

Perception of Emotions from Crowd Dynamics

M.W. Baig^{*†}, Mirza Sulman Baig^{*}, V. Bastani^{*}, E.I. Barakova[†], L. Marcenaro^{*}, C. S. Regazzoni^{*} and M. Rauterberg[†]

^{*}Department of Electrical, Electronic, Telecommunications Engineering and Naval Architecture - DITEN
University of Genova, Genova, Italy.

[†]Department of Industrial Design, Eindhoven University of Technology
Eindhoven, Netherlands.

[†]Email:(M.W.Baig, E.I.Barakova, G.W.M.Rauterberg)@tue.nl

^{*}Email:(Waqar, Sulman.Baig, Vahid.Bastani, Lucio.Marcenaro, Carlo)@ginevra.dibe.unige.it

Abstract—Perceiving crowd emotions and understand the situation is vital to control the situations in surveillance applications. This paper introduces the evolution of methods for crowd emotion perception based on bio-inspired probabilistic models. The emotions have been perceived both in an offline and online manner from the crowd. We focus on the perception of emotion from crowd behavior and dynamics. The paper explains few probabilistic algorithms and compares these for detection of emotion of crowds and proposes a probabilistic modelling approach which is trained on data to perceive the emotions of the crowd in an area under surveillance. Emotions are defined as evolving dynamic patterns arising due to interaction of people in an environment with their relationships to the past interaction patterns. Camera sensors are used to track the motion of the individuals within a crowd scenario under observation. The data mining techniques are used to distinguish between different behaviors and events into positive and negative emotions. The results have been evaluated using simulated data from a proposed office environment.

Keywords-Crowd emotion detection, Event based Dynamic Bayesian Network, Log-likelihood ratio, Crowd simulation, Autobiographical Memory, Instantaneous Topological Map

I. INTRODUCTION

The research on emotion is an area of emerging interest due to two main reasons. First, to get the insights of human nature evolution and second, for security, control and management purposes. The latter one is much important due to the reasons of security threat analysis and for well being of individuals and to prevent disasters. Dangerous situations such as stampede, one of which occurred in Multan, Pakistan [1] during political rally last year and one at Shanghai, China [2] during new year celebrations this year, urges the need of online emotion estimation of the crowd. This should be done in a way, that emotions changes are detected in runtime and actions can be done by experienced crowd controllers to control the situations.

During surveillance situations, there are a number of displays covering different cameras and crowd controllers and these have to be diligent to monitor situations at every display, making a map of the whole situation. This should be done by combining live video instants sequentially coming from displays and his/her situation analyzing experience. If these displays are not intelligent enough to detect the emotion changes in their viewpoints, and alert the crowd controllers/managers, then crowd control situations will be more actively

monitored and controlled.

The detection of emotions of crowd is an emerging area of interest. Previously, research has been carried out in single person emotion detection and classification such as facial emotion expressions detection systems using the distinctive facial features and calibrating them into emotions [3] [4] of individuals. Facial Action Coding System (FACS) [5] uses set of action units (AUs) that model emotions on independent motion of face. Voice expressions have also been used to detect emotions of individuals. The audio-based affect recognition using the features of pitch and context has been developed such as [6] [7]. The body language and posture is also one of the most useful feature in detection of emotion of people. The predefined body language patterns and postures are interpreted by the activity recognizers. The transition of body patterns over a specific interval of time, posture patterns to describe the interest towards some specific events are quite useful in detecting emotions [8] [9]. Physiology based affect detection is quite broad approach, which requires the inclusion of delicate and calibrated sensors on body [10] and in some cases quite strict supervision, as in case of brain imaging and electroencephalography (EEG) signals [11]. The physiology based methods are quite accurate but they are usually used to calibrate the ground truth for emotion and affect detection. They cannot be used in real life situations. Apart from these techniques, there are also multi-modal techniques which fuse many features such as face, voice, and posture etc., the data coming from different sensors is fused to make a decision of emotion [12] [13]. The brief overview presented so far, is used for individual persons emotion detection. These systems are not being employed in crowd emotion detection due to computational and feasibility limitations.

