MODELING CROWDS AS SINGLE-MINDED ENTITIES

Oscar J. Urizar[1,2], *Emilia I. Barakova*[2], *Carlo S. Regazzoni*[1], *Matthias Rauterberg*[2]

[1]Department of Electrical, Electronic, Telecommunications Engineering and Naval Architecture (DITEN) University of Genoa - Genoa, Italy.

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

ABSTRACT

Inspired by Gustave Lebon's idea of crowds as single-minded entities, we present a novel approach to describe the behavior of a crowd as a single entity, based on the global movement of the entire aggregate of people conforming the crowd. The present work significantly differs from existing literature where the behavior of single individuals within the crowd are the building blocks to describe crowd behavior. A bi-dimensional neural gas network is implemented to learn the topology of the physical environment in an unsupervised fashion, then a self-organizing map and a Bayesian network are used to describe the behavior of the crowd as a single entity. Experiments were conducted using footage from New York Grand Central Station to test the accuracy of our model to learn and identify different behaviors of the crowd. Results show high accuracy to identify behaviors under usual circumstances and low but consistently increasing accuracy over time on less common cases.

Index Terms— Crowd Behavior, Crowd Modeling, Single-Entity Crowd Model

1. INTRODUCTION

The phenomenon of crowds has become commonplace in urban areas, as seen in massive audiences gathering to enjoy a concert or sport event, political manifestations, long queues to buy the newest smart-phone or simply as large aggregates of people in shopping malls or train stations. However, crowds are unstable and highly emotional, hence understanding the dynamics of crowds is essential to avoid dangerous situations as well as for crowd control and of particular interest for the development of dynamic cognitive systems intended for smart cities [1, 2, 3].

Previous approaches tend to focus on pedestrian tracking in order to model crowds and crowd behaviors [4, 5]. Unfortunately, performance of current techniques for tracking decreases as the number of pedestrian increases [6, 7]. On the other hand, algorithms intended for pedestrian detection are able to perform well even on high density scenarios [8, 9].

Our proposed approach relies on pedestrian detection techniques to model a crowd as a single entity, where the behaviors are described by changes in density distribution within the crowd. Hence our method is able to learn and identify behaviors with high accuracy regardless of changes in crowd density levels. Starting with input signals representing the position of individuals at a given time, we first obtain a topological representation of the environment partitioned into small regions using a Growing Neural Gas (GNG) [10]. The configuration of the crowd at a certain moment is described by the number of people in each region of the environment. The set of different configurations of the crowd are clustered into states using a Self-Organizing Map (SOM) [11]. Hence, the GNG clusters input signals to provide a topological representation enabling us to describe a crowd's configuration as a uniform observation state vector whereas the SOM clusters the observation state vectors into states to describe the crowd's dynamics as state transitions in a Markov process. Using a Dynamic Bayesian Network approach [12], a separate model is constructed for each learned behavior.

1.1. Literature Review

Multiple approaches have been proposed to model crowd behavior, either focusing on the individual as agent-based models or to a higher level of abstraction as in flow-based models. Suport Vector Machine (SVM) approaches have been widely adopted for the task of crowd behavior modeling as in [13] for static crowd behavior intended for crowd control. Also, the work in [14] presented a structural SVM learning method for detecting groups in crowded environments. A dynamic agent-based method was presented in [15] to model collective behavior patterns of individuals in crowded scenes. The authors of [16] made use of motion-flow

This work was partially supported 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.

models to analyze behavioral patterns in crowds using flow estimates for behavior recognition. An approach to cluster trajectories using GNG is presented in [17]. Similarly [18] used a GNG to provide a topological representation of the environment but made use of trajectories to identify major pedestrian flows, whereas our approach considers only density distribution changes to model the movement of the crowd which is advantageous when trajectories are fragmented due to high density levels. Our previous work in the subject as presented in [19] started to explore the idea of describing crowd behaviors using a Bayesian approach but it was limited to describe a crowd as a Markov process without formally modeling behaviors.

The rest of this work is organized as follow: Our proposed approach is described in section 2. Results and discussion of conducted experiments to validate our approach are shown in section 3. Finally, in section 4 we present our conclusions and intentions for future work.

