INFORMATION BOTTLENECK-BASED RELEVANT KNOWLEDGE REPRESENTATION IN LARGE-SCALE VIDEO SURVEILLANCE SYSTEMS

Simone Chiappino, Lucio Marcenaro, Carlo S. Regazzoni

DITEN, University of Genova
Via Opera Pia 11A 16145 Genoa - Italy
{s.chiappino, mlucio, carlo}@ginevra.dibe.unige.it

ABSTRACT

Extraction and representation of relevant information from large-scale surveillance systems constitute fundamental processes for allowing automatic interpretation of complex scenes. In particular, when the amount of information increases (i.e., due to a larger number of monitored areas), attention focusing techniques are needed to highlight most relevant parts within the overall acquired data. When wide area surveillance systems are considered, one of the major problems in event detections is the reconstruction of the scene as a whole, from spatially limited observations. In this paper, a novel representation technique for sparse information, based on information theory, is presented. Self Organizing Maps (SOMs) have been used for classifying and correlating observed sparse data time series. By means of *Information Bottleneck* theory, it is possible to determine the optimal data representation in the SOM-space as a trade-off between the signal reconstruction capabilities and the original data statistical similarities preservation. Proposed experiments show how the so called *information bottleneck-based SOM selection* for knowledge modelling, can be applied to the field of crowd monitoring for people density map estimation and event detection. Results are presented on synthetic and real video sequences.

Index Terms— Information bottleneck, Cognitive systems, Anomalous event detections, Crowd monitoring, Self-Organizing Maps

1. INTRODUCTION

Automatic representation, analysis and detection of abnormal events is a central issue for last generation video surveillance systems. In this context, distributed interactive and intelligent systems embedded in physical environments can represent a breakthrough in the design of people-oriented services applied to different application domains, among which, crowd monitoring in large-scale environments is becoming one of the most relevant. Several works have been devoted in the last decade to link traditional computer vision tasks to high-level context aware functionalities such as scene understanding, interaction classification and recognition of possible threats or dangerous situations. For instance, in [1] a method for crowd behaviour analysis based on social forces and optical flow is proposed. In [2] the authors present an innovative method based on people flow estimation. In order to detect crowd events, a new abstract viscous fluid field has been proposed in [3]. More recently, in [4], a people trajectories based social force model has been proposed for describing interactions among the individual members of a group of people. Different features have been considered for automatic crowd analysis: local features (e.g., *features from accelerated segment test - FAST*) can be used for people detection, while optical flow efficiently estimates human motion [5]. Considering such features it is possible to evaluate the density of a crowd ([6, 7]).

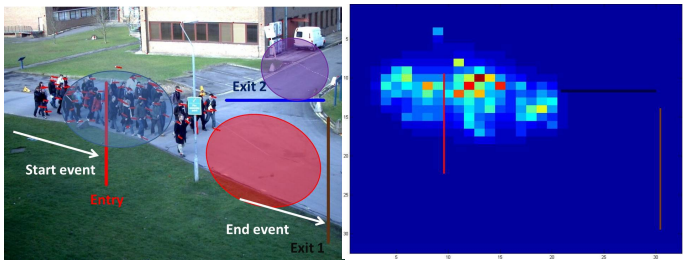
Video crowd analysis frameworks in the state-of-the-art typically do not address two major problems that arise when an higher number of sensors are used: rigorous methods for obtaining 1) optimal *information representation* able to maintain the *informativeness* of acquired low level features, as well as 2) compact description to reduce the processing *complexity* due to an ever increasing amount of information, are needed.

The problem of information overload can be avoided through an automatic method for selecting subparts of the guarded environment and focusing operators' attention on most informative regions, such as the one proposed in [8]. In Figure 1.a an example of relevant information extraction, which is defined as sparse information, is shown. The main problem, for event detection and classification mechanisms, is related to the reconstruction accuracy of original data from incomplete and limited observations. Typical tasks in crowd monitoring applications consist in recognizing particular events within the crowd itself, such as presence of crush in forbidden areas or suspicious movements. For instance, in Figure 1.a two events of interest can be defined when the crowd flow crosses red and brown lines, respectively. An approach based on *Self Organizing Maps (SOMs)* for reconstructing observed signals is presented in [9]; more recently, a SOM-based algorithm for defective image restoration is proposed [10]. In particular, it is highlighted how the SOM-based method performances depend on Kohonen-layer size. Other artificial neural networks derived from SOMs, such as *Growing Neural Gas (GNG)* (see [11]) and *Growing Hierarchical SOM (GH-SOM)* (see [12]), can automatically adapt the dimension of the Kohonen-layer. More in details, GNG computes an accumulated local error, which represents a distance measurements between two neuronal weights, and increases the number of neurons if this is considered too large. Similarly, in GH-SOM the increase of the number of neurons and layers is based on distance measurements between neuronal weights and input data. Another type of neural network (*Neural Gas (NG)*) can improve the input data topology preservation through an adaptive method based on learning of neighbourhood relationships between the weight vector (associated with neuronal unit) and each external stimuli (associated with input vector).