The detection of crowd emotion, where the crowd is a single entity is a challenging task. There has been a lot of research on crowd behavior management. The crowd emotion detection was not extensively investigated before due to computationally expensive implementation issues. The main shortcomings of detection of emotions in a crowd is that we have to depend on one sensor such as video cameras; we cannot use delicate and detailed features such as face emotion features [5], activity features [9] and voice features [7]. The processing speed and viewpoint differences are main constraints in computation in crowd and this gets even worse during abnormal situations.

Therefore, the features available for crowd emotion detection are quite crude, but using the knowledge from psychology, learning the context and history, we are able to develop a model that can detect the emotions of the crowd. Gustave Le Bon [14] described that crowd acts like a collective mind and acts and reacts. This phenomenon becomes more prominent when some abnormal situation arises. Therefore, our evaluations mostly correspond to the accuracy due to features we have selected.

The main contribution of this paper is to discuss the existing models of crowd emotion detection and their comparison with a proposed online crowd emotion estimation model. This model is used to detect the emotions of the crowd in an online manner. The rest of the paper is organized as follows. Section II describes crowd emotion detection algorithms and Section III explains the result and comparisons of algorithms and Section IV concludes the paper along with some future directions.

II. CROWD EMOTION RECOGNITION

Crowd emotion detection algorithm based on behaviors is presented [15]. The evolution of this algorithm is based on the following steps. First, the framework is based on the collecting the local features from crowd and using Damasio [16] findings to make these observations in causal relationships based on dynamic Bayesian networks. These causal relationships are then being clustered using Self-Organized maps (SOM) [17] to define events. Using the Ortony, Clore and Collins (OCC) model [18] to define causes of emotions for learning and classification and for calibrating the model Russell et al. circumplex model of affect [19] is being used. In first algorithm, the surveillance area is partitioned using Instantaneous Topological Mapping (ITM) [20] into different zones. All possible interactions that occur in the chosen area among people and environment and their reactions due to these actions are memorized using a bio-inspired approach. These interactions are generated on the basis of cause and effect relationships that happen between people and people with environment. The interactions can be learned using autobiographical memory as proposed by Damasio [16]. To describe the effect of interacting quantities internal and external states. The two states proto (internal) and core (mirror of external) states are defined. The proto and core observations are defined as proto (entity) and core (crowd). This makes a chain of temporally and spatially aligned proto \vec{x}_p and core state \vec{x}_c vectors. The proto and core events can be defined as ϵ_p and ϵ_c using the probabilistic model to develop autobiographical memory (AM). The triplets of events for passive and active interactions are $\{\epsilon_p^-, \epsilon_c, \epsilon_p^+\}$ and $\{\epsilon_c^-, \epsilon_p, \epsilon_c^+\}$ respectively. These represents the causal relationship in terms of initial situation (first event $\epsilon_{p,c}^-$), the cause (second event $\epsilon_{c,p}^+$), and consequent effect of the examined entity (third event $\epsilon_{p,c}^+$).

The Dynamic Bayesian Networks (DBN) are used to model the interaction. The conditional probability densities (CPD) $p(\epsilon_t^p | \epsilon_{t-1}^p)$ and $p(\epsilon_t^c | \epsilon_{t-1}^c)$ simulates the motion tracking and

crowdedness patterns. The interactions among the two interacting objects can be modelled using conditional probability densities (CPDs):

$$p(\epsilon_t^p | \epsilon_{t-\Delta t^c}^c) \quad (1)$$

$$p(\epsilon_t^c | \epsilon_{t-\Delta t^p}^p) \quad (2)$$

The Eq. 1 represents the probability that events ϵ_c occurred at time $t - \Delta t^c$, by the interacting object which is related to the core context and vice versa in case of Eq. 2. The casual relationships between the two interacting objects are modelled using two conditional probabilities (CPDs):

$$p(\epsilon_t^p | \epsilon_{t-\Delta t^c}^c, \epsilon_{t-\Delta t^p}^p) \quad (3)$$

$$p(\epsilon_t^c | \epsilon_{t-\Delta t^p}^p, \epsilon_{t-\Delta t^c}^c) \quad (4)$$

