Crowd Emotion Detection
Using Dynamic Probabilistic Models

Mirza Waqar Baig[1,2], Emilia I. Barakova[1] , Lucio Marcenaro[2],
Matthias Rauterberg[1], and Carlo S. Regazzoni[2]

[1] Department of Industrial Design, Eindhoven University of Technology,
Eindhoven, Netherlands
[2] Department of Electronics Engineering, University of Genoa, Italy
`{M.W.Baig,E.I.Barakova,G.W.M.Rauterberg}@tue.nl,`
`{Waqar,Lucio.Marcenaro,Carlo}@ginevra.unige.it`

Abstract. Detecting emotions of a crowd to control the situation is an area of emerging interest. The purpose of this paper is to present a novel idea to detect the emotions of the crowd. Emotions are defined as evolving quantities arising from the reaction to contextual situations in a set of dynamic pattern of events. These events depend on internal and external interaction states in an already mapped space. The emotions of multiple people constituting a crowd in any surveillance environment are estimated by their social and collective behaviors using sensor signals e.g., a camera, which captures and tracks their motion. The feature space is constructed based on local features to model the contextual situations and the different interactions corresponding to different emergent behaviors are modeled using bio-inspired dynamic model. The changes in emotions correspond to behavioral changes which are produced to regulate behaviors under different encountered situations. Proposed algorithm involves the probabilistic signal processing modelling techniques for analysis of different types of collective behaviors based on interactions among people and classification models to estimate emotions as positive or negative. The evaluations are performed on simulated data show the proposed algorithm effectively recognizes the emotions of the crowd under specific situations.

Keywords: Crowd Emotions Detection, Collective Behavior, Dynamic Bayesian Networks, Autobiographical Memory, Dynamic event modeling.

1   Introduction

Emotion is a feeling evoked by environmental stimuli or by internal body states [1], which modulate human behavior in terms of actions in the environment or changes in the internal status of the body. The idea of emotional expressions as scientific subject was first recognized by Darwin, who found out that the natural selection process is not only applicable to anatomic patterns but also can be applied to animal's behaviors and mind. The detection of collective emotions of a crowd is a new research area that has not been explored systematically so far. Research has been carried out in single

person emotion detection and classification such as facial expressions systems using the distinctive facial features and calibrating them into emotions [2] of individuals. 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 [3]. The body language and posture is also one of most useful feature in detection of emotion of people. The body language and posture guessed by the activity recognizers [4] [5]. Physiology based affect detection is quite broad approach, which requires the inclusion of delicate and calibrated sensors on body [6]. The physiology based methods are quite accurate but they require inclusion of delicate and calibrated sensors on body [7] and cannot be used in real life situations. Apart from these techniques, there are also multimodal techniques which fuse many features such as face, voice, and posture etc., the data coming from different sensors is fused to make a decision on type of emotion [8]. All the compact review mentioned above is in use for individual emotion detection, these systems cannot be employed in crowd emotion detection due to richness of features obtained with timing constraints of evaluation and calibration. The detection of emotions of the crowd is useful in crowd behavior management for safety and security and also in robot responses towards emotions of crowd.

The main contribution of this paper is to develop the first model to detect the emotions of the crowd. This paper presents the dynamic probabilistic modelling of behaviors that are being generated from autobiographical memory (AM). The AM is developed on modelling of events which got it concept from psychological theories. The evaluations have been performed for training and testing purposes. The paper explains the Emotion detection and classification algorithm in Section 2. The Section 3 consists of evaluations and results, Section 4 is conclusions, Section 5 consists of acknowledgement and Section 6 consists of references.

2   Crowd Emotion Computation

The proposed framework is based on Ortony, Clore and Collins (OCC) [17] theory of emotion. The estimation of emotions is based on Russell et al. [16] circumplex model of affect.   The local features such as motion tracking and density of crowd are used to make the possible causal relationship based features space. This feature space after reduction is used to generate emergent possible behaviours in an unsupervised way. These behaviours are then trained into different models according to context and situation, which are used to detect emotions. The algorithm flow first shows the topology based partitioning of observation space and then using the local features to make event based memory. Using this event based memory model possible behaviours which are then used to estimate emotions of the crowd.

