A BIO-INSPIRED LOGICAL PROCESS FOR SALIENCY DETECTIONS IN COGNITIVE CROWD MONITORING

Simone Chiappino *Andrea Mazzù* *Lucio Marcenaro* *Carlo S. Regazzoni*

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

ABSTRACT

It is well known from physiological studies that the level of human attention for adult individuals rapidly decreases after five to twenty minutes [1]. Attention retention for a surveillance operator represents a crucial aspect in Video Surveillance applications and could have a significant impact in identifying relevance, especially in crowded situations. In this field, advanced mechanisms for selection and extraction of saliency information can improve the performances of autonomous video surveillance systems and increase the effectiveness of human operator support. In particular, crowd monitoring represents a central aspect in many practical applications for managing and preventing emergencies due to panic and overcrowding.

In this paper, an adaptive inductive reasoning mechanism for saliency extraction and information reconstruction for a distributed camera sensors network is presented. The proposed system, by means of Self Organizing Maps (SOMs) [2], can learn the correlation of the observed data and then recover the whole information from a subset of available sensors.

Experimental results show how the proposed system can reconstruct the information about the non-observed parts starting from relevant data acquired from observed areas of the environment.

Index Terms— Self Organizing Maps, saliency detection, cognitive crowd monitoring, machine learning, human reasoning

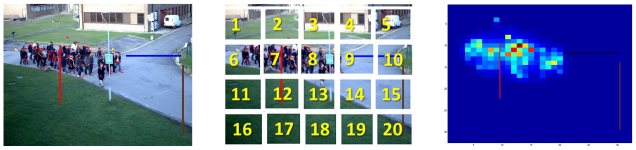
1. INTRODUCTION

There is currently a great interest in methodologies which combine computer vision tasks with high-level situation awareness modules for environment monitoring. Techniques such as behavior analysis, event detection, or recognition of possibly dangerous situations can have a high impact in different sectors, ranging from social (e.g. security-related) to technological ones (e.g. resources management). Several works in the literature try to link physical and social aspects to video surveillance tasks [3]. The main and common problem is that they require monitoring of events in wide video surveillance networks for long periods of time [4].

In order to detect people, local features can be used (e.g., *features from accelerated segment test - FAST* [5]), while optical flow efficiently estimates human motion [6]. Considering such features, a new abstract viscous fluid field has been proposed in [7]. More recently, in [8], a social force model based on people trajectories has been proposed for describing interactions among the individual members of a group of people.

These techniques often require a large number of networked sensors for acquiring data. Adaptable intelligent systems can be exploited in order to balance the amount of needed resources and sensory relevant information. Human perception mechanisms for data processing, integrated in artificial systems, represent a breakthrough for designing a new generation of algorithms inspired by neuroscience. Such frameworks can optimize resource management and improve camera identification by acquiring (i.e. perceiving) relevant information to be combined with inductive reasoning, and machine learning techniques. The main concept in cognitive neuroscience is the Perception Action Cycle (PAC) also referred to as the *Fuster's Paradigm*, which is based on five fundamental building blocks, namely *perception*, *memory*, *attention*, *intelligence* and *language*, [9]. The PAC describes how an entity can perceive, learn and change itself through continuous interaction experiences with the external world. In particular the perception and attention blocks are meant to optimize the information flow within the system. In such a context, Haykin et al. in [10] have presented an algorithmic implementation of the human perceptual attention process in order to separate the relevant from irrelevant information. The challenge of this scheme is to bridge the gap between cognitive science and engineering in order to design intelligent system called *Cognitive Dynamic Systems* (CDS), originally formalized by Haykin in the cognitive radio domain, [11] and later generalized in [12]. Only recently the scheme was applied in the crowd monitoring field [13].