2. METHODS

This method describes the behavior of a crowd by the changes in density distribution over time. First, the topology of the observed environment is learned with a GNG and divided into regions to have a formal description of the crowd's configuration (density distribution) at a given time. Second, we train a SOM to cluster similar configurations of the crowd to a given state, enabling us to describe the dynamics of a crowd by state transitions as a Markov process. Third, we train a separated Bayesian model for each behavior we desire to identify.

2.1. Topological Representation

A GNG is used to learn the observed environment's topology in an unsupervised fashion. The set of input signals employed to learn the topology is defined as

$$\mathbf{y}_{t:\tau} = \{\hat{y}_{1_t}, ..., \hat{y}_{N_t}\}_{t=1}^{\tau} \quad (1)$$

where $\hat{y}_{i_t} \in \mathbb{R}^2$ is the estimated position of person i at time t during an observation period from 1 to τ (total number of observations). The GNG is defined by a set A of nodes and a set M of edges. The structure of the GNG is described with a set of M unweighted edges. The GNG starts with two nodes initialized to random positions and nodes are added or removed by evaluating the input signals from $\mathbf{y}_{t:\tau}$ until the number of nodes converge to a maximum accumulated euclidean distance between a node and its associated input signals. Upon completion of the adaptation phase, the set $A = \{n_1, ..., n_Q\}$ provides a topological representation where n_k corresponds to a physical region in the environment. The use of the GNG is illustrated in Figure 2(d) by yellow circles representing a node n_k placed above its associated physical region. Once we have obtained the toplogical representation, a function g is defined to classify the set of all input signals at time t

$$g(\mathbf{y}_t) = \mathbf{z}_t \quad (2)$$

where $\mathbf{z}_t = \{r_{i_t}, ..., r_{Q_t}\}$ is the observation vector of the crowd at time t and $r_{i_t} \in \mathbb{R}$ contains the number of people in the region n_i at time t, for all Q regions depicted by the GNG. The size of $\mathbf{y}_t$ changes as people come and go but $\mathbf{z}_t$ remains with size Q at all time. Hence $\mathbf{z}_t$ provides a uniform description of the crowd's density distribution at a given time.

2.2. Dynamic Bayesian Network

In this work the dynamics of a crowd are described as a Markov process. A dynamic Bayesian network of type Hidden Markov Model (HMM), as shown in Figure 1, is implemented to learn different behaviors of a crowd. The hidden vector state $\mathbf{x}_t$ depicting the configuration of the crowd as a whole is estimated from the observation vector $\mathbf{z}_t$ defined in eq.2. A collection of estimated hidden vector states over a discrete period of observation, $\mathbf{x}_{t:\tau} = \{\hat{\mathbf{x}}_t, ...\hat{\mathbf{x}}_\tau\}$, is used to train a SOM comprised by a set S of neurons and a set E of edges. The SOM is configured in an hexagonal topology with p rows and q columns for a total of v neurons. Neuron's weights are initialized randomly and distance among neurons is measured by the number of edges between them. After completion of training, we can use SOM to classify any state vector estimation $\hat{\mathbf{x}}_t$ to a state (neuron) $s_k \in S$. At this point we are able to describe the dynamics of a crowd observed from input signals $\mathbf{y}_{t:\tau}$, converted to observation vectors $\mathbf{z}_{t:\tau}$ and estimated hidden vector states $\hat{\mathbf{x}}_{t:\tau}$ as a sequence of state transitions $\{s_a, ..., s_k; s_j \in S\}$.