The probability densities in Eq. 3 and Eq. 4 consider interaction of Eq. 1 and Eq. 2 as well as initial situation $\epsilon_{t-\Delta t^c}^c$ and $\epsilon_{t-\Delta t^p}^p$. A self organized map is used to convert these conditional interactions into super-states to reduce the feature space. Therefore the events generated by these super-states are $\{\theta_p^-, \theta_c, \theta_p^+\}$. These triplets are saved for each positive and negative emotion events in the autobiographical memory. The Gaussian mixture model (GMM) is used to encode the temporal evolution of events. In particular, the Figueiredo and Jain [21] GMM model is used. The CPDs of the model of emotions are given by:

$$p(\Delta t^p | \theta^{p(j,k)}, \theta^{c(m,n)}, \theta^{p(i,j)}) = \sum_{N_k}^{i=1} \pi_i \mathcal{N}(\Delta t^p | \mu_i, \sum_i \theta_n) \quad (5)$$

$$p(\Delta t^c | \theta^{c(m,n)}, \theta^{p(j,k)}, \theta^{c(n,p)}) = \sum_{N_k}^{i=1} \pi_i \mathcal{N}(\Delta t^c | \mu_i, \sum_i \theta_n) \quad (6)$$

where N_k number of modes in GMM. The emotions are recognized as an accumulative measure of online emotion classification. Whenever an external observed event is detected θ_{obs}^c , the proto CPD map is analysed which was the internal event θ_w^p . The AM is then examined to describe the internal event, its belonging to the emotion class w , which could be positive or negative. The probability distribution of detected event estimated by GMM and the one from the map is evaluated for matching by Hellinger Distance [29]. The Hellinger distance is evaluated for each detected behavior triplet.

In second algorithm [22], we detected the events belonging to positive and negative emotions, without time dependence. The conditional CPDs are same as previous. But these CPDs triplets saved in AM are developed for different kind of known behavioral events. The triplet pattern for AM for trajectory behavior sequence is given by

$$W_N = \{w_1, \dots, w_n\} \quad (7)$$

The training dataset consists of N triplet vectors

$$U = \{W_1, \dots, W_n, \dots, W_N\} \quad (8)$$

W_N is the total behavioral patterns gathered. In order to gather the natural clustering of training trajectories on which the

model should be constructed, we need to cluster the data into classes. As number of clusters is unknown, we use the DBN model to cluster the data into behavioral patterns using affinity matrix approach. To calculate the affinity between two trajectory sequences, W_a and W_b , the D_a and D_b two DBNs, are trained using expectation maximization (EM) algorithm [23]. The affinity between W_a and W_b is:

$$Q_{ab} = \frac{1}{2} \left\{ \frac{1}{s_a} \log p(W_a|D_b) + \frac{1}{s_b} \log p(W_b|D_a) \right\} \quad (9)$$

(6) where $p(W_a|D_b)$ is the likelihood of perceiving W_a given D_b , s_a and s_b are the lengths of W_a and W_b respectively. A $N \times N$ infinity matrix is obtained $Q = [Q_{ab}]$, where $a, b \leq N$. This makes a new representation of dataset U as

$$Z_q = \{q_{k1}, \dots, q_{kn}, q_{kN}\} \quad (10)$$

where k is the cluster or class label. Each behavioural pattern is modelled by a feature vector which is dynamically warped by DBNs. This defines the scene/ situation based clusters in the model. Considering the different clusters, if we have only sparse data, then it will be difficult to fit the model. The training set Z_q in this case consists of K clusters. The distribution of behavioral patterns is modeled based on Gaussian mixture model (GMM). Given the K -th mixture component of GMM, the log likelihood of observing the training dataset Z_q is defined as:

$$\log p(Z_q|\partial) = \sum_{n=1}^N \log \left(\sum_{K=1}^K p(q_{kn}|\partial_k) \right) \quad (11)$$

where $p(q_{kn}|\partial_k)$ is the Gaussian distribution of K -th mixture component. The parameters ∂ of the model are evaluated using EM algorithm. For training initialization of positive E_p and negative emotion E_n models, the labelled dataset is used. The positive emotion E_p model and negative emotion model E_n are trained separately using labelled data. An accumulative measure of online emotion classification is proposed in this task. Whenever an external observed behavior Q_{new} is detected, the probability distribution estimated by GMM of detected behavior and the positive and negative emotion models is evaluated by using the likelihood ratio test (LRT).

$$\psi(Q_{new}) = \frac{p(Q_{new}|E_p)}{p(Q_{new}|E_n)} \begin{cases} \geq \alpha_{th} : \text{PositiveEmotion} \\ \leq \alpha_{th} : \text{NegativeEmotion} \end{cases} \quad (12)$$

where $p(Q_{new}|E_p)$ and $p(Q_{new}|E_n)$ are calculated using equations, where α_{th1} and α_{th2} are the thresholds that are being tuned based on the model and false positive rate of detection of emotions. Using this procedure, several responses are calculated for time T_c and the decision is made with maximum response of people with positive and negative emotions based on detection of emotion of people detected in maximum in an area under observation.