In the video-surveillance domain many frameworks are dealing with saliency extraction from video sequences. In [14], a crowd anomaly identification technique is proposed through a saliency detector based on center-surround discriminant descriptor (i.e. color, intensity, orientation, scale). Other frameworks for saliency extraction use the Histogram of Oriented Gradient (HOG) combined with Support Vector Machine (SVM) classifier for pedestrian detection [15] and for people counting [16]. This paper proposes an adaptive inductive reasoning mechanism for a saliency-based information reconstruction from a distributed camera network. The common goal of the saliency detection frameworks is to extract relevant information from guarded environments. In our method the relevant information, which is identified directly from environmental observations, is defined through the learning of the relationships among sensory data. Unlike other methods for relevance detection, the proposed algorithm shows how it is possible to use SOMs for learning data correlations. Moreover it is presented how through such a procedure, it is possible to reduce the number of observes areas; by using such limited observations the proposed system can reconstruct the whole information. The presented system works at two different logical levels: by means of SOMs, it is able to learn the correlation of



(a) Optical flow (b) Controlled areas (c) Crowd density

Fig. 1: Information extraction and crowd density map estimation

the salient observed data. Moreover, relying on an adaptive SOM-based inductive reasoning process, it is able to recover the whole information from an optimal subset of sensors, allowing a minimum distortion.

The remainder of this work is organized as follows. Section 2 presents the proposed bio-inspired inductive reasoning model based on SOMs, for extracting relevant data and recovering complete information. Recovering capabilities of the proposed scheme are evaluated in Section 3 by using real data from "Performance Evaluation of Tracking and Surveillance" (PETS) database [17], while conclusions are drawn in 4.

2. DATA CORRELATION REPRESENTATION FOR BIO-INSPIRED INDUCTIVE REASONING ALGORITHM

Inductive reasoning is a mental process allowing humans to draw conclusions according to pre-defined models. In this application, the model describes the relations among different sub-parts of the guarded environment and it is directly derived from acquired sensors information. This section proposes a SOM-based data correlation modeling for inductive reasoning algorithm for relevant information representation. In particular we suppose to apply it to the crowd monitoring domain. In Figure 1a an example of information extraction from a crowded environment is shown. The controlled area can be divided into sub-parts (Figure 1b) and by applying tools such as Lucas-Kanade optical flow [18] it is possible estimate the density of the people within each cell. In this way, a crowd density map, (Figure 1c) can be sketched. This map be used for evaluating people density within the scene.

Let us define the available information $\mathbb{X}$, acquired from a video surveillance network, as the set of crowding state vectors

$$\mathbf{X} = [X_1, X_2, \cdots, X_N]^T \quad \text{where} \quad \mathbf{X} \in \mathbb{R}^N. \tag{1}$$

X_i represents the number of people in each i^{th} monitored area with $i = 1, \cdots, N$, estimated through a people counter embedded on cameras. N also represents the maximum number of sensors, as we assume that one camera is associated with one controlled area.

During a training phase, a SOM provides an adaptive mechanism for acquiring different correlation models, according to similar dynamical evolution of people captured by neurons. The SOM maps the data (i.e. $\mathbf{X}$) into a dimensionally reduced space. It provides a semantic representation of the input vectors, by associating similar (though not necessary identical) crowding states to the same neuronal unit. Each neuron represents a codeword i.e. a prototype (or weight) vector $\mathbf{W}_k = [(W_k)_1, (W_k)_2, \cdots, (W_k)_N]^T$ and $\mathbf{W}_k \in \mathbb{R}^N$ where $k \in \{1, \cdots, K\}$ and K is the maximum number of prototype vectors within the neuronal layer (i.e. the number of neurons in the SOM).

The external stimuli (i.e. observed state vectors) activate specific neuronal units in the SOM-map. Therefore, the SOM can be used for dividing the training data $\mathbb{X}$ into different multivariate sets $\{\mathbb{X}_k\}_{k=1}^K$ where $\mathbb{X}_k = \{\mathbf{X}_{1,k}, \cdots, \mathbf{X}_{n,k}\}$ is associated to the $k-th$ neuron, and $\{\mathbb{X}_k\}_{k=1}^K$ is a *partition* of $\mathbb{X}$, i.e. $\mathbb{X}_k \cup \mathbb{X}_z = \emptyset$, $\forall k \neq z$ and $\bigcup_{k=1}^K \mathbb{X}_k = \mathbb{X}$. By examining the possible relationships among these sub-sequences of training samples, the system can build correlation structures C_k (i.e. correlation matrices) embedded in each SOM-neuron. The relative effects of the number of people in the i^{th} monitored area on people on j^{th} zone is stored in the correlation coefficients $c_k(i, j)$.

$$C_k(\mathbf{X}) = (diag(\Sigma_k))^{-1/2} \Sigma_k (diag(\Sigma_k))^{-1/2} \tag{2}$$

where Σ_k is the covariance matrix in the k^{th} node of the SOM computed as $\Sigma_k = E[(\mathbf{X} - \mathbf{W}_k)(\mathbf{X} - \mathbf{W}_k)^T]$ and $diag(\Sigma_k)$ is the matrix of the diagonal elements of Σ_k.

