Inexpensive user tracking using Boltzmann machines

Inexpensive user tracking using Boltzmann Machines

Elena Mocanu*, Decebal Constantin Mocanu*, Haitham Bou Ammar†, Zoran Zivkovic ‡
Antonio Liotta * and Evgueni Smirnov§
*Department of Electrical Engineering, Eindhoven University of Technology
†Computer and Information Science Department, University of Pennsylvania
‡ NXP Semiconductors, Central R&D, Netherlands
§Department of Knowledge Engineering, Maastricht University

***Abstract*—Inexpensive user tracking is an important problem in various application domains such as healthcare, human-computer interaction, energy savings, safety, robotics, security and so on. Yet, it cannot be easily solved due to its probabilistic nature, high level of abstraction and uncertainties, on the one side, and to the limitations of our current technologies and learning algorithms, on the other side. In this paper, we tackle this problem by using the Multi-integrated Sensor Technology, which comes at a low price. At the same time, we are aiming to address the lightweight learning requirements by investigating Factored Conditional Restricted Boltzmann Machines (FCRBMs), a form of Deep Learning, that has proven to be an efficient and effective machine learning framework. However, due to their construction properties, the conventional FCRBMs are only capable of performing predictions but are not capable of making classification. Herein, we are proposing *extended* FCRBMs (eFCRBMs), which incorporate a novel classification scheme, to solve this problem. Experiments performed on both artificially generated as well as real-world data demonstrate the effectiveness and efficiency of the proposed technique. We show that eFCRBMs outperform popular approaches including Support Vector Machines, Naive Bayes, AdaBoost, and Gaussian Mixture Models.**

I. Introduction

Detecting and tracking people has proven to be an important problem in various areas such as healthcare, human-computer interaction, energy savings (e.g. light control in unoccupied places), safety (e.g. locating people in emergency situations), robotics, and security [1], [2]. In order to identify the movements and activities of people in various closed environments such, as buildings, various detection methods are used. These methods usually employ sensor measurements and knowledge of the environment. Due to their many facets, probabilistic nature, high level of abstraction and high uncertainties, finding methods that perform well, is still a difficult research task in computer science. This task is even more challenging when low cost and energy efficiency need to be considered in combination. In this paper, people detection is mainly considered in the context of Home and Building Automation (HABA). HABA becomes increasingly complex and provides advanced services, which are usually spread over a broad range of applications. In their simplest form, they are able to solve individual problems such as opening a door, turning the lights off when a room is empty, locking a computer when the user moves away, or air conditioning systems using inputs from proximity sensors such as passive infra-red sensors (PIRs). Actually, these services offer integrated solutions, but a cheap and reliable solution with low power consumption for both detection, and localization is not yet available. Thus, to accurately accomplish the aforementioned goals two main challenges have to be solved: the technological and the logical ones. From the technological point of view, a wide range of sensors is already available on the market, each with its own advantages and disadvantages. Cameras are among the most widespread sensors, but these are often not usable for user tracking, due to privacy constraint and high energy consumption. Another solution is coming from robotics. Laser range scanners are widely used for tasks like localization and position estimation but have been also used for people detection [3] and place classification [4]. Recent approaches to combine visual and laser data for classification and object detection tasks show promising results [3], [5]. Besides that, networks of sensors [6] or fingerprint-based approach [7] have proven to be a good solution for people detection. Still, all of the aforementioned technologies are too expensive to be used in large scale applications. In the scope of these arguments, we are investigating one low-energy and low-cost multi-sensor, made by combining several different sensors, namely Multi-integrated Sensor Technology (MIST1431).