These mechanisms of adapting layer sizes and topology preservation are mainly addressed towards original data reconstruction. The problem of recovering the signal from sparse data, requires more than just reconstruction accuracy: it is indeed necessary to preserve the *similarities* between relevant information and original data (i.e., input signals). This SOM-layer size optimization problem can be represented by means of specific cost function, which relies on *In-*

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

formation Bottleneck theory [13].



(a) Relevant events. (b) Crowd density map.

Fig. 1: Relevant information extraction and crowd density map estimation. a. The coloured circles specify important subparts of the scene; optical flow features represent sparse relevant information. b. Crowd density map estimation by Lucas-Kanade [14] optical flow features.

The contributions of this paper are as follows. A novel approach is presented, based on information bottleneck, for designing a cost function able to quantify the SOM trade-off between the capability to recover original signals and preserve statistical similarities between sparse relevant information and original data. It will be shown how SOMs' correlation abilities can be measured through a mixture of local linear regressive models associated to each neuron. Such models can be used for predicting the future values based on previous states. Finally, by means of the proposed cost function, an algorithm is described for *information bottleneck-based SOM selection (IB-SOM)*. The proposed framework has been applied to crowd monitoring domain for people density estimation and event recognition on video real sequences extracted from public database PETS [15]. Moreover, proposed approach is compared to other neural networks, such as NG, GNG and GH-SOM. The remainder of the paper is organized as follows: in Section 2 the information bottleneck-based SOM selection for relevant knowledge representation is presented; experimental results are described in Section 3, while conclusions are drawn in Section 4.

2. INFORMATION BOTTLENECK-BASED RELEVANT INFORMATION REPRESENTATION

This section describes the proposed relevant information representation method applied to video-surveillance. In the communication system theory, encoding a time-varying multi-dimensional signal $X(t)$ is a common approach for extracting relevant information from it. The available information acquired from video-surveillance network can be defined as the vector $X = [X_1, X_2, \cdots, X_N]$ where $X \in \mathbf{X}$ with $X \in \mathbb{R}^N$ and X is a sample of $X(t)$, which has been acquired at sampling time t. For crowd monitoring applications, X_i represents people density in each $i-th$ monitored area with $i = 1, \cdots, N$ and N is the maximum number of guarded zones. Such a vector describes the crowd density map. It is possible to define the relevant information $\tilde{X}$, extracted at time instant t by using the attention focusing algorithm in [8], as a subset of X: $\tilde{X} \subseteq X$ where $\tilde{X} \in \mathbb{R}^M$ and $M \leq N$. Figure 2 shows how relevant sparse information $\overline{X}$ (i.e., crowd density sparse map) can be reconstructed from $\tilde{X}$. The percentage of controlled area can be computed as a ratio between the number of significant values in $\overline{X}$ and total available information contained in X. The quantity $\overline{X} \in \mathbb{R}^N$ is a sample of the reconstructed signal $\overline{X} \approx X$. The SOM projects input data (i.e., $\mathbf{X}$) into reduced dimensionality space. The neural network has the ability to semantically represent input vectors by selecting a similar but not necessary identical crowding density maps within the same neuronal unit. Each neuron represents a codeword associated to a prototype vector, i.e., a weight $W_k \in \mathbf{W}$ with $W_k \in \mathbb{R}^N$, where $k \in 1, \cdots, K$ and K is the maximum number of prototype vectors within SOM-layer.

The problem can be formalized as that of representing with the same best match unit (BMU) $\hat{W}_k$ the vector X and the corresponding sparse vector $\overline{X}$, as shown in Figure 2. The central task is here to establish the firing properties of neuronal patterns by balancing *reconstruction* capabilities and *correlation* properties, between sparse information and neurons. To this end, a new variable $\mathbf{Y}$ has been defined in order to quantify the differences between sparse and original signals. Such a variable is as informative as possible of $\mathbf{X}$. Reconstruction and correlation attributes lead to the information bottleneck concept, which is defined as a trade-off between two average mutual information (AMI) $I(W,X)$ and $I(W,Y)$.