A *saliency* masking matrix M can be defined as a diagonal matrix with dimension $N \times N$ where the elements on the diagonal are equal to one if the estimated people density in the corresponding area is available, zero otherwise. It is possible to hide (i.e. mask) some components of $\mathbf{X}$ defining a new vector $\overline{\mathbf{X}} \in \mathbb{R}^N$ as $\overline{\mathbf{X}} = M \cdot \mathbf{X}$. The i^{th} component of $\overline{\mathbf{X}}$ is $\overline{X}_i = M_i \cdot X_i$, $i = 1, \cdots, N$. In this way, $\overline{\mathbf{X}}$ will retain the components of $\mathbf{X}$ when $M_i = 1$ and will be set to zero elsewhere.

A "masked" distance between two vectors can be defined as a modified version of the euclidean distance in this case:

$$d_M(\mathbf{X}, \mathbf{Y}) = \|M\mathbf{X} - M\mathbf{Y}\| = \sqrt{\sum_{i=1}^{N} (M_i \cdot X_i - M_i \cdot Y_i)^2}. \tag{3}$$

The role of the proposed algorithm is to deduce the people number inside the non-observed zones. This is accomplished by minimizing the following cost function:

$$F(\hat{\mathbf{X}}) = d_M(\hat{\mathbf{X}}, \overline{\mathbf{X}}) + \|(\hat{\mathbf{X}} - \mathbf{W}_k) - \Omega_k \cdot (\overline{\mathbf{X}} - \mathbf{W}_k)\| \tag{4}$$

where $\hat{\mathbf{X}}$ the state vector to be estimated, $\mathbf{W}_k$ is the weight vector of the k^{th} SOM neuron activated by $\overline{\mathbf{X}}$ and Ω_k is an influence matrix where each element is given by $\omega_k(i, j) = c_k(i, j) \cdot \sigma_j / \sigma_i$. Such a matrix takes into consideration the correlation $c_k(i, j)$ between the zones and their standard deviations σ_i and σ_j which are calculated during training phase.

The cost function F has two terms: the first $F_1(\hat{\mathbf{X}}) = d_M(\hat{\mathbf{X}}, \overline{\mathbf{X}})$ is clearly minimized by $\hat{\mathbf{X}} = \overline{\mathbf{X}}$, i.e. each i^{th} component is equivalent to the observation. The second term $F_2(\hat{\mathbf{X}}) = \|(\hat{\mathbf{X}} - \mathbf{W}_k) - \Omega_k \cdot (\overline{\mathbf{X}} - \mathbf{W}_k)\|$ represents the objective function introduced for estimating the number of people within each non-observed component through the data acquired from the observed areas. This second term takes into account that each non observed component can be estimated as follows: $(I - M) \cdot \hat{\mathbf{X}} = (I - M) \cdot (\Omega_k \cdot [\overline{\mathbf{X}} - (\mathbf{W}_k)] + (\mathbf{W}_k))$ with I identity matrix $N \times N$.

The minimization of the proposed cost function is an optimization problem. The fundamental idea behind this method is to use a gradient descent approach in order to find a local minimum of the function $F(\hat{\mathbf{X}})$. This cost function depends from the reconstructed information $\hat{\mathbf{X}}$, but also implicitly from the masking matrix M that is representing the most informative areas within the guarded environment. This problem can be solved iteratively by searching for the minimum of F by altering $\hat{\mathbf{X}}$ and then modify the available information through the selection of a different masking matrix M with a constraint on the number of observed areas (i.e., the number of elements equal to 1 in matrix M). In the following we will concentrate on finding the local minimum of F for a fixed set of available sensors (i.e. without modifying the masking matrix M at each iteration).