2.3. Crowd Models

A separated model b_i is created for each crowd behaviour we intend to learn. For each model b_i we require a training set $\mathbf{s}^{b_i} = \{s_a, ..., s_k; s_j \in S\}$ obtained from input signals $\mathbf{y}_{t:\tau}$ displaying the crowd's behavior intended to model. We create a transition matrix $TRANS^{b_i}$ of dimensions $v \times v$, where v is the total number of states (neurons) in SOM. We employ $TRANS^{b_i}$ to learn the transitions as presented in $\mathbf{s}^{b_i}$. This procedure is repeated for each behavior. Once the transition matrices for all behaviors have been created, we employ the training sets $\{\mathbf{s}^{b_1}, ..., \mathbf{s}^{b_u}\}$ to produce the emission matrix $EMIS$ with dimensions of $v \times u$ where v is the

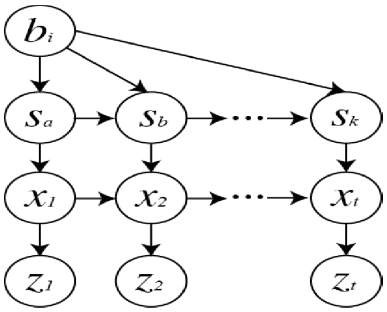

Fig. 1: Dynamic Bayesian network.

total number of states from SOM and u is the total number of modeled behaviors.

The dynamic bayesian network b_i is characterized with a conditional probability distribution function defined as

$$\Phi^{b_i}_{s_t,s_{t-1}} = e(b_i|s_t)P^{b_i}(s_t|s_{t-1}) \quad (3)$$

Where $e(b_i|s_t)$ is the emission probability of b_i and $P^{b_i}(s_t|s_{t-1})$ is the probability of state transition from s_{t-1} to s_t given from $TRANS^{b_i}$. Prior probabilities $e(b_i|s_0)$ and $P^{b_i}(s_0)$ are assumed to be uniform densities and learned from training data. In the above definition, the observation sequence is limited to previous (s_{t-1}) and current states (s_t), but this can be extended to a higher order m and calculate $P^{b_i}(s_t|s_{t-1},...,s_{t-m})$ by recursion.

3. EXPERIMENTS AND RESULTS

To validate our proposed model's capability to learn and identify different behaviors in a crowd we employ the Grand Central Station dataset [15]. This dataset provides the ground truth (manual annotations) of the observed trajectories for each individual and it is used as the input signals as defined in eq.1.

The GNG is trained with the following parameters: $input\ signals = 273,000$, $\lambda = 50$, $\epsilon_b = 0.2$, $\epsilon_n = 0.005$, $\alpha = 0.5$, $a_max = 20$, $d = 0.995$. A snapshot of the footage, plotting of input signals, the trained GNG and the partitioned environment are shown in Figure 2.

The SOM is configured in an hexagonal arrangement with 100 neurons (10 columns and 10 rows). Neurons weights are initialized randomly within the input space and the initial neighborhood size is 3 with 100 steps in the ordering phase. Training phase is performed over 500 epochs by competitive layer without bias and using 14,000 observation vector samples as input signals.

Employing the GNG and SOM, a total of seven training sets were prepared, one for each different behavior to be modeled. Each training set contains 2,000 observation vector samples. A behavior is characterized

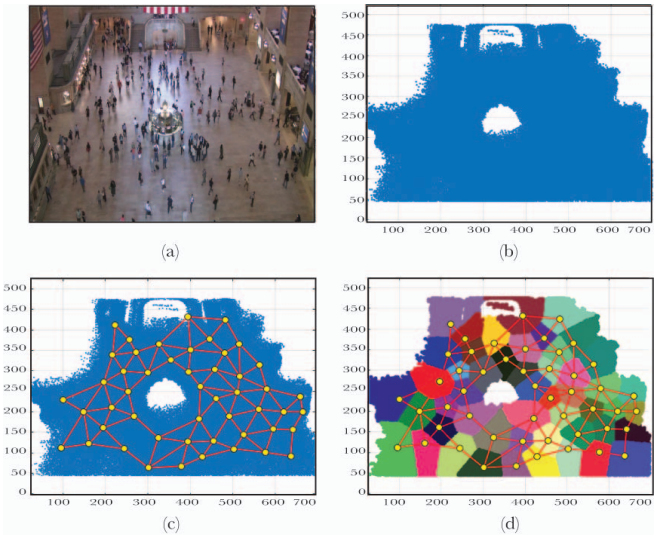

Fig. 2: a) Footage of New York Grand Central station. b) Sample points of observed trajectories. c) Topology learned with growing neural gas networks. d) Environment representation divided by regions.

by all detected individuals heading to the same direction, in the experiments presented here we employ 7 different directions: north-east, north, north-west, south-west, south, south-east, mixed directions. A plot of each behavior is presented in Figure 3.

