. Majecka, Statistical Models of Pedestrian Behaviour in the Forum (MSc Dissertation), School of Informatics, University of Edinburgh, 2009, URL https://api.semanticscholar.org/CorpusID:15616262.

[50] J. Ferryman, A. Shahrokni, Pets2009: Dataset and challenge, in: IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS-Winter), 2009, pp. 1–6, http://dx.doi.org/10.1109/PETS-WINTER.2009.5399556.

[51] D. Yang, L. Li, K. Redmill, U. Özgüner, Top-view trajectories: A pedestrian dataset of vehicle-crowd interaction from controlled experiments and crowded campus, in: IEEE Intelligent Vehicles Symposium, IV, 2019, pp. 899–904, http://dx.doi.org/10.1109/IVS.2019.8814092.

[52] D. Brščić, T. Kanda, T. Ikeda, T. Miyashita, Person tracking in large public spaces using 3-d range sensors, IEEE Trans. Hum.–Mach. Syst. 43 (6) (2013) 522–534, http://dx.doi.org/10.1109/THMS.2013.2283945.

[53] K. Chen, C.C. Loy, S. Gong, T. Xiang, Feature mining for localised crowd counting, in: British Machine Vision Conference, BMVC, 2012, pp. 1–11, http://dx.doi.org/10.5244/C.26.21.

[54] B. Zhou, X. Wang, X. Tang, Understanding collective crowd behaviors: Learning a mixture model of dynamic pedestrian-agents, in: IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 2871–2878, http://dx.doi.org/10.1109/CVPR.2012.6248013.

[55] S. Yi, H. Li, X. Wang, Pedestrian behavior modeling from stationary crowds with applications to intelligent surveillance, IEEE Trans. Image Process. 25 (9) (2016) 4354–4368, http://dx.doi.org/10.1109/TIP.2016.2590322.

[56] L. Leal-Taixé, A. Milan, I. Reid, S. Roth, K. Schindler, Motchallenge 2015: Towards a benchmark for multi-target tracking, 2015, arXiv:1504.01942.

[57] A. Milan, L. Leal-Taixe, I. Reid, S. Roth, K. Schindler, Mot16: A benchmark for multi-object tracking, 2016, arXiv:1603.00831.

[58] P. Dendorfer, H. Rezatofighi, A. Milan, J. Shi, D. Cremers, I. Reid, S. Roth, K. Schindler, L. Leal-Taixé, Mot20: A benchmark for multi object tracking in crowded scenes, 2020, arXiv:2003.09003.

[59] S. Oh, A. Hoogs, A. Perera, N. Cuntoor, C.-C. Chen, J.T. Lee, S. Mukherjee, J.K. Aggarwal, H. Lee, L. Davis, E. Swears, X. Wang, Q. Ji, K. Reddy, M. Shah, C. Vondrick, H. Pirsiavash, D. Ramanan, J. Yuen, A. Torralba, B. Song, A. Fong, A. Roy-Chowdhury, M. Desai, A large-scale benchmark dataset for event recognition in surveillance video, in: IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2011, pp. 3153–3160, http://dx.doi.org/10.1109/CVPR.2011.5995586.

[60] J. Amirian, B. Zhang, F.V. Castro, J.J. Baldelomar, J.-B. Hayet, J. Pettré, Opentraj: Assessing prediction complexity in human trajectories datasets, in: 2020 Asian Conference on Computer Vision, ACCV, 2021, pp. 566–582, http://dx.doi.org/10.1007/978-3-030-69544-6_34.

[61] A. Seyfried, O. Passon, B. Steffen, M. Boltes, T. Rupprecht, W. Klingsch, New insights into pedestrian flow through bottlenecks, Transp. Sci. 43 (3) (2009) 395–406, http://dx.doi.org/10.1287/trsc.1090.0263.

[62] A. Rudenko, T.P. Kucner, C.S. Swaminathan, R.T. Chadalavada, K.O. Arras, A.J. Lilienthal, Thör: Human–robot navigation data collection and accurate motion trajectories dataset, IEEE Robot. Autom. Lett. 5 (2) (2020) 676–682, http://dx.doi.org/10.1109/LRA.2020.2965416.

[63] A. Rudenko, L. Palmieri, M. Herman, K.M. Kitani, D.M. Gavrila, K.O. Arras, Human motion trajectory prediction: a survey, Int. J. Robot. Res. 39 (8) (2020) 895–935, http://dx.doi.org/10.1177/0278364920917446.

[64] P. Kothari, S. Kreiss, A. Alahi, Human trajectory forecasting in crowds: A deep learning perspective, IEEE Trans. Intell. Transp. Syst. 23 (7) (2022) 7386–7400, http://dx.doi.org/10.1109/TITS.2021.3069362.

[65] A. Sieben, J. Schumann, A. Seyfried, Collective phenomena in crowds—where pedestrian dynamics need social psychology, PLoS One 12 (6) (2017) 1–19, http://dx.doi.org/10.1371/journal.pone.0177328.

[66] C. Appert-Rolland, J. Cividini, H.J. Hilhorst, P. Degond, Pedestrian flows: From individuals to crowds, Transp. Res. Procedia 2 (2014) 468–476, http://dx.doi.org/10.1016/j.trpro.2014.09.062.

[67] T. Osaragi, Modeling of pedestrian behavior and its applications to spatial evaluation, in: Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS, 2004, pp. 836–843, URL https://ieeexplore.ieee.org/document/1373600.