2.1   Topological Partitioning

In order to extract the effective information from the crowd, we need to discretize the space under observation into smaller acceptable areas, in which the rich data information should be preserved. The topology representing networks (TRNs) algorithms are possible solution to divide the observation space into discrete areas. The instantaneous

topological map (ITM) is employed to divide the observation space into zones as shown in Figure 1. In [9], it has been demonstrated that ITM is an appropriate solution to discretize the zones of correlated trajectories.
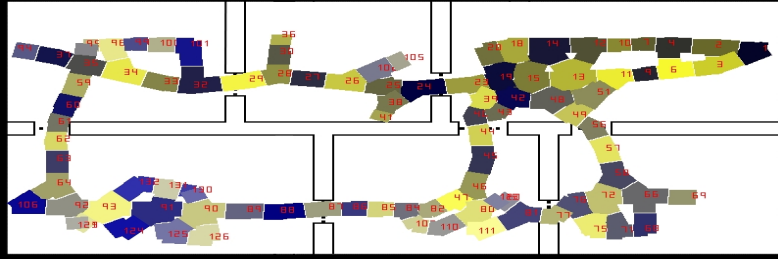
Fig. 1. With $\tau = 500$, τ defines the minimum distance between a given zone given all the observations for creating new zone

2.2   Modelling Events

The ability to predict the emotions requires understanding of actions in an area under surveillance in order to define the interactions and activities that represent emotions. To reach this goal, a probable technique is to memorize all possible interactions that occur in related area among people and environment and their reactions due to these actions. A bio-inspired technique is used which has been mentioned in [10][11][12][13]. These interactions are generated due to causal relationships that happen among persons and environment. These types of interactions can be learned using AM as mentioned by Damasio [14]. The two states proto (internal) and core (mirror of external) state based 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 AM. The triplets of events for passive and active interactions are $\{\varepsilon_p^-, \varepsilon_c, \varepsilon_p^+\}$ and $\{\varepsilon_c^-, \varepsilon_p, \varepsilon_c^+\}$ respectively. These represents the causal relationship in terms of initial situation (first event $\varepsilon_{p,c}^-$), the cause (second event $\varepsilon_{c,p}$), and consequent effect of the examined entity (third event $\varepsilon_{p,c}^+$). This sequence of these states can be decided by using a probabilistic graphical model that describes the relationships among them using statistical and mathematical similarities among interactions. The interactions consist of temporal sequences of interdependent states described by two probability distributions: $p(\varepsilon_p^+|\varepsilon_c, \varepsilon_p^-)$ and $p(\varepsilon_c^+|\varepsilon_p, \varepsilon_c^-)$. The Dynamic Bayesian networks (DBNs) are used to model the interaction. The conditional probability densities (CPD) $p(\varepsilon_t^p|\varepsilon_{t-1}^p)$ and $p(\varepsilon_t^c|\varepsilon_{t-1}^c)$ simulates the motion tracking and crowdedness patterns. The interactions among the two interacting objects can be modelled using conditional probability densities (CPDs):

$$p(\varepsilon_t^p|\varepsilon_{t-\Delta t^c}^c); p(\varepsilon_t^c|\varepsilon_{t-\Delta t^p}^p) \qquad (1)$$

The (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. The casual relationships

between the two interacting objects are modelled using two conditional probabilities (CPDs):

$$p\left(\varepsilon_t^p\middle|\varepsilon_{t-\Delta t}^c{}^c, \varepsilon_{t-\Delta t}^p{}^p\right); p\left(\varepsilon_t^c\middle|\varepsilon_{t-\Delta t}^p{}^p, \varepsilon_{t-\Delta t}^c{}^c\right) \tag{2}$$

The probability densities in (2) consider interaction of (1) as well as initial situation $\varepsilon_{t-\Delta t}^c{}^c$ and $\varepsilon_t^p{}_{-\Delta t}^p$. The observations associated with proto and core are ε_t^p and $\varepsilon_t^c = [d_t]$. The proto and core states are $\vec{x}_p(t)$ and $\vec{x}_c(t)$. In order to represent these elements with respect to contextual information, a clustering technique of dimensionality reduction such as Self organized map (SOM) [15], unsupervised classifier is used to convert multidimensional proto $\vec{x}_p(t)$ and core vectors $\vec{x}_c(t)$ into low dimension W-Z, where W is the dimension of map (layer). The clusterization process, maps proto and core states into 2-D vectors, which correspond to the neurons of SOM map. These are called core super states S_{x_c} and proto Super states S_{x_p}. SOM allows clustering the proto $\vec{x}_p(t)$ and core vectors $\vec{x}_c(t)$ into corresponding neuron map as super states, which are then known as proto super-state S_{x_p} and core super-state S_{x_c}. The parameter M is to be tuned. The labels associated are given by:

$$S_{x_p}^i \mapsto l_p^i, \quad i = 1, \dots, n_p; \ S_{x_c}^j \mapsto l_c^j, \quad j = 1, \dots, n_c \tag{3}$$

where $S_{x_p}^i$ and $S_{x_c}^j$ are the ith and jth super states, and n_p and n_c are total number of proto and core states generated during mapping. The proto and core super-states are labelled with the semantic labels from the ITM zones. The event is defined as when a proto or core super-state changes its zone. This gives rise to AM associated with each zone. Using the SOM representation, it is possible to detect changes in the map through super states, where super-states are connected by local features for particular instances. This representation encompasses the changes in state vectors $\vec{x}_p(t)$ and $\vec{x}_c(t)$ for every time instatnt as movements in map. If changes in state vectors $\vec{x}_p(t)$ and $\vec{x}_c(t)$ do not imply changes in super-state labels $S_{x_p}^i \mapsto l_p^i$ and $S_{x_c}^j \mapsto l_c^j$, then the SOM mapping needs to be recalibrated as the semantics defined are not the correct   representation of events. When super states $S_{x_p}^i$ and $S_{x_c}^j$ change during specific time instants, their contextual modification entails an event. Therefore, an event is defined as: $\theta_e^t = l_{p,c}^{i,j}(t-1) \mapsto l_{p,c}^{i,j}(t)$, where $i, j = 1, \dots, n_{p,c}$, with timing constraints T_{max}. There are also null events (null changes in super states $i = j$, can be defined as $\theta_e^t = \emptyset_e$. This gives rise to AM memory model, this is modelled by the events in which learning the changes from proto super-states to core-super-states and subsequent modification of core super-state is memorized. The following three types of events can be memorized. (i) $\theta_p^- = S_{x_p^0} \rightarrow S_{x_p^-}$ is the proto event at initial time instant. This represents the alteration of the proto super-state into $S_{x_p^0} \mapsto l_p^i$ to $S_{x_p^-} \mapsto l_p^j$ that can happen before the core event. The event θ_p^- also remembers the time window T_{max}^-. The two labels l_p^i and l_p^j and are related with super states $S_{x_p^0}$ and $S_{x_p^-}$. The event θ_p^- also remembers the time window T_{max}^-. (ii) $\theta_c = S_{x_c^-} \rightarrow S_{x_c^+}$

is the core event. This shows the alteration of external super-state into $S_{x_c^-} \mapsto l_c^m$ to $S_{x_c^+} \mapsto l_c^n$. (iii) $\theta_p^+ = S_{x_p^-} \rightarrow S_{x_p^+}$: is the proto event following the core event. It is also associated with the super states as $S_{x_p^-} \mapsto l_p^j$ to $S_{x_p^+} \mapsto l_p^k$. The following triplet $W = \{\theta_p^-, \theta_c, \theta_p^+\}$ represents the self-abstraction, which is related with AM. The AM represents the core conscious in Damasio work.

2.3   Emotion Modelling in Crowd

Emotion modelling in crowd is based on OCC theory [16], which states that the emotions are valanced reactions to events, agents and objects. These events, agents and objects can be either pleasing or displeasing to the subject under consideration. We use AM events which are already coupled by probability density of observations from agents and events. To scale the emotions into positive and negative, we use the Russell et. al. [17] two dimensional map of mental space of emotions. On this map, we only use valance coordinate to define emotions into positive and negative emotions. In order to cluster AM event stream triplets $\{\theta_p^-, \theta_c, \theta_p^+\}$ which are based on trajectory behavioral patterns, which is later used to classify into emotions. The triplet pattern for AM for trajectory behavior sequence is given by

$$W_N = \{w_1, \dots w_n\} \tag{4}$$

The training dataset consists of N triplet vectors

$$U = \{W_1, \dots. W_n, \dots. W_N\} \tag{5}$$

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 modelling 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 [18]. The affinity between W_a and W_b is:

$$Q_{ab} = \frac{1}{2}\{\frac{1}{s_a}\log p(W_a|D_b) + \frac{1}{s_b}\log p(W_b|D_a)\} \tag{6}$$