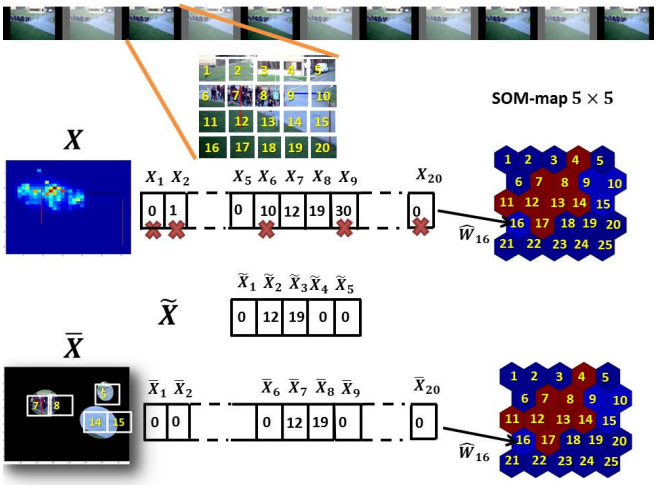


Fig. 2: Relevant information extraction and crowding density map projections into SOM-space. The grid cells, laid over the image plane, can be seen as the set of controlled areas. Each cell is associated to a number of feature extracted by Lucas-Kanade optical flow and used for estimating the crowd density map. In this example $X \in \mathbb{R}^{20}$, $\tilde{X} \in \mathbb{R}^5$ and the sparse vector $\overline{X} \in \mathbb{R}^{20}$. X and $\overline{X}$ are mapped into the same unit $\hat{W}_{16}$. The percentage of controlled area corresponds to 40% (i.e. 2|5).

The first term $I(W,X)$ denotes the reconstruction measurement; according to the *Rate Distortion theory*, it should be minimized depending on the allowed distortion d_W introduced by the mapping process. Such a distortion is measured by the conditional entropy $H(X|W)$: more details are given in [16]. The second term $I(W,Y)$ represents the correlation measurement between sparse and original data, which should be maximized. A practical measure of the correlation is proposed as the difference between statistical relationships of data, described by $p(Y,X)$, and neuronal unit correlation capabilities which are defined by $p(Y,W)$. Where $p(Y,X)$ and $p(Y,W)$ are two joint probabilities.

The quantities $p(Y,W)$ and $p(Y,X)$ can be estimated by using the SOMs for dividing the set of input data (i.e., $\mathbf{X}$) into different multivariate time series $\{\mathbf{X}_k\}_{k=1}^K$ where $\mathbf{X}_k = \{X_{1,k}, \cdots, X_{n,k}\}$ associated to the $k-th$ neuron, such as $\mathbf{X}_k \cup \mathbf{X}_j = \varnothing$ with $k \neq j$ and $\bigcup_{k=1}^K \mathbf{X}_k = \mathbf{X}$ [17]. These sub-sequences of vectors can be modelled

by local *Vector Auto Regressive* (VAR) models [18]. The number of generated VAR models corresponds to the number of neurons of the SOM. Considering a multivariate time series $\mathbf{X}_k$, an auto regressive model of order m, denoted as $VAR(m)$, describes the $i-th$ vector $X_{i,k}$ as linear combination of the previous state vectors:

$$X_{i,k} = \Phi_0 + \Phi_1 X_{i-1,k} + \Phi_2 X_{i-2,k} + \cdots + \Phi_p X_{i-p,k} + \varepsilon_{i,k}, \qquad (1)$$

where $\Phi_0, \cdots, \Phi_p$ are $(N \times N)$ parameter matrices and ε_i represents a $(N \times 1)$ Gaussian noise. By the multivariate time series $\mathbf{X}_k$ we have modelled a $VAR(2)$ as $X_{i,k} = \hat{\Phi}_0 + \hat{\Phi}_1 X_{i-1,k} + \hat{\Phi}_2 X_{i-2,k} + \varepsilon_{i,k}$, where $\hat{\Phi}_0$, $\hat{\Phi}_1$ and $\hat{\Phi}_2$ are estimated coefficient matrices which are stored in each SOM node. In order to determine the fitting of the data to the $VAR(2)$ models, error terms are estimated as follows: $\hat{\varepsilon}_{i,k} = X_{i,k} - [\hat{\Phi}_0 + \hat{\Phi}_1 X_{i-1,k} + \hat{\Phi}_2 X_{i-2,k}]$.