To evaluate the above created models, 7 testing sets were prepared with 2,000 observation vector samples for each set. Additionally, three variations of each testing set were used for evaluation where 100%, 75% and 50% of people behave according to the testing set's intended behavior and the remainder percentage of people behave in different behaviors. The performance of the models capability to identify each behavior are recorded after different periods of observation: 1 observation, 10 observations, 20 observations, 30 observations. Results are shown in Table 1, separated into 3 sub-tables for each variation of the testing sets, by rows for each behavior and by columns for each length of observation period.

The results in Table 1(a) show a low accuracy to identify most behaviors after only one observation but consistently increase after more observations, this is reasonable as 100% of people in the testing set moving in the same behavior is a very unlikely scenario from the provided data. However, in Table 1(b) the results after just one observation are high as only 75% of people follow the detected behavior, which is a more common case. Results in Table 1(c) where only 50% of people follow the intended behavior are still high for most of the behaviors. A testing set with less than 50% of peo-

ple following the same behavior would imply that this is not the predominant behavior in the crowd.

Behaviour	e_i $P(S_t \mid S_{t-1})$	e_i $P(S_t \mid S_{t-10},...,S_{t-1})$	e_i $P(S_t \mid S_{t-20},...,S_{t-1})$	e_i $P(S_t \mid S_{t-30},...,S_{t-1})$
north-east	0.0560	0.3098	0.4416	0.5337
north	0.0815	0.2732	0.3891	0.4951
north-west	0.5423	0.6259	0.6477	0.6531
south-west	0.0325	0.2006	0.2860	0.3316
south	0.5055	0.5685	0.6446	0.7331
south-east	0.2890	0.4620	0.5689	0.6539
mixed	0.9600	0.9954	1.0000	1.0000

(a) Test set with 100% of trajectories moving according to each of the tested behaviour.

Behaviour	e_i $P(S_t \mid S_{t-1})$	e_i $P(S_t \mid S_{t-10},...,S_{t-1})$	e_i $P(S_t \mid S_{t-20},...,S_{t-1})$	e_i $P(S_t \mid S_{t-30},...,S_{t-1})$
north-east	0.9730	0.9984	1.0000	1.0000
north	0.9760	0.9989	1.0000	1.0000
north-west	0.9775	0.9969	1.0000	1.0000
south-west	0.9309	0.9914	1.0000	1.0000
south	0.9955	1.0000	1.0000	1.0000
south-east	0.9845	0.9849	0.9934	0.9969
mixed	0.8980	0.9869	1.0000	1.0000

(b) Test set with 75% of trajectories moving according to each of the tested behaviour.

Behaviour	e_i $P(S_t \mid S_{t-1})$	e_i $P(S_t \mid S_{t-10},...,S_{t-1})$	e_i $P(S_t \mid S_{t-20},...,S_{t-1})$	e_i $P(S_t \mid S_{t-30},...,S_{t-1})$
north-east	0.9870	0.9984	1.0000	1.0000
north	0.9760	0.9989	1.0000	1.0000
north-west	0.9775	0.9969	1.0000	1.0000
south-west	0.9309	0.9914	0.9994	1.0000
south	0.9901	1.0000	1.0000	1.0000
south-east	0.3310	0.3761	0.4406	0.4550
mixed	0.6620	0.6238	0.5593	0.5449

(c) Test set with 50% of trajectories moving according to each of the tested behaviour.

Table 1: Results of model's performance to identify learned behaviour after S_{t-1}, S_{t-10}, S_{t-20} and S_{t-30} observations. Three different test sets are employed to evaluate each behaviour as indicated in (a),(b) and (c).

A direct comparison between our and other related methods would be inaccurate since our method models the predominant behaviors of the crowd as a whole whereas existing methods model one or several behaviors based on partial features of the crowd. Additionally, a significant portion of literature that studies crowd behavior is focused on abnormality detection, which is not covered in the experiments presented here. It is also important to notice that the conducted experiments made use of the ground truth provided for this dataset, therefore the results presented do not account for the accuracy error added by the underlying algorithm employed for pedestrian detection.