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 $N \, X \, N$ infinity matrix is obtained $Q = [Q_{ab}]$, where $a, b \leq N$. This makes a new representation of dataset U as

$$Z_q = \{q_{k1}, \dots, q_{kn}, \dots. q_{kN}\} \tag{7}$$

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:

$$\log p(Z_q|\partial) = \sum_{n=1}^{N}(\log \sum_{k=1}^{K} p(q_{kn}|\partial_k)) \tag{8}$$

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. To design optimal number of classes for the mixture model, Bayesian information criterion [19]. For any given K, the BIC is given by:

$$BIC = -\log p(Q|\partial) + \frac{F_K}{2}\log N \tag{9}$$

where F_K is the number of parameters required to model K-th component of GMM. Now N numbers of behavioral patterns are labelled to one of the behavioural clustered class. To define positive and negative emotional behaviors, we need to have labelled dataset for positive and negative emotions.
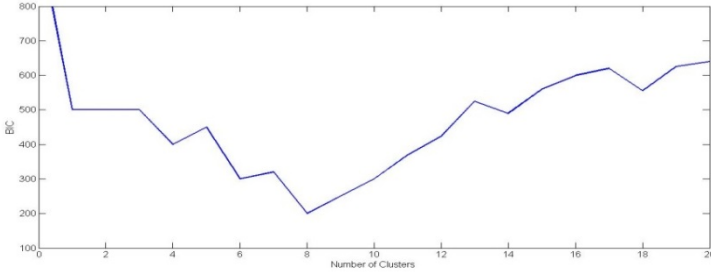
BIC versus cluster selection

Fig. 2. BIC model selection for optimal number of clusters of model, this BIC is modelled for negative emotion model E_n for affinity matrix of order N=60

For training initialization of positive E_p and negative emotion E_n models, the labelled dataset is used. The positive emotion E_p model is given for any behaviour O as:

$$p(O|E_p) = \sum_{q=1}^{Q_p} \delta_{pq}p(O|D_{pq}) \tag{10}$$

where δ_{pq} is the mixing probability of qth mixture component with $\partial_{pq} = \sum_{q=1}^{Q_p} \delta_{pq} = 1$, and D_{pq} is the qth DBN corresponding to qth normal behavior. Similarly for negative emotions, the behavior **O** belonging to negative emotions model E_n is given by:

$$p(O|E_n) = \sum_{q=1}^{Q_n} \delta_{nq}p(O|D_{nq}) \tag{11}$$

The models for both positive and negative emotions situations 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) [22].

$$\psi(Q_{new}) = \frac{p(Q_{new}|E_p)}{p(Q_{new}|E_n)} \begin{cases} \geq \propto_{th} & Positive\ Emotion \\ < \propto_{th} & Negative\ Emotion \end{cases} \tag{12}$$

where $p(Q_{new}|E_p)$ and $p(Q_{new}|E_n)$ are calculated using equations, where $\propto_{th1}$ and $\propto_{th2}$ 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.

3   Evaluation and Results

To evaluate the proposed computational model of crowd emotions, we design a simulation model in which we can produce the evacuation and panic situations. People are presented as agents that move under different behaviors and a realistic crowd behaviors and RBAS (Realistic Behavioral Agent Simulation) model is proposed [20]. The algorithm is based on social force model SFM [21], simulates the agents as ellipse with particular sizes. The agents have sense of environment and plan their own path to avoid collisions. The shaking and repelling effect in agents have been reduced using the body contact and sliding forces. The interactions among agents have been modeled based on personal reaction space. These changes in SFM model makes the model more realistic as every parameter has been modeled based on findings from psychology and video tracking. The crowd simulation model has been developed on C++. To train the negative emotions model, the following situations produced, which are common to panic and evacuation events was produced a) Herding b) Faster is Slower c) Turbulent flow d) Stop and go waves. The events have been captured and different kinds of behaviors have been detected during these flows according to the context such as passing through doors during high density, changing the motivation etc. The positive and negative emotion situations are labelled and normal behaviors are also labelled and two models for both positive and negative emotions have been trained. The labelled data for emotions produces the following clustered global negative behavior situations. The commonly observed behavior cluster classes are clustered at doors of rooms a) C_1 to C_3 b) C_2 to C_3 c) C_1 to C_4 d) C_1 to C_3 e) C_2 to C_4 f) C_3 to C_5 g) C_5 to C_6 h) C_4 to C_6, where C_1 to C_6 are the rooms in the simulation. There are also many other classes which are representation of sudden events. The optimal cluster selection is done using the BIC as shown in figure 2. The log likelihood ratio is plotted against frames which are taken at random. The emotions detected are then followed by comparison with the labelled data and it shows that detection ratio follows the actual labels as shown in figure 4.