In Algorithm 1 the pseudo code of the proposed method is presented, by considering one single step of the optimization procedure for a fixed $M = M^*$. The algorithm parameters are: the precision value *dmin*, which defines a stop criterion based on minimum threshold of updating; *maxint*, which defines the maximum number of allowed iterations; α, which specifies the step size weighting the value of the cost function.

```
**Data**: $\overline{\mathbf{X}}$, $\mathbf{W}_k$ and $M^*$
**Result**: $\hat{\mathbf{X}}$
Initialization: $\hat{\mathbf{X}}^{new} = \overline{\mathbf{X}} + (I - M)\mathbf{W}_k, dmin, maxint, \alpha$;
Definition:
$F(\hat{\mathbf{X}}) = d_{M^*}(\hat{\mathbf{X}}, \overline{\mathbf{X}}) + \|(\hat{\mathbf{X}} - \mathbf{W}_k) - \Omega_k \cdot (\overline{\mathbf{X}} - \mathbf{W}_k)\|$;
for $j = 1 : N$ do
    if $M^*_j = 0$ with $M^*_j \in M^*$ then
        while $dx \geqslant dmin$ AND $nint \leqslant maxint$ do
            $\hat{\mathbf{X}}^{old} = \hat{\mathbf{X}}^{new}$;
            $\hat{\mathbf{X}}^{new} = \hat{\mathbf{X}}^{old} - \alpha \cdot \frac{\partial F(\hat{\mathbf{X}})}{\partial \hat{\mathbf{X}}}\Big|_{\hat{\mathbf{X}} = \hat{\mathbf{X}}_{old}}$;
            $dx = \|\hat{\mathbf{X}}^{new} - \hat{\mathbf{X}}^{old}\|$;
            $nint++$;
        end
    end
end
```

Algorithm 1: Computing the optimal value for $\hat{\mathbf{X}}$ with a specific masking matrix $M = M^*$

3. EXPERIMENTAL RESULTS

In this section the performance of the proposed algorithm for data recovering are described. In order to give consistence to the proposed approach, an experiment has been conducted on three available video sequences from the PETS workshop dataset for single camera [17]. According to the main purpose of this work, performances of the method are evaluated for selecting the optimal subset of sensors in real crowding scenes, in order to recover the whole information. Five different SOMs were considered, with $K = 25, 16, 9, 4, 1$ neuronal units respectively, and the following layer topologies: 5×5, 4×4, 3×3 and 2×2. The parameters presented in the algorithm 1 are set as follows: $dmin = 0.01$, $maxint = 60$, $\alpha = 0.1$. Finally, the proposed algorithm has been compared with the standard SOM capability to recover missing values, where restoration is based on the rough representation of the loss data by the value provided by the prototype or by the "temporal influence" of the BMU, as defined in [19]. Such a method has been applied for the reconstruction of areas covered by clouds in a time sequence of optical satellite images. We will refer to this method as SOM-BMU method, table 1.

Figure 2 shows the correlations between two locations: for distant zones the correlation is smaller then for neighboring cells.

In Figure 3, an example of how the method works is shown: the set of sparse data are acquired in proximity to entry and exit points. The training set is generated by using a crowd simulator. The same previous five different SOMs were used. The sparse vector $\overline{\mathbf{X}}$ is projected 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 $\mathbf{X} \in \mathbb{R}^{20}$ while $\overline{\mathbf{X}}$ contains 40% of the total information. By comparing the cost functions it is possible to identify the optimal SOM (in this case SOM-9).

	SOM-1	SOM-4	SOM-9	SOM-16	SOM-25
S1 L2Time 14-06 40% of data loss					
MSE	0,911	0.845	0.727	0,651	**0,455**
Cost Function	0,85	0,731	0,698	0,602	**0,53**
SOM-BMU	0,945	0.596	0.581	0.546	0.559
S1 L2Time 14-31 40% of data loss					
MSE	0,908	0.826	0.802	0,688	**0,501**
Cost Function	0,899	0,831	0,788	0,762	**0,677**
SOM-BMU	0,901	0.663	0.645	0.639	0.630
S3 14-33 40% of data loss					
MSE	0,853	0.783	0.713	0,694	**0,655**
Cost Function	0,885	0,811	0,806	0,787	**0,706**
SOM-BMU	0,891	0.799	0.735	0.714	0.701