In the third algorithm, we propose the emotions of the individuals are evaluated in an area under surveillance with an online estimation yielded by event based Dynamic Bayesian Network based on motions dynamics of the people. The

environment is divided into smaller local areas, in which the model and inferences are carried out. The ITM is not used in it to discretize the zones. The probabilistic models for both positive and negative emotions are trained using big data of trajectories collected corresponding to two different emotions. The positive emotions are mostly based on motivation for reaching the destination using parameters such as position, time etc. We have defined two information layers. One low-level takes the information of tracking of individuals in crowd using Kalman filter and upper level infers this tracking information into high-level, which is based on zone transitions and time spent in a zone.

The surveillance area is divided in N mutually exclusive and exhaustive partitions $\{\mathcal{P}_n\}_{n=1}^N$, as zones. Note that the temporal scale between tracker level sequence and transition level sequence is different.

An Event-Based Dynamic Bayesian Network (E-DBN) used for analyzing emotion of individuals. The variable $r \in \{-1, +1\}$ is an indicator variable representing the emotion of the person either be positive emotion $r = +1$ or negative emotion $r = -1$. The E-DBN is characterized by two conditional probability distribution function

$$\Pi_{z_k, z_{k-1}}^r = p(z_k|z_{k-1}, r) \quad (13)$$

and

$$\Omega_{\Delta_k, z_k, z_{k-1}}^r = p(\Delta_k|z_k, z_{k-1}, r) \quad (14)$$

and prior probabilities $p(r)$ and $p(z_0|r)$. Applying Bayes rule, the joint Probability Density Function (PDF) of the DBN can be calculated

$$p(r, z_{0:k}, \Delta_{1:k}) = p(r)p(z_0|r) \prod_{i=1}^k p(z_i|z_{i-1}, r)p(\Delta_i|z_i, z_{i-1}, r) \quad (15)$$

In this paper, the prior probabilities $p(r)$ and $p(z_0|r)$ are assumed to be uniform densities. However, two model parameters $\Pi_{z_k, z_{k-1}}^r$ and $\Omega_{\Delta_k, z_k, z_{k-1}}^r$ are learned from historical data in real life situation or simulated scenarios.

The zone transition probabilities $\Pi_{z_k, z_{k-1}}^r$ are two square matrices of $N \times N$ dimension for $r = +1$ and $r = -1$. For learning the elements of this matrix, in the training data labeled by r we count n_i^r the number of time a trajectory enters to zone i and $n_{i,j}^r$ the number of time a transition occurred from zone i to j , then

$$\Pi_{i,j}^r = \frac{n_{i,j}^r}{n_i^r} \quad (16)$$

which is the optimum estimate of this parameter in maximum data likelihood sense.

In order to detect the emotional state of an individual from its trajectory $x_{0:t}$ we proposed a log-likelihood ratio test:

$$\eta = \log \frac{\gamma(t, r = +1)}{\gamma(t, r = -1)} \quad (17)$$

where $\gamma(t, r = +1)$ and $\gamma(t, r = -1)$ are the likelihood of the sequence $s_{0:t}$ given positive and negative trajectory models

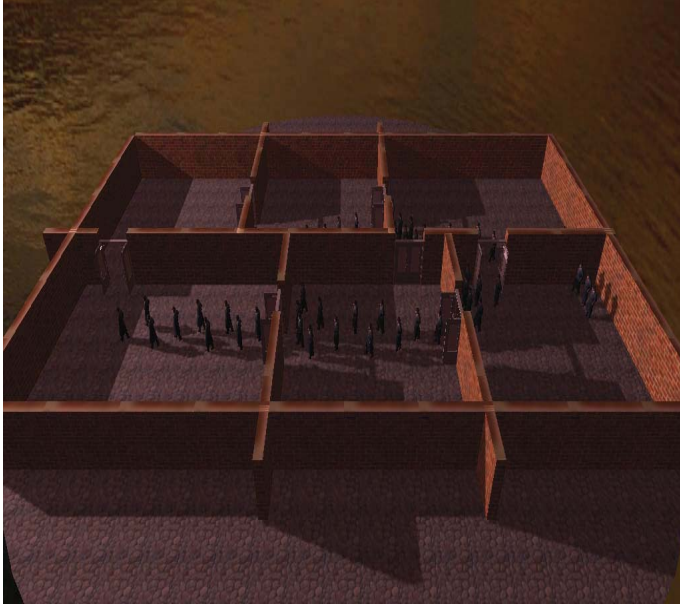


Fig. 1: Scenario of an office environment in crowd simulator.