The error vector associated to each neuron is denoted with $\hat{\varepsilon}_k$. The average of $\hat{\varepsilon}_k$ is denoted with $Y_k = \frac{1}{N} \sum_{c=1}^{N} \hat{\varepsilon}_{k,c}$, where $\hat{\varepsilon}_{k,c}$ is $c-th$ component of $\hat{\varepsilon}_k$.

It is supposed that $p(Y,W) = \mathcal{N}(0, \sigma_{Y,W})$ is the joint pdf between $\mathbf{Y}$ and $\mathbf{W}$, where $\sigma_{Y,W} = E\{Y_k{}^2\} - E\{Y_k\}^2$. For expressing the quantity $p(Y,X)$, it is sufficient to define only one $VAR(2)$ and this can be used for modelling all input data $\mathbf{X}$. The same approach can be used for estimating $p(Y,X)$.

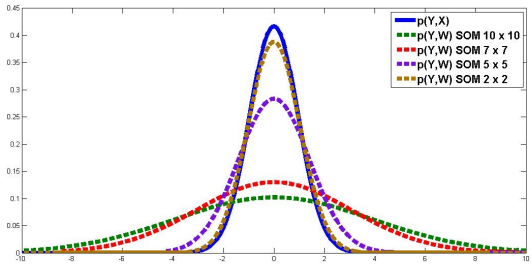


Fig. 3: Kullback-Leibler divergences D_{KL} for two Gaussian probability density functions $p(Y,W)$ and $p(Y,X)$. The information lost has been represented by distance metric between $p(Y,W)$ and $p(Y,X)$ (see Table 1).

The optimal SOM-layer size can be obtained by minimizing the modified cost function based on information bottleneck, as follows:

$$\mathscr{F} = \min_{p(W,X)} \{H(X|W) + \lambda D_{KL}[p(Y,W) \| p(Y,X)]\}; \qquad (2)$$

where Kullback-Leibler divergence $D_{KL}[p(Y,W) \| p(Y,X)]$ is a measure of the difference between the two joint probability distributions $p(Y,W)$ and $p(Y,X)$. Figure 3 shows that a larger SOM-layer (e.g. $K = 100$ neurons) present higher D_{KL} values (i.e., poor correlation quality) than smaller SOM (e.g., $K = 4$ neurons).

It can be noticed that D_{KL} describes an effective distortion measurement. In Equation 2 the λ parameter was introduced, which can balance the information bottleneck. In particular when $\lambda \to 0$, the cost function privileges reconstruction capabilities of the SOMs, (i.e., larger SOM-layers will be selected). Vice versa when $\lambda \to \infty$, $\mathscr{F}$ selects the correlation properties of the SOMs, (i.e., smaller SOM-layers will be selected).

Table 1: Cost function parameters. For evaluating D_{KL}, two divergence normalized density functions $p(Y,X) = N(0, 0.9996)$ and $p(Y,W_K)$ are being considered.

SOM	$H(W\|X)$	D_{KL}	$p(Y,W)$	dr_W	dp_W
2×2 $\lambda \in (1.28 \div \infty)$	2,38	0,005	$N(0, 0.94)$	0,842	0,3201
5×5 $\lambda \in (0.48 \div 1.28]$	2,04	0,27	$N(0, 1.88)$	0,623	0,4018
7×7 $\lambda \in (0.32 \div 0.48]$	1,89	0,60	$N(0, 3)$	0,417	0,53
10×10 $\lambda \in (0 \div 0.32]$	1,78	0,96	$N(0, 4)$	0,209	0,78

3. EXPERIMENTAL RESULTS

This section is divided into two subparts: in the first part, SOMs training and information bottleneck based cost function evaluation is carried out through synthetic data. Then, performances of the proposed IB-SOM selection algorithm are compared with other neural networks (GH-SOM, GNG and NG), for crowding density reconstruction on real video sequences.