Table 1: Quantitative results for PETS video sequences. It is important to note that the minimum value of the Cost Function F proposed in this paper, corresponds to the minimum value of MSE

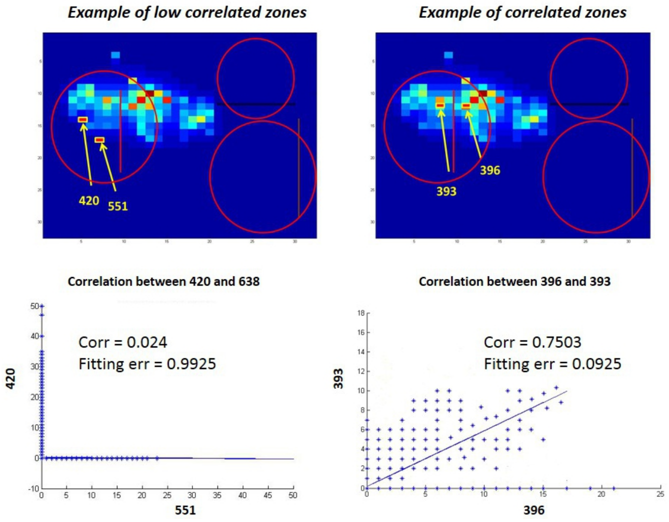


Fig. 2: Example of correlations analysis between locations pairs. The data are represented by the number of feature extracted by Lucas-Kanade optical flow associated to each cell

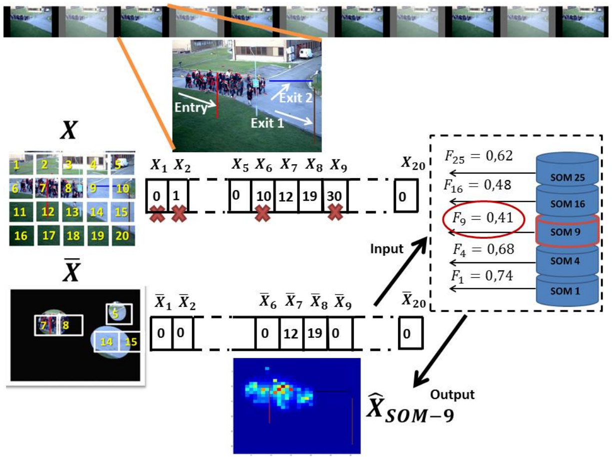


Fig. 3: Salient information extraction and crowding density map reconstruction

By the neuronal activation process, each SOM-K provides a reconstructed crowding density vector $\mathbf{X}_K$ with a corresponding value of the cost function $F_K(\hat{\mathbf{X}}_K)$ (see equation (4) and algorithm 1). In table 1 a quantitative results of the reconstruction are presented. In