respectively and they are recursively calculated as new position of the person is received from the tracker. The variable η in Eq.17 is positive when the data is more likely to come from a trajectory with positive emotion and vice versa. In addition the magnitude of this value shows how much close is the trajectory to positive or negative model.

III. EXPERIMENTAL RESULTS AND DISCUSSION

To test and validate the proposed model of emotions, we developed a realistic crowd behavior simulator and the Realistic Behavioral Agent Simulation (RBAS) [24]. In crowd situations, it is difficult to know how individuals will move and pass through a confined space and how the buildings and pathways should be built such that people passing through them do not have a situation of congestion and jamming and they pass easily without injuries or other crowd induced problems. It is cumbersome and dangerous to bring a large number of people along and to perform the experiments; therefore it is feasible to make the realistic simulation and making the model and then collecting a real user data based on the results from that model. The Figure 1 shows a snapshot of the developed simulation model. The motivation of the agents in the simulator environment is set in such a way that the simulated humans have to pass through the office area, from one side to another. During simulations, the agents experience different kinds of interactions, as we covered every aspect and made learning useful by including every possible interaction. For training purposes, we used the data obtained for simulation with different configurations of people and events for 3000 minutes. The trajectories that reach the destination in an average fixed simulation time $T_{max} = 360sec$ are considered as trajectories of positive emotions, while the other are considered as negative emotion trajectories. The testing is performed using multiple test trajectories of positive and

TABLE I: Confusion Matrix of Emotion recognition concluded at the end of each trajectory for algorithm 1

		Prediction outcome	
		p	n
p	p	TP 90.23%	FP 9.77%
	n	FN 14.88%	TN 85.12%

TABLE II: Confusion Matrix of Emotion recognition concluded at the end of each trajectory for algorithm 2

		Prediction outcome	
		p	n
p	p	TP 92.01%	FP 7.99%
	n	FN 11.64%	TN 88.36%

negative emotions. The testing is performed using 30 test trajectories.

The three algorithms have been tested and positive and negative trajectories have been evaluated. The confusion matrices of these algorithms show that all the three algorithms perform well for evaluating trajectories at the end of simulation as in Tables I, II, III. The main difference is evident in evaluations among these algorithms. When test trajectories, both positive and negative ones, have been evaluated with these three algorithms, the proposed third algorithm evaluates the emotion well than the other two algorithms. The Table. IV shows the emotion recognized by these algorithms at each time stamp and it clearly shows that online estimation of crowd emotion in third algorithm performs well than other two algorithms.

IV. CONCLUSION AND FUTURE WORK

This paper presents the overview of crowd emotion detection techniques and evaluates the proposed online crowd emotion detection algorithm with other developed crowd emotion algorithms. The testing shows that algorithms for positive and negative emotions perform well for any kind of interactions and scenarios. The proposed online emotion

TABLE III: Confusion Matrix of Emotion recognition concluded at the end of each trajectory for algorithm 3

		Prediction outcome	
		p	n
p	p	TP 94.54%	FP 5.46%
	n	FN 10.37%	TN 89.63%

TABLE IV: Emotion Recognized at each time stamp/Total time stamps

Emotion(Positive and Negative)	Algorithm 1	Algorithm 2	Algorithm 3
Time Percentage	56	64	79

detection algorithm detects the emotions of the individuals with a significant timing difference than other two algorithms. The future work includes the designing of experiments to include complex interactions and motivations. Therefore, different kind of situations can happen and people can have different motivations. Hence making a complex probabilistic model for such situations is needed.