3.1. Training of SOMs and cost function evaluation on synthetic data

A common training set is generated by using a simulator where crowd behaviours are generated based on Social Forces model [19]. The simulator has the capability to add virtual sensors able to *acquire* data coming from different subparts of the monitored scene. A virtual image processing algorithm has been implemented for obtaining a plausible crowd density map for each frame. Generated map is a 32×32 matrix, which corresponds to a vector $\mathbf{X}$ with $N = 1024$ components. Four different SOMs were used, with $K = 100, 49, 25, 4$ number of neurons respectively and the following layer topologies: 10×10, 7×7, 5×5 and 2×2.

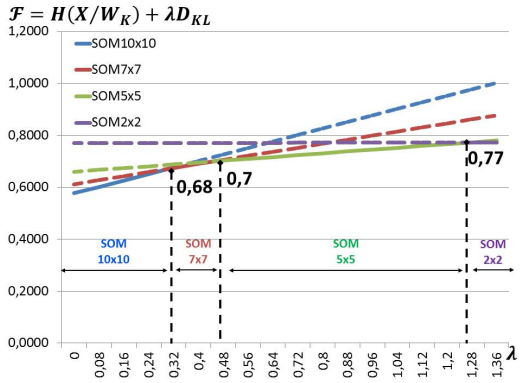


Fig. 4: Normalized cost function average trends for different SOMs. The experiments have been conducted on 100 sequences of synthetic data provided by the simulator. Total time of each simulation is $1000secs$. The validity regions are defined by intervals of λ.

By using the common training set, SOMs and other neural networks have been trained. Finally, the parameters for evaluating the normalized information bottleneck-based cost function curves $\mathscr{F}$ are

Table 2: Comparisons between the proposed IB-SOM selection and other neural networks. Results are presented for different percentages of controlled areas. In the table normalized reconstruction errors are shown. For IB-SOM reconstruction dr_{W_K} and prediction dp_{W_K} errors have been computed. The last two rows show the averages and the variances of the errors.

	IB-SOM dr_W—dp_W	GH-SOM dr_W	GNG dr_W	NG100 dr_W	NG49 dr_W	NG4 dr_W
100%	0,331 — 0,841	0,403	0,396	**0,234**	0,409	0,754
80%	0,343 — 0,5743	0,409	0,408	**0,335**	0,451	0,754
60%	**0,355** — 0,5006	0,423	0,412	0,391	0,511	0,811
40%	**0,431** — 0,4089	0,685	0,639	0,588	0,533	0,853
Average	**0,365 — 0,5412**	0,480	0,463	0,387	0,476	0,793
Variance	**0,0015 — 0,0099**	0,0140	0,0102	0,0166	0,0023	0,0017

determined (see Figure 4). In Table 1, dr_W and dp_W represent average reconstruction and prediction errors obtained by different value intervals of λ parameter. dr_W is the average error (i.e., Euclidean distance) between the input data $\mathbf{X}$ and its representation $\mathbf{W}$. Each *VAR* model can be used as linear predictor filter. The dp_W has been defined by an average measurement of the fitting between one period ahead forecast sequences $\hat{X}_k$ (obtained by $VAR(2)$ local models) and the training data X_k: $dp_W = \frac{\sum_{i=1}^{n} \|X_{i,k} - \hat{X}_{i,k}\|}{\sum_{i=1}^{n} \|X_{i,k} - E\{X_k\}\|}$.

For small λ values, such as $(0 \div 0.32]$, the minimum of $\mathscr{F}$ is given by the SOM 10×10 (i.e., the reconstruction capabilities will be preserved and $dr_W > dp_W$). Vice versa for higher λ values, such as more than 1.2, the minimum of $\mathscr{F}$ is given by the SOM 2×2 (i.e., the correlation proprieties will be maintained and $dp_W > dr_W$).

3.2. Crowd density reconstruction on real video sequences

An experiment has been conducted on three available video sequences from PETS dataset for single camera (S1 L2 Time 14 : 06 and 14 : 31, S3 High Level Time 14 : 33 View_0001; sequences length are 200, 130 and 377 frames respectively and frame rate is ~ 7 [fps]). The information bottleneck theory is adopted as a practical strategy for optimal SOM selection; by using this approach it is possible to limit the reconstruction error by varying the percentages of controlled areas.