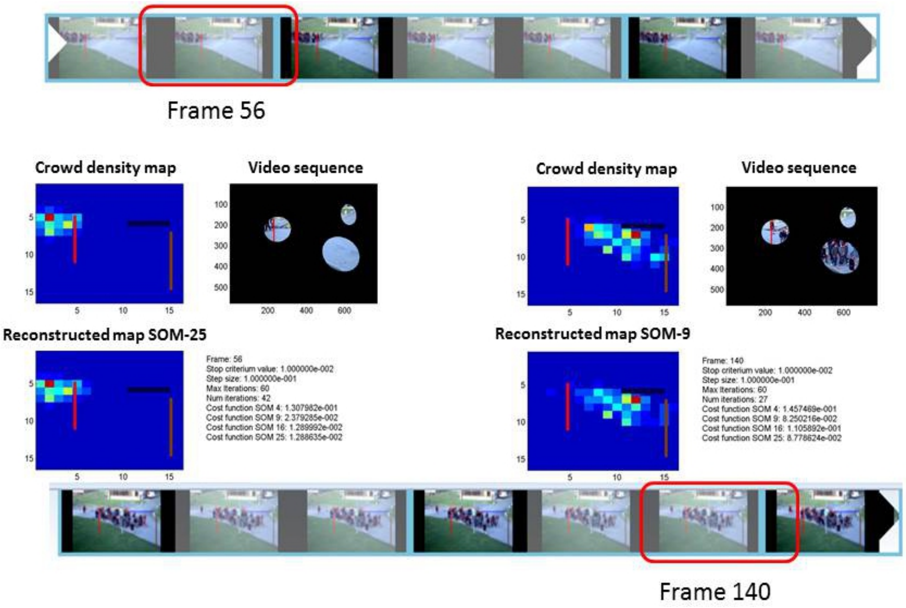

Frame 56
Crowd density map
Video sequence
Reconstructed map SOM-25
Frame: 56
Stop criterium value: 1.000000e-002
Step size: 1.000000e-001
Max iterations: 60
Num iterations: 42
Cost function SOM 4: 1.307982e-001
Cost function SOM 9: 2.379285e-002
Cost function SOM 16: 1.289992e-002
Cost function SOM 25: 1.288635e-002
Crowd density map
Video sequence
Reconstructed map SOM-9
Frame: 140
Stop criterium value: 1.000000e-002
Step size: 1.000000e-001
Max iterations: 60
Num iterations: 27
Cost function SOM 4: 1.457469e-001
Cost function SOM 9: 8.250216e-002
Cost function SOM 16: 1.105892e-001
Cost function SOM 25: 8.778624e-002
Frame 140

Fig. 4: An example of qualitative and quantitative results for PETS sequence S1 L2 Time 14 : 06 (frames 56 and 140) using 40% of the controlled area

this case the SOM-25 provides the lower average cost function value, which corresponds to the minimum (normalized) reconstruction error (MSE).

In Figure 4 a qualitative results of the SOM selection process for crowd density map reconstruction are presented. In the proposed layer dynamical selection process different SOMs can be chosen in order to provide the optimal information recovery. During the initialization phase, the system acquires the data form all the zones. In this case the proposed framework evaluates the optimum SOM-K based on the minimum MSE between the prototype $\mathbf{W}_k$ of the BMU k and the observed data $\mathbf{X}$.

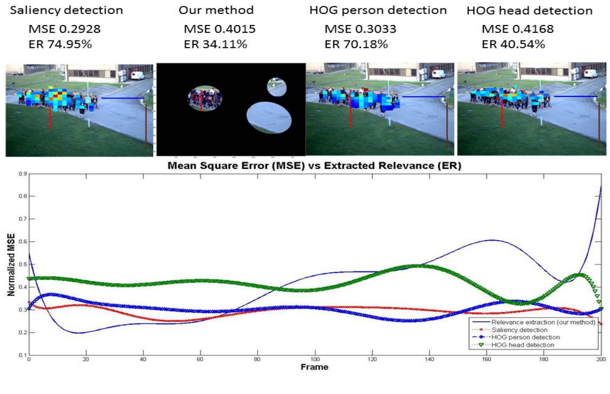


Fig. 5: Reconstruction comparison for different saliency extraction methods: saliency detection presented in [14]; HOG descriptors [15] combined with a learning-based SVM classifier for person (pedestrian) and head detections. The average mean square error normalized to 1 and the percentage of the crowd density map average acquired from the system (i.e. Extracted Relevance-ER) are shown.

Finally, in Figure 5 we compare the proposed strategy for relevant information extraction with other approaches for detecting saliency in video sequences. In particular, we test the recovering algorithm using the saliency detected by the method proposed in [14] and using the HOG descriptors combined with SVM for pedestrian [15] and head detections. The results underline how the presented relevant extraction, based on the correlations among the parts of the environment, is able to drastically reduce the total amount of the acquired information (34.11%) then saliency detection (74.95%) and the pedestrian detection by HOG descriptors (70.18%). We point out like an higher percentage of crowd density saliency extraction is not always synonymous of better reconstruction. Considering an extracted relevance provided by the head detection method (40.54%), our system is able to better recover the whole density map (MSE 0.4015 vs 0.4168).