4. CONCLUSIONS

The main contribution of this paper is a novel approach to learn and model the behaviors of a crowd as whole by a combination of existing methodologies, namely, Growing Neural Gas, Self-Organizing Maps and Bayesian Networks. Behaviors are modeled by changes in density distribution in the crowd rather than by observing single trajectories of individuals. This approach is advantageous when crowd density is high as in such circumstances people counting approaches perform better than pedestrian tracking techniques. Also, seen the crowd as a whole provides a novel and perhaps more comprehensive understanding of the crowd's dynamics. The experiments yielded high accuracy to identify different behaviors in crowds even when the predominant behavior is exhibit by as low as only half of the individuals in the crowd. Future work will include the use of this approach to estimate the emotional state of the crowd and to learn causalities in emotional state changes.

5. REFERENCES

[1] T. Franke, P. Lukowicz, and U. Blanke, "Smart crowds in smart cities: real life, city scale deployments of a smartphone based participatory crowd management platform," *Journal of Internet Services and Applications*, vol. 6, no. 1, pp. 1–19, 2015.

[2] G. Cardone, A. Cirri, A. Corradi, L. Foschini, R. Ianniello, and R. Montanari, "Crowdsensing in urban areas for city-scale mass gathering management: Geofencing and activity recognition," *IEEE Sensors Journal*, vol. 14, no. 12, pp. 4185–4195, Dec 2014.

[3] S. Chiappino, P. Morerio, L. Marcenaro, E. Fuiano, G. Repetto, and C. S. Regazzoni, "A multi-sensor cognitive approach for active security monitoring of abnormal overcrowding situations," in *15th International Conference on Information Fusion, FUSION 2012, Singapore, July 9-12, 2012*. 2012, pp. 2215–2222, IEEE.

[4] M. Rodriguez, I. Laptev, J. Sivic, and J.-Y. Audibert, "Density-aware person detection and tracking in crowds," in *Proceedings of the 2011 International Conference on Computer Vision*, Washington, DC, USA, 2011, ICCV '11, pp. 2423–2430, IEEE Computer Society.

[5] A. Bera and D. Manocha, "Realtime multilevel crowd tracking using reciprocal velocity obstacles," in *22nd International Conference on Pattern*

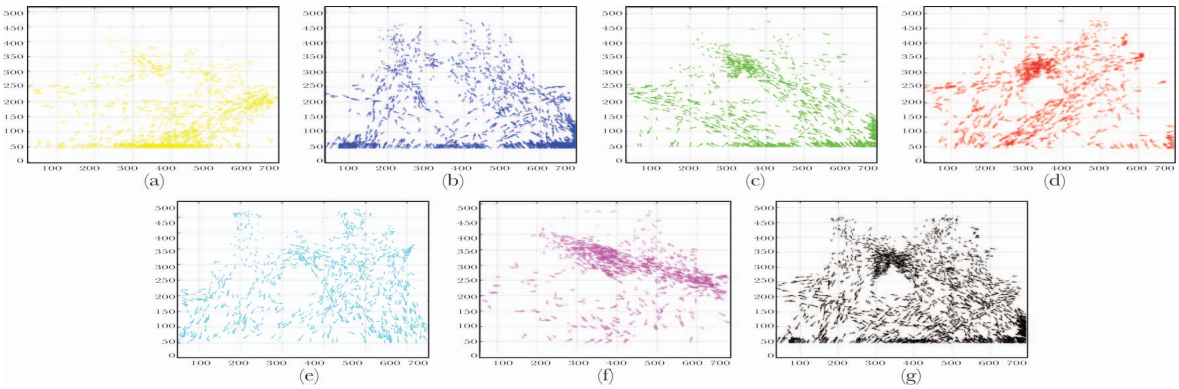

Fig. 3: List of behaviours learned from training dataset. a) Trajectories heading north-east. b) Trajectories heading north. c) Trajectories heading north-west. d) Trajectories heading south-west. e) Trajectories heading south. f) Trajectories heading south-east. g) Trajectories heading to multiple directions.