Under the hypothesis that the data are acquired and processed at the same PETS sequence frame rates, λ parameter (see Equation 2) can be defined as follows: $\lambda \propto d(X, W_k)|_{p(W,X)}$. Such a value can automatically balance the bottleneck through a distortion $d(\cdot)$ (i.e., Euclidean distance), which is due to mapping process $p(W,X)$, between observed vector X and its representation W_k. In particular, when $d(X,W_k)$ is low, the reconstruction capabilities will be preserved (i.e., larger SOM-layers will be selected). Vice versa, when $d(X,W_k)$ is high, the correlation properties of the SOMs will be maintained (i.e., smaller SOM-layers will be selected).

The table 2 shows how the NG100 has the minimum reconstruction errors in 100% and 80% of controlled area percentages. Vice versa when the controlled area percentages decrease (i.e., 60% and 40%) the distortions of this neural network increase. In these situations, proposed method can find the optimal SOM size. IB-SOM selection restricts reconstruction accuracy reduction, i.e., it is able to maintain a minimum reconstruction average error. Moreover, the proposed approach delimits error variations, due to different percentages of controlled areas (i.e., error variance).

On the other hand, when the SOM-map sizes are reduced the prediction errors decrease as well. Finally, in Figure 5 quantitative and qualitative comparison measurements for event recognition are presented. In particular, using limited observations as well, the proposed IB-SOM presents smaller distortion errors. The density map reconstructions show how all the neural networks can identify the first event, while only through the proposed approach it is possible to recognize the second event.

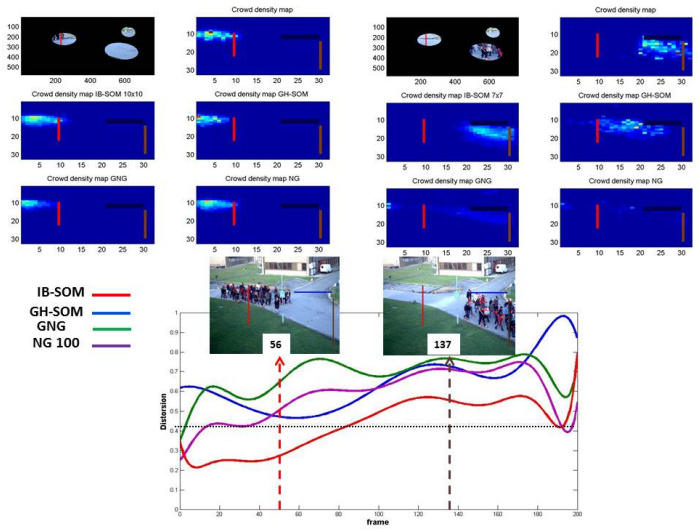


Fig. 5: Qualitative and quantitative results for event recognition for PETS sequence S1 L2 Time 14 : 06 using 40% of the controlled area. The figure shows the comparison between the proposed method and other approaches: on the upper part crowding density map reconstructions are presented; on the lower part distortion curves are shown. The whole video is available on https://www.youtube.com/watch?v=KN2aYZ64TTw.

4. CONCLUSIONS

This paper presented a novel approach for information representation. It has been applied for sparse data within a crowd monitoring application. The proposed algorithm is a method encompassing different steps, which involves the application of information theory and neural networks such as SOMs. First of all, by means of information bottleneck paradigm, a cost function has been designed in order to balance the data reconstruction and correlation capabilities of different SOMs. An information bottleneck based strategy for SOM selection was proposed.

Finally, the IB-SOM selection method has been tested on public datasets. The results show that the proposed approach outperforms other neural networks (such as NG, GNG, GH-SOM) in crowd density reconstruction from very sparse observations.

Furthermore, it has been shown how such a knowledge representation can recover original crowding density maps in order to recognize particular events on real video sequences.

Future developments of this work will include a detailed study on the impact of the information bottleneck on the GH-SOM. It can lead to improve the GH-SOM strategy for selecting the knowledge representation among different hierarchical layers.

5. REFERENCES

[1] R. Mehran, A. Oyama, and M. Shah, "Abnormal crowd behavior detection using social force model," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, Jun. 2009, pp. 935 –942.

[2] B. Solmaz, B. E. Moore, and M. Shah, "Identifying behaviors in crowd scenes using stability analysis for dynamical systems," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 10, pp. 2064 –2070, oct. 2012.

[3] H. Su, H. Yang, S. Zheng, Y. Fan, and S. Wei, "Crowd event perception based on spatio-temporal viscous fluid field," in *Advanced Video and Signal-Based Surveillance (AVSS), 2012 IEEE Ninth International Conference on*, sept. 2012, pp. 458 –463.