To evaluate the performance of the algorithm, ROC (receiver operating curves) are obtained, these are obtained by 50 different trails of simulations with two different thresholds of likelihood ratios for both models of positive and negative emotions. The ROC curves evaluations in figure 3 and table 1 show the classification accuracy is quite acceptable.

4   Conclusion

A situation based emotion detection and classification technique has been proposed. Using the models for positive and negative emotions, the method utilizes a dynamic

probabilistic clustering technique to model responses in different situations. The context based information is modeled based on the bio-inspired autobiographical memory. The likelihood ratio test used to classify emotions in crowds has been demonstrated. To our knowledge, the proposed model is the first one in the literature to present algorithm on detection of emotions of the crowd and results have shown that the model is capable of detecting the emotions of the people from individual level to collective level with high accuracy. This system could be useful in surveillance environment to predict the crowd situations before they happen. The future work includes the collection of datasets with the help of experts and making general context information for every kind of situation.

Table 1. Results from ROC for positive emotion model E_p and negative emotion model E_n

| Parameters	Area under ROC curve for positive emotion model E_p	Area under ROC curve for negative emotion model E_n
$\propto_{th1}$	0.886	0.858
$\propto_{th2}$	0.845	0.812

Table 1. Results from ROC for positive emotion model E_p and negative emotion model E_n
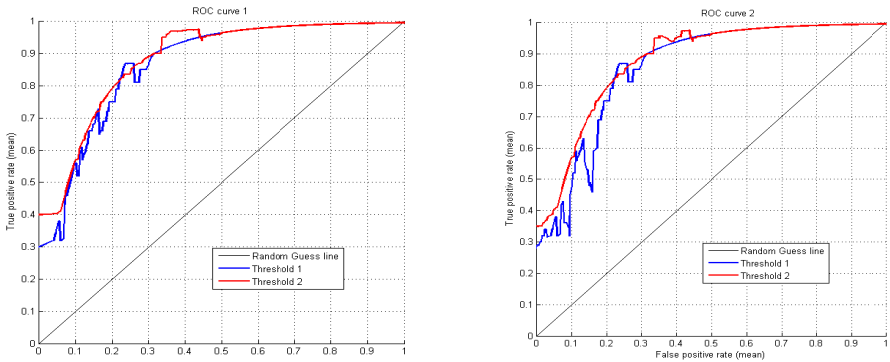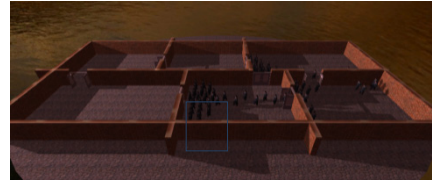

(a)   ROC curve for positive emotion model E_p (b) ROC curve for negative emotion model E_n

Fig. 3. ROC curves obtained using 50 different experimental scenarios and using two different thresholds as in equation with $\propto_{th1} = -0.25$ and $\propto_{th2} = -0.30$. (a) Positive Emotion model E_p (b) Negative emotion model E_n.
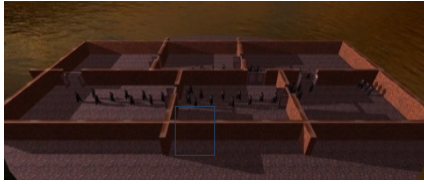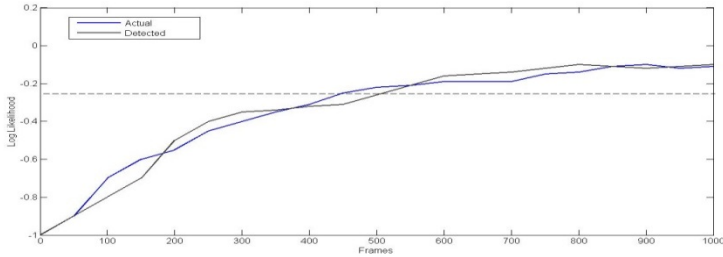
(a) Positive Emotions Detected of the crowd at (350th frame)