Recognition, ICPR 2014, Stockholm, Sweden, August 24-28, 2014. 2014, pp. 4164–4169, IEEE.

[6] N. Courty, P. Allain, C. Creusot, and T. Corpetti, "Using the agoraset dataset: Assessing for the quality of crowd video analysis methods," *Pattern Recognition Letters*, vol. 44, pp. 161 – 170, 2014, Pattern Recognition and Crowd Analysis.

[7] P. Morerio, L. Marcenaro, and C. S Regazzoni, "People count estimation in small crowds," in *Advanced video and signal-based surveillance (AVSS), 2012 IEEE Ninth International Conference on*. IEEE, 2012, pp. 476–480.

[8] A. Dehghan, H. Idrees, A. R. Zamir, and M. Shah, "Automatic detection and tracking of pedestrians in videos with various crowd densities," in *Pedestrian and Evacuation Dynamics 2012*, U. Weidmann, U. Kirsch, and M. Schreckenberg, Eds., pp. 3–19. Springer International Publishing, 2014.

[9] B. Leibe, E. Seemann, and B. Schiele, "Pedestrian detection in crowded scenes," in *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 1 - Volume 01*, Washington, DC, USA, 2005, CVPR '05, pp. 878–885, IEEE Computer Society.

[10] B. Fritzke, "A growing neural gas network learns topologies," in *Advances in Neural Information Processing Systems 7*. 1995, pp. 625–632, MIT Press.

[11] T. Kohonen, "The self-organizing map," *Proceedings of the IEEE*, vol. 78, pp. 1464–1480, 1990.

[12] F. Castaldo, F. A. N. Palmieri, V. Bastani, L. Marcenaro, and C. S. Regazzoni, "Abnormal vessel behavior detection in port areas based on dynamic bayesian networks," in *17th International Conference on Information Fusion, FUSION 2014, Salamanca, Spain, July 7-10, 2014*. 2014, pp. 1–7, IEEE.

[13] F. Solera and S. Calderara, "Social groups detection in crowd through shape-augmented structured learning," in *Image Analysis and Processing - ICIAP 2013 - 17th International Conference, Naples, Italy, September 9-13, 2013. Proceedings, Part I*, Alfredo Petrosino, Ed. 2013, vol. 8156 of *Lecture Notes in Computer Science*, pp. 542–551, Springer.

[14] M. Manfredi, R. Vezzani, S. Calderara, and R. Cucchiara, "Detection of static groups and crowds gathered in open spaces by texture classification," *Pattern Recogn. Lett.*, vol. 44, no. C, pp. 39–48, July 2014.

[15] B. Zhou, X. Wang, and X. Tang, "Understanding collective crowd behaviors: Learning a mixture model of dynamic pedestrian-agents.," in *CVPR*. 2012, pp. 2871–2878, IEEE Computer Society.

[16] B. Solmaz, B. E. Moore, and M. Shah, "Identifying behaviors in crowd scenes using stability analysis for dynamical systems," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 10, pp. 2064–2070, Oct. 2012.

[17] J. Acevedo-Rodríguez, S. Maldonado-Bascón, R. Javier López-Sastre, P. Gil-Jiménez, and A. Fernández-Caballero, "Clustering of trajectories in video surveillance using growing neural gas.," in *Proceedings of the 4th international conference on Interplay between natural and artificial computation - IWINAC11*. 2011, vol. 6686, pp. 461–470, Springer.

[18] P. Widhalm and N. Brändle, "Learning major pedestrian flows in crowded scenes.," in *20th International Conference on Pattern Recognition, ICPR 2010, Istanbul, Turkey, August 23-26, 2010*. 2010, pp. 4064–4067, IEEE Computer Society.

[19] O. J. Urizar, M. S. Baig, E. I. Barakova, C. S. Regazzoni, L. Marcenaro, and M. Rauterberg, "A hierarchical bayesian model for crowd emotions," *Frontiers in Computational Neuroscience*, vol. 10, pp. 63, 2016.