[68] M. Moussaïd, J.D. Nelson, Simple heuristics and the modelling of crowd behaviours, in: Pedestrian and Evacuation Dynamics 2012, 2014, pp. 75–90, http://dx.doi.org/10.1007/978-3-319-02447-9_5.

[69] Y. Peng, G. Zhang, J. Shi, B. Xu, L. Zheng, Srai-lstm: A social relation attention-based interaction-aware lstm for human trajectory prediction, Neurocomputing 490 (2022) 258–268, http://dx.doi.org/10.1016/j.neucom.2021.11.089.

[70] T. Do, M. Haghani, M. Sarvi, Group and single pedestrian behavior in crowd dynamics, Transp. Res. Rec. 2540 (1) (2016) 13–19, http://dx.doi.org/10.3141/2540-02.

[71] S. Yi, H. Li, X. Wang, Understanding pedestrian behaviors from stationary crowd groups, in: IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2015, pp. 3488–3496, http://dx.doi.org/10.1109/CVPR.2015.7298971.

[72] W.H. Warren, Collective motion in human crowds, Curr. Dir. Psychol. Sci. 27 (4) (2018) 232–240, http://dx.doi.org/10.1177/0963721417746743.

[73] J. Bian, D. Tian, Y. Tang, D. Tao, Trajectory data classification: A review, ACM Trans. Intell. Syst. Technol. 10 (4) (2019) 1–34, http://dx.doi.org/10.1145/3330138.

[74] A. Stergiou, R. Poppe, Analyzing human–human interactions: A survey, Comput. Vis. Image Underst. 188 (2019) 1–20, http://dx.doi.org/10.1016/j.cviu.2019.102799.

[75] D. Xu, Y. Yan, E. Ricci, N. Sebe, Detecting anomalous events in videos by learning deep representations of appearance and motion, Comput. Vis. Image Underst. 156 (2017) 117–127, http://dx.doi.org/10.1016/j.cviu.2016.10.010.

[76] A.M. Kanu-Asiegbu, R. Vasudevan, X. Du, Leveraging trajectory prediction for pedestrian video anomaly detection, in: IEEE Symposium Series on Computational Intelligence, SSCI, 2021, pp. 01–08, http://dx.doi.org/10.1109/SSCI50451.2021.9660004.

[77] F. Martinez-Gil, M. Lozano, I. García-Fernández, P. Romero, D. Serra, R. Sebastián, Using inverse reinforcement learning with real trajectories to get more trustworthy pedestrian simulations, Mathematics 8 (9) (2020) http://dx.doi.org/10.3390/math8091479.

[78] G. Di Gennaro, A. Buonanno, F. Verolla, G. Fioretti, F.A.N. Palmieri, K.R. Pattipati, Imitation learning through prior injection in markov decision processes, in: A. Esposito, M. Faundez-Zanuy, F.C. Morabito, E. Pasero (Eds.), Applications of Artificial Intelligence and Neural Systems To Data Science, Springer, Singapore, 2023, pp. 103–113, http://dx.doi.org/10.1007/978-981-99-3592-5_10.

[79] G. Di Gennaro, A. Buonanno, G. Fioretti, F. Verolla, K.R. Pattipati, F.A.N. Palmieri, Probabilistic inference and dynamic programming: A unified approach to multi-agent autonomous coordination in complex and uncertain environments, Front. Phys. 10 (2022) 1–12, http://dx.doi.org/10.3389/fphy.2022.944157.

[80] G. Di Gennaro, A. Buonanno, F.A.N. Palmieri, K.R. Pattipati, M. Merola, Path planning of multiple agents through probability flow, in: IEEE 33rd International Workshop on Machine Learning for Signal Processing, MLSP, 2023, pp. 1–6, http://dx.doi.org/10.1109/MLSP55844.2023.10285946.

[81] T. Salzmann, B. Ivanovic, P. Chakravarty, M. Pavone, Trajectron++: Dynamically-feasible trajectory forecasting with heterogeneous data, in: 2020 European Conference on Computer Vision, ECCV, 2020, pp. 683–700, http://dx.doi.org/10.1007/978-3-030-58523-5_40.

[82] M. Gulzar, Y. Muhammad, N. Muhammad, A survey on motion prediction of pedestrians and vehicles for autonomous driving, IEEE Access 9 (2021) 137957–137969, http://dx.doi.org/10.1109/ACCESS.2021.3118224.

[83] Y. Zhang, X. Shen, P. Raksincharoensak, Study on collision avoidance strategies based on social force model considering stochastic motion of pedestrians in mixed traffic scenario, J. Robot. Mechatronics 35 (2) (2023) 240–254, http://dx.doi.org/10.20965/jrm.2023.p0240.

[84] D. Yang, U. Özgüner, Combining social force model with model predictive control for vehicle's longitudinal speed regulation in pedestrian-dense scenarios, in: 8th Biennial Workshop on Digital Signal Processing for in-Vehicle Systems, 2018, pp. 1–8, URL https://dongfang-steven-yang.github.io/files/2018-DSPinVehicles.pdf.

[85] R.G.N. Lakmali, A.A.B.D.P. Abewardhana, P.V. Genovese, Pedestrian movement tracking and tracing in public space, in: 13th KDU International Research Conference, KDUIRC, 2020, pp. 32–44, URL http://ir.kdu.ac.lk/handle/345/3064.