In addition to the technical challenge, user tracking with multiple sensors involves an algorithmic issue too, as proper mathematical models and learning algorithms are required. User tracking could generally be seen as a time series classification problem. A time series is a sequence of observations $\mathbf{x}_t$, taken at regular intervals in time. The goals in time series analysis are: 1) modeling time series, thus obtaining insights into the mechanisms generating such data, 2) forecasting/predicting future value(s) of these variables, and/or 3) classifying time series into corresponding classes of activities. Different algorithms [8], [9] dealing with each sub-problem separately have been proposed. Unfortunately, most of these techniques suffer from difficulties when it comes to complex high dimensional and nonlinear time series data. To remedy such problems, Mnih et.al [10] introduced the Conditional Restricted Boltzmann machine (CRBM), a form of deep learning, as a method for structured output predictions (e.g., time series predictions). Taylor et.al [11] proposed the Factored CRBMs (FCRBMs). FCRBMs preserved the prediction advantages of CRBMs, but are able to predict different types of time series. Apart from predictions, time series classification is also an important challenge in such analysis. Although the prediction efficiency and effectiveness of FCRBMs has been shown in different scenarios [11], these models are not capable of performing classification since its construction was intended for predictions. Therefore, if FCRBMs are to be extended to classification, a novel layer-wise construction is essential.

In this paper, the aim is to create a low cost framework capable to do accurate people detection and localization, while

offering them full privacy. Therefore, we contribute by: 1) using an inexpensive and energy efficient innovative combination of sensors (i.e. MIST1431 and PIR), 2) introducing the novel *extended* FCRBM (eFCRBM). eFCRBMs extend FCRBMs to classification, while retaining FCRBMs' prediction capabilities. Namely, two novel variations to FCRBMs are introduced: a) a layer-wise modification to incorporate *class* labels, and b) an introduction of an additional classification scheme. To test the efficiency and effectiveness of eFCRBMs two sets of experiments were performed. In the first, artificial data was used. In the second experiment, real-world sensory data for human detection and localization were used.

The content of this paper is organized as follows. Section II presents the background knowledge and Section III describes the formulation and derivation of the mathematical models proposed. In Section IV the experimental validation of the methods is shown. Finally, in Section V, conclusions are drawn and recommendations for future research are given.

II. Background

Restricted Boltzmann Machines (RBMs) [12] have been applied in different machine learning fields including, multi-class classification [13], collaborative filtering [14], information retrieval [15], among others. They are energy-based models for unsupervised learning. These models are stochastic with stochastic nodes and layers, making them less vulnerable to local minima [11]. Further, due to their neural configurations, RBMs possess excellent generalization capabilities [16]. Formally, an RBM consists of visible and hidden binary layers. The visible layer represents the data, while the hidden increases the learning capacity by enlarging the class of distributions that can be represented to an arbitrary complexity [11]. Formally, an RBM has an energy function $E(v,h) = -\sum_{i,j} v_i h_j w_{ij} - \sum_i v_i a_i - \sum_j h_j b_j$, where i represents the indices of the visible layer, j those of the hidden layer, and $w_{i,j}$ denotes the weight connection between the i^{th} visible and j^{th} hidden unit. Further, v_i and h_j denote the state of the i^{th} visible and j^{th} hidden unit, respectively, a_i and b_j represent the biases of the visible and hidden layers.

To train such models, Contrastive Divergence (CD) was introduced in [17]. In CD learning follows the gradient of $CD_n \propto D_{KL}(p_0(\mathbf{x})||p_\infty(\mathbf{x})) - D_{KL}(p_n(\mathbf{x})||p_\infty(\mathbf{x}))$, where, $p_n(.)$ is the distribution of a Markov chain running for n steps. Since the visible units are conditionally independent given the hidden units and vice versa, learning can be performed using one step Gibbs sampling, which is carried in two half-steps: 1) update all the hidden units, and 2) update all the visible units. Thus, in CD_n the weight updates are done as follows: $w_{ij}^{\tau+1} = w_{ij}^{\tau} + \alpha\left(\left\langle\langle h_j v_i\rangle_{p(\mathbf{h}|\mathbf{v};\mathbf{W})}\right\rangle_0 - \langle h_j v_i\rangle_n\right)$ where τ is the iteration, α is the learning rate, subscript $_{(n)}$ indicates that the states are obtained after n iterations of Gibbs sampling from the Markov chain starting at $p_0(\cdot)$.