ACKNOWLEDGEMENTS

This work was supported in part by the Erasmus Mundus Joint Doctorate in Interactive and Cognitive Environments, which is funded by the EACEA Agency of the European Commission under EMJD ICE FPA n 2010-0012

REFERENCES

- [1] *Stampede at political procession in Multan, Pakistan*, <http://tribune.com.pk/story/773657/pti-multan-administration-trade-blame-after-8-people-killed-in-qasim-bagh-stampede/>.
- [2] *Stampede at new year celebrations in Shanghai, China*, <http://www.bbc.com/news/world-asia-china-30646918>.
- [3] Paul Ekman. An argument for basic emotions. *Cognition & emotion*, 6(3-4):169–200, 1992.
- [4] Paul Ekman. Expression and the nature of emotion. *Approaches to emotion*, 3:19–344, 1984.
- [5] Paul Ekman and Wallace V Friesen. Facial action coding system. 1977.
- [6] Patrik N Juslin and Klaus R Scherer. *Vocal expression of affect*. Oxford University Press, 2005.
- [7] Klaus R Scherer. Vocal communication of emotion: A review of research paradigms. *Speech communication*, 40(1):227–256, 2003.
- [8] Sidney K DMello, Scotty D Craig, Amy Witherspoon, Bethany McDaniel, and Arthur Graesser. Automatic detection of learners affect from conversational cues. *User modeling and user-adapted interaction*, 18(1-2):45–80, 2008.
- [9] Selene Mota and Rosalind W Picard. Automated posture analysis for detecting learner's interest level. In *Computer Vision and Pattern Recognition Workshop, 2003. CVPRW'03. Conference on*, volume 5, pages 49–49. IEEE, 2003.
- [10] Emilia I Barakova and Tino Lourens. Expressing and interpreting emotional movements in social games with robots. *Personal and ubiquitous computing*, 14(5):457–467, 2010.
- [11] Olivier Villon and Christine Lisetti. A user-modeling approach to build user's psycho-physiological maps of emotions using bio-sensors. In *Robot and Human Interactive Communication, 2006. ROMAN 2006. The 15th IEEE International Symposium on*, pages 269–276. IEEE, 2006.
- [12] Jaak Panksepp. *Affective neuroscience: The foundations of human and animal emotions*. Oxford university press, 1998.
- [13] Jaak Panksepp. Emotions as natural kinds within the mammalian brain. *Handbook of emotions*, 2:137–156, 2000.
- [14] Ashish Kapoor and Rosalind W Picard. Multimodal affect recognition in learning environments. In *Proceedings of the 13th annual ACM international conference on Multimedia*, pages 677–682. ACM, 2005.
- [15] Mirza Waqar Baig, Emilia I Barakova, Lucio Marcenaro, Carlo S Regazzoni, and Matthias Rauterberg. Bio-inspired probabilistic model for crowd emotion detection. In *Neural Networks (IJCNN), 2014 International Joint Conference on*, pages 3966–3973. IEEE, 2014.
- [16] Antonio Damasio and Raymond J Dolan. The feeling of what happens. *Nature*, 401(6756):847–847, 1999.
- [17] Teuvo Kohonen. *Self-organizing maps*, volume 30. Springer Science & Business Media, 2001.
- [18] Andrew Ortony. *The cognitive structure of emotions*. Cambridge university press, 1990.
- [19] James A Russell, Maria Lewicka, and Toomas Niit. A cross-cultural study of a circumplex model of affect. *Journal of personality and social psychology*, 57(5):848, 1989.
- [20] Jan Jockusch and Helge Ritter. An instantaneous topological mapping model for correlated stimuli. In *Neural Networks, 1999. IJCNN'99. International Joint Conference on*, volume 1, pages 529–534. IEEE, 1999.
- [21] Mario AT Figueiredo and Anil K. Jain. Unsupervised learning of finite mixture models. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(3):381–396, 2002.
- [22] Mirza Waqar Baig, Emilia I Barakova, Lucio Marcenaro, Matthias Rauterberg, and Carlo S Regazzoni. Crowd emotion detection using dynamic probabilistic models. In *From Animals to Animats 13*, pages 328–337. Springer, 2014.
- [23] Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society. Series B (methodological)*, pages 1–38, 1977.
- [24] Mirza Waqar Baig, Emilia Barakova, Carlo S Regazzoni, and Matthias Rauterberg. Realistic modeling of agents in crowd simulations. In *Proceedings of the fifth International Conference on Intelligent Systems, Modelling and Simulation, IEEE*, pages 507–512, 2014.