[4] R. Mazzon, F. Poiesi, and A. Cavallaro, "Detection and tracking of groups in crowd," in *Proc. of IEEE Int. Conference on Advanced Video and Signal based Surveillance (AVSS)*, Krakow, Poland, 27–30 August 2013.

[5] R. Perko, T. Schnabel, G. Fritz, A. Almer, and L. Paletta, "Counting people from above: Airborne video based crowd analysis," *CoRR*, vol. abs/1304.6213, 2013.

[6] H. Fradi, V. Eiselein, I. Keller, J.-L. Dugelay, and T. Sikora, "Crowd context-dependent privacy protection filters," in *DSP 2013, 18th International Conference on Digital Signal Processing, 1-3 July 2013, Santorini, Greece*, Santorini, GRÈCE, 07 2013. [Online]. Available: https://www.eurecom.fr/publication/3995

[7] S. A. Rao, J. Gubbi, S. Marusic, P. Stanley, and M. Palaniswami, "Crowd density estimation based on optical flow and hierarchical clustering," in *ICACCI 2013 International Conference on Advances in Computing Communications and Informatics , 22-25 August 2013, Mysore, India*, Mysore, India, Aug. 2013. [Online]. Available: http://people.eng.unimelb.edu.au/jgl/Publication.html

[8] S. Chiappino, L. Marcenaro, and C. Regazzoni, "Selective attention automatic focus for cognitive crowd monitoring," in *Proc. of IEEE Int. Conference on Advanced Video and Signal based Surveillance (AVSS)*, Krakow, Poland, 27–30 August 2013.

[9] J. Cho, A. R. C. Paiva, S.-P. Kim, J. C. Sanchez, and J. C. Príncipe, "Self-organizing maps with dynamic learning for signal reconstruction," *Neural Netw.*, vol. 20, no. 2, pp. 274–284, Mar. 2007. [Online]. Available: http://dx.doi.org/10.1016/j.neunet.2006.12.002

[10] M. Maeda, "Restoration model with inference capability of self-organizing maps," in *Advances in Self-Organizing Maps*, ser. Advances in Intelligent Systems and Computing, P. A. Estvez, J. C. Prncipe, and P. Zegers, Eds. Springer Berlin Heidelberg, 2013, vol. 198, pp. 153–162. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-35230-0_16

[11] F. Canales and M. Chacn, "Modification of the growing neural gas algorithm for cluster analysis." in *CIARP*, ser. Lecture Notes in Computer Science, L. Rueda, D. Mery, and J. Kittler, Eds., vol. 4756. Springer, 2007, pp. 684–693. [Online]. Available: http://dblp.uni-trier.de/db/conf/ciarp/ciarp2007.html

[12] A. Rauber, D. Merkl, and M. Dittenbach, "The growing hierarchical self-organizing map: Exploratory analysis of high-dimensional data," *IEEE Transactions on Neural Networks*, vol. 13, pp. 1331–1341, 2002.

[13] N. Tishby, F. C. Pereira, and W. Bialek, "The information bottleneck method," 1999, pp. 368–377.

[14] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," 1981, pp. 674–679.

[15] J. Ferryman and J. L. Crowley, Eds., *Eleventh IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, PETS 2009*, 2009. [Online]. Available: http://www.cvg.rdg.ac.uk/PETS2009

[16] S. Chiappino, P. Morerio, L. Marcenaro, and C. S. Regazzoni, "Event definition for stability preservation in bio-inspired cognitive crowd monitoring," in *Digital Signal Processing (DSP), 2013 18th International Conference on*, Jul. 2013, pp. 1–6.

[17] H. NI and H. YIN, "Self-organising mixture autoregressive model for non-stationary time series modelling," *International Journal of Neural Systems*, vol. 18, no. 06, pp. 469–480, 2008, pMID: 19145663. [Online]. Available: http://www.worldscientific.com/doi/abs/10.1142/S0129065708001737

[18] B. Pfaff, "Var, svar and svec models: Implementation within r package vars," *Journal of Statistical Software*, vol. 27, no. 4, pp. 1–32, 7 2008. [Online]. Available: http://www.jstatsoft.org/v27/i04

[19] S. Chiappino, L. Marcenaro, P. Morerio, and C. Regazzoni, "Event based switched dynamic bayesian networks for autonomous cognitive crowd monitoring," in *Augmented Vision and Reality*, ser. Augmented Vision and Reality. Springer Berlin Heidelberg, 2013, pp. 1–30.