(b) Herding Situation (Negative Emotions Detected) (850th frame)

(c) Comparison of Log likelihood ratio and Frames

Fig. 4. Comparison of Log likelihood ratio and emotion deduced in different frames (a) Normal situation depicts positive emotion. The blue box represents the group of people whose emotions are averaged to give the result. (b) Herding behavior produced at C_3 to C_5 and negative emotions were detected on average in blue box. (c) Comparison of Log likelihood ratio and video frames using $\propto_{th} = -0.25$. The detected curve is by our model and actual curve is obtained by comparison of labelled data. The blue box shows the area under observation.

References

1. Bower, G.H., Cohen, P.R.: Emotional influences in memory and thinking: Data and theory. In: Affect and Cognition, vol. 1 (1982)
2. Ekman, P.: Expression and the nature of emotion. Approaches to Emotion 3, 19–344 (1984)
3. Juslin, P.N., Scherer, K.R.: Vocal expression of affect. The new handbook of Methods in Nonverbal Behavior Research, 65–135 (2005)
4. Scherer, K.R.: Vocal communication of emotion: A review of research paradigms. Speech Communication 40, 227–256 (2003)
5. Mota, S., Picard, R.W.: Automated posture analysis for detecting learner's interest level. In: 2003 Conference on Computer Vision and Pattern Recognition Workshop, CVPRW 2003, vol. 5. IEEE (2003)

6. Villon, O., Lisetti, C.: A user-modeling approach to build user's psycho-physiological maps of emotions using bio-sensors. In: The 15th IEEE International Symposium on Robot and Human Interactive Communication, ROMAN 2006 (2006)

7. Villon, O., Lisetti, C.: A user-modeling approach to build user's psycho-physiological maps of emotions using bio-sensors. In: The 15th IEEE International Symposium on Robot and Human Interactive Communication, ROMAN 2006, pp. 269–276. IEEE (2006)

8. Bänziger, T., Grandjean, D., Scherer, K.R.: Emotion recognition from expressions in face, voice, and body: The Multimodal Emotion Recognition Test (MERT). Emotion 9(5), 691 (2009)

9. Jockusch, J., Ritter, H.: An instantaneous topological mapping model for correlated stimuli. In: International Joint Conference on. Neural Networks, IJCNN 1999, vol. 1. IEEE (1999)

10. Chiappino, S., et al.: Event based switched dynamic bayesian networks for autonomous cognitive crowd monitoring. In: Wide Area Surveillance, pp. 93–122. Springer, Heidelberg (2014)

11. Dore, A., et al.: A bio-inspired system model for interactive surveillance applications. Journal of Ambient Intelligence and Smart Environments 3(2), 147–163 (2011)

12. Dore, A., Regazzoni, C.S.: Interaction Analysis with a Bayesian Trajectory Model. IEEE Intelligent Systems 25(3), 32–40 (2010)

13. Dore, A., Cattoni, A.F., Regazzoni, C.S.: Interaction modeling and prediction in smart spaces: a bio-inspired approach based on autobiographical memory. IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans 40(6), 1191–1205 (2010)

14. Damasio, A.R.: The feeling of what happens: Body and emotion in the making of consciousness. EBook (1999)

15. Kohonen, T.: Self-organizing maps, vol. 30. Springer (2001)

16. Ortony, A.: The cognitive structure of emotions. Cambridge University Press (1990)

17. Russell, J.A., Lewicka, M., Niit, T.: A cross-cultural study of a circumplex model of affect. Journal of personality and social psychology 57(5), 848 (1989)

18. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal statistical Society 39(1), 1–38 (1977)

19. Schwarz, G.: Estimating the dimension of a model. The annals of statistics 6(2), 461–464 (1978)

20. Baig, M.W., Barakova, E., Regazzoni, C.S., Rauterberg, M.: Realistic Modeling of Agents in Crowd Simulations. In: 2014 5th International Conference on Intelligent Systems Modelling & Simulation (ISMS), pp. 507–512. IEEE (2014)

21. Helbing, D., Molnar, P.: Social force model for pedestrian dynamics. Physical review E 51(5), 4282 (1995)

22. Neyman, J., Pearson, E.S.: On the problem of the most efficient tests of statistical hypotheses. Springer, New York (1992)