4. CONCLUSIONS

This paper has presented an adaptive inductive reasoning mechanism able to reconstruct missing information from a set of observations. Salient information has been defined as the optimal subset of observations which are able to reconstruct the lost information which is therefore "redundant" for that specific model. The main goal of the proposed framework is to adapt the inductive reasoning mechanism and therefore to emphasize the saliency. Accordingly, the reconstruction of the non-observed data (i.e. loss values) is performed by means of different correlation models acquired by different SOMs. By minimizing the proposed cost function, the method is able to regenerate the missing data and establish the optimal salient information. In order to evaluate the reconstruction capability, we have tested the algorithm for crowd state recovering by considering a limited subset of observations (i.e. excluding a part of the available camera sensors). The proposed approach optimizes the estimation procedure of the original values by considering not only the prototypes (weights of the BMU) but also the influence of the variations of neighbours' values. Moreover, by minimizing the proposed cost function, by varying the SOM-layer size, it is possible to identify the optimum filter (i.e. SOM) which corresponds to the minimum MSE calculated *a posteriori* (i.e. the minimum introduced distortion).

5. REFERENCES

[1] M. W. Green and N. Sandia National Labs., Albuquerque, *The Appropriate and Effective Use of Security Technologies in U.S. Schools. A Guide for Schools and Law Enforcement Agencies [microform] / Mary W. Green.* Distributed by ERIC Clearinghouse [Washington, D.C.], 1999. [Online]. Available: http://www.eric.ed.gov/contentdelivery/servlet/ERICServlet?accno=ED436943

[2] T. Kohonen, "The self-organizing map," *Proceedings of the IEEE*, vol. 78, no. 9, pp. 1464 –1480, Sep. 1990.

[3] R. Mehran, A. Oyama, and M. Shah, "Abnormal crowd behavior detection using social force model," in *CVPR*. IEEE Computer Society, 2009.

[4] Y. Nam, S. Rho, and J. H. Park, "Intelligent video surveillance system: 3-tier context-aware surveillance system with metadata," *Multimedia Tools Appl.*, vol. 57, no. 2, pp. 315–334, Mar. 2012. [Online]. Available: http://dx.doi.org/10.1007/s11042-010-0677-x

[5] E. Rosten and T. Drummond, "Fusing points and lines for high performance tracking," in *IN INTERNATIONAL CONFERENCE ON COMPUTER VISION*. Springer, 2005, pp. 1508–1515.

[6] 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.

[7] 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.

[8] 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.

[9] J. Fuster, *Cortex and Mind: Unifying Cognition*. Oxford University Press, USA, 2005. [Online]. Available: http://books.google.it/books?id=3R-9dcFw95wC

[10] A. Amiri and S. Haykin, "Improved Sparse Coding Under the Influence of Perceptual Attention," *Neural Computation*, pp. 1–44, Nov. 2013. [Online]. Available: http://dx.doi.org/10.1162/neco_a_00546

[11] S. Haykin, "Cognitive dynamic systems: Radar, control, and radio," *Proceedings of the IEEE*, vol. 100, no. 7, pp. 2095–2103, 2012.

[12] S. Haykin and J. Fuster, "On cognitive dynamic systems: Cognitive neuroscience and engineering learning from each other," *Proceedings of the IEEE*, vol. 102, no. 4, pp. 608–628, April 2014.

[13] S. Chiappino, P. Morerio, L. Marcenaro, and C. Regazzoni, "Bio-inspired relevant interaction modelling in cognitive crowd management," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–22, 2014. [Online]. Available: http://dx.doi.org/10.1007/s12652-014-0224-0

[14] V. Mahadevan, W. Li, V. Bhalodia, and N. Vasconcelos, "Anomaly detection in crowded scenes," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, June 2010, pp. 1975–1981.

[15] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1, 2005, pp. 886–893 vol. 1.

[16] C. Zeng and H. Ma, "Robust head-shoulder detection by pca-based multilevel hog-lbp detector for people counting," in *Pattern Recognition (ICPR), 2010 20th International Conference on*, Aug 2010, pp. 2069–2072.

[17] 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

[18] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proceedings of the 7th International Joint Conference on Artificial Intelligence - Volume 2*, ser. IJCAI'81. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1981, pp. 674–679.

[19] M. Jouini, S. Thiria, and M. Crépon, "Images reconstruction using an iterative som based algorithm," in *Artificial Neural Networks Computational Intelligence and Machine Learning (ESANN), 2012 European Symposium on*, Bruges, Belgium, April 2012, pp. 25–27.