To model time series data and human activities, Taylor and Hinton introduced an extension over RBM, namely Factored Condition Restricted Boltzmann Machine (FCRBM) [11]. FCRBM defines a joint probability distribution over the visible, $\mathbf{v_t}$, and hidden, $\mathbf{h_t}$, neurons, conditioned on the past N observations, $\mathbf{v_{<t}}$, model parameters, $\mathbf{\Theta}$, style and features layers. Although successful, FCRBMs are not capable of performing classification due to their architecture and because it was out of their original scope. The proposed method, explained next, solves this problem.

III. Proposed Method

This section introduces Extended Factored Conditional Restricted Boltzmann Machines (eFCRBMs), shown in Figure 1. Firstly, an intuition describing the model as well as the configuration is discussed. Secondly, eFCRBMs' mathematical details including, the energy function, probabilistic inference, and learning/update rules are detailed. Thirdly, the novel classification procedure is introduced.

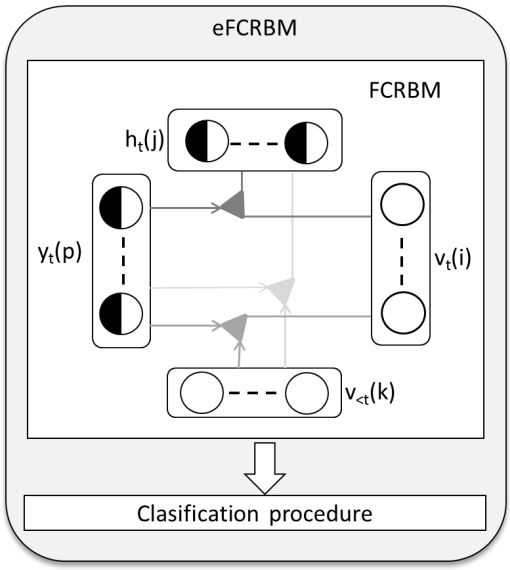

Fig. 1. The general architecture of eFCRBM which include FCRBM (see [11]). The main differences between the proposed model and that of [11] are: 1) replacing the features and style layers with a class layer, and 2) the incorporation of a classification procedure.

A. Configuration & Intuition

To enable classification and predictions in one unified framework, two modifications to FCRBMs [11] are required. Firstly, a joint *class* layer out of both the style and feature layers of an FCRBM is introduced. The second modification is the incorporation of a classification procedure. Based on the learned predictions, the classification procedure ensures accurate segregation to corresponding classes. Next, the details of each of the above steps are described. Due to changes in the FCRBM structure, new mathematical derivations are required.

B. eFCRBM - Mathematical Details

To formalise eFCRBMs, three main ingredients are required. Firstly, an energy function providing scalar values for a given configuration of the network is essential. Secondly, probabilistic inference needs to be detailed. Finally, the update/learning rules required for fitting free parameters have to be derived. Before diving into the mathematical details, however, the notation used in this paper is first introduced. As shown in Figure 1, eFCRBMs consist of four layers: 1) a real valued visible layer v_t, 2) a real valued history layer $v_{<t}$

(i.e., $v_{<t} = v_{t-N:t-1}$, where $N \in \mathbb{N}$), 3) a binary hidden layer h_t, and 4) a binary class layer y_t. Each of the above layers is essential for the success of eFCRBMs. The visible layer encodes the current time step which needs to be predicted. The history, being the basis of such predictions, is encoded on the history layer. The hidden layer guarantees the discovery of important features of the data, while the label layer encodes different classes. To learn the inherent relations between these layers, undirected or directed weights and factors, as shown in Figure 1, are used as connections.

1) Total energy: The total energy, $E(\mathbf{v}_t, \mathbf{h}_t|\mathbf{v}_{<t}, \mathbf{y}_t)$, of eFCRBMs can be used to: 1) define joint and conditional distributions as shown in Section III-B2, and 2) derive update/learning rules of the free parameters (see Section III-B3). This function is defined as the sum of the first and third order energy terms:

$$E(v_t,h_t|v_{<t},y_t) = \overbrace{\frac{1}{2}\sum_{i=1}^{n_1}(v_{i,t}-\hat{a}_{i,t})^2 - \sum_{j=1}^{n_2}\hat{b}_{j,t}h_{j,t}}^{\text{first order terms}} \quad (1)$$
$$\underbrace{-\sum_{f=1}^{F}\left[\sum_{i=1}^{n_1}\left[\sum_{j=1}^{n_2}\left[\sum_{p=1}^{n_3}W_{if}^{v}W_{jf}^{h}W_{pf}^{y}v_{i,t}h_{j,t}y_{p,t}\right]\right]\right]}_{\text{third order terms}}$$

where F, n_1, n_2, and n_3, represent the total number of factors, and the number of units in each of the visible, hidden, and label layers, respectively. Moreover, each vertex of the factors is the product of the weighted sums at the other two vertices. Therefore, W_{if}^{v}, W_{jf}^{h}, and W_{pf}^{y} connect each of the present, hidden, and label units to the factors, respectively. The terms $\hat{a}_{i,t}$ and $\hat{b}_{j,t}$ are called dynamic biases, which are defined as:

$$\begin{aligned}\hat{a}_{i,t} &= a_i + \sum_{m}\sum_{k,p}A_{im}^{v}A_{km}^{v<t}A_{pm}^{y}v_{k,<t}y_{p,t} \quad (2)\\ &= a_i + \sum_{m}A_{im}^{v}\sum_{k}A_{km}^{v<t}v_{k,<t}\sum_{p}A_{pm}^{y}y_{p,t}\\ \hat{b}_{j,t} &= b_j + \sum_{n}\sum_{k,l}B_{jn}^{h}B_{kn}^{v<t}B_{p,n}^{y}v_{k,<t}y_{p,t} \quad (3)\\ &= b_j + \sum_{n}B_{jn}^{h}\sum_{k}B_{kn}^{v<t}v_{k,<t}\sum_{p}B_{pn}^{y}y_{p,t}\end{aligned}$$

with A_{im}^{v}, $A_{km}^{v<t}$, A_{pm}^{y}, B_{jn}^{h}, $B_{kn}^{v<t}$, B_{pn}^{y}, being the connections between biases factors and layers.

2) Probabilistic Inference: In eFCRBMs probabilistic inference means determining two conditional distributions. The first is the probability of the hidden layer conditioned on all the other layers (i.e., $p(h_{j,t}|\mathbf{v}_t, \mathbf{v}_{<t}, \mathbf{y}_t)$), while the second is the probability of the present layer conditioned on the others (i.e., $p(v_{i,t}|\mathbf{h}_t, \mathbf{v}_{<t}, \mathbf{y}_t)$). Since there are no connections between the neurons in the same layer, inference can be done in parallel for each unit type, leading to:

a) Probability of the hidden neurons: $p(h_{j,t} = 1|\mathbf{v}_t, \mathbf{v}_{<t}, \mathbf{y}_t)$ is given by a sigmoid function evaluated on the total input to each hidden unit via the factors. Formally: $p(h_{j,t} = 1|\mathbf{v}_t, \mathbf{v}_{<t}, \mathbf{y}_t) = \text{sig}\Big(\hat{b}_{j,t} + \sum_{f}W_{if}^{h}\sum_{i}W_{if}^{v}v_{i,t}\sum_{p}W_{pf}^{y}y_{p,t}\Big)$ where $\text{sig}(x) = 1/1 + \exp(-x)$.

b) Probability of the visible neurons: $p(v_{i,t}|\mathbf{h}_t, \mathbf{v}_{<t}, \mathbf{y}_t)$ is given by a Gaussian distribution over the total input to each visible unit via the factors: $p(v_{i,t}|\mathbf{h}_t, \mathbf{v}_{<t}, \mathbf{y}_t) = \mathcal{N}\Big(\hat{a}_{i,t} + \sum_{f}W_{if}^{v}\sum_{j}W_{jf}^{h}h_{j,t}\sum_{p}W_{pf}^{y}y_{p,t}, 1\Big)$

3) Learning & Update Rules: The model free parameters (i.e., dynamical biases and weights) are learned using CD, discussed previously. The update rules for each of the matrices can be computed by deriving the energy function with respect to each of the variables, yielding:

$$\begin{aligned}\triangle W_{if}^{v} &\propto \sum_{t}\Big(\langle v_{i,t}\sum_{j}W_{jf}^{h}h_{j,t}\sum_{p}W_{pf}^{y}y_{p,t}\rangle_{data}\\ &\quad - \langle v_{i,t}\sum_{j}W_{jf}^{h}h_{j,t}\sum_{p}W_{pf}^{y}y_{p,t}\rangle_{recon}\Big)\\ \triangle W_{jf}^{h} &\propto \sum_{t}\Big(\langle h_{j,t}\sum_{i}W_{if}^{v}v_{i,t}\sum_{p}W_{pf}^{y}y_{p,t}\rangle_{data}\\ &\quad - \langle h_{j,t}\sum_{i}W_{if}^{v}v_{i,t}\sum_{p}W_{pf}^{y}y_{p,t}\rangle_{recon}\Big)\\ \triangle W_{pf}^{y} &\propto \sum_{t}\Big(\langle y_{p,t}\sum_{i}W_{if}^{v}v_{i,t}\sum_{j}W_{jf}^{h}h_{j,t}\rangle_{data}\\ &\quad - \langle y_{p,t}\sum_{i}W_{if}^{v}v_{i,t}\sum_{j}W_{jf}^{h}h_{j,t}\rangle_{recon}\Big)\\ \triangle A_{im}^{v} &\propto \sum_{t}\Big(\langle v_{i,t}\sum_{k}A_{km}^{v<t}v_{k,<t}\sum_{p}A_{pm}^{y}y_{p,t}\rangle_{data}\\ &\quad - \langle v_{i,t}\sum_{k}A_{km}^{v<t}v_{k,<t}\sum_{p}A_{pm}^{y}y_{p,t}\rangle_{recon}\Big)\\ \triangle A_{km}^{v<t} &\propto \sum_{t}\Big(\langle v_{k,<t}\sum_{i}A_{im}^{v}v_{i,t}\sum_{p}A_{pm}^{y}y_{p,t}\rangle_{data}\\ &\quad - \langle v_{k,<t}\sum_{i}A_{im}^{v}v_{i,t}\sum_{p}A_{pm}^{y}y_{p,t}\rangle_{recon}\Big)\\ \triangle A_{pm}^{y} &\propto \sum_{t}\Big(\langle y_{p,t}\sum_{i}A_{im}^{v}v_{i,t}\sum_{k}A_{km}^{v<t}v_{k,<t}\rangle_{data}\\ &\quad - \langle y_{p,t}\sum_{i}A_{im}^{v}v_{i,t}\sum_{k}A_{km}^{v<t}v_{k,<t}\rangle_{recon}\Big)\\ \triangle B_{jn}^{h} &\propto \sum_{t}\Big(\langle h_{j,t}\sum_{k}B_{kn}^{v<t}v_{k,<t}\sum_{p}B_{pn}^{y}y_{p,t}\rangle_{data}\\ &\quad - \langle h_{j,t}\sum_{k}B_{kn}^{v<t}v_{k,<t}\sum_{p}B_{pn}^{y}y_{p,t}\rangle_{recon}\Big)\\ \triangle B_{kn}^{v<t} &\propto \sum_{t}\Big(\langle v_{k,<t}\sum_{j}B_{jn}^{h}h_{j,t}\sum_{p}B_{pn}^{y}y_{p,t}\rangle_{data}\\ &\quad - \langle v_{k,<t}\sum_{j}B_{jn}^{h}h_{j,t}\sum_{p}B_{pn}^{y}y_{p,t}\rangle_{recon}\Big)\\ \triangle B_{pn}^{y} &\propto \sum_{t}\Big(\langle y_{p,t}\sum_{j}B_{jn}^{h}h_{j,t}\sum_{k}B_{kn}^{v<t}v_{k,<t}\rangle_{data}\\ &\quad - \langle y_{p,t}\sum_{j}B_{jn}^{h}h_{j,t}\sum_{k}B_{kn}^{v<t}v_{k,<t}\rangle_{recon}\Big)\end{aligned}$$

C. Classification procedure

Classification is based on the predicted values of eFCRBMs. The intuition behind it, derives from the following reasoning: if a trained FCRBM model is capable of predicting the next values for different types of time series, using different labels and history windows associated with each type, then, if a wrong label is used for predictions, these will be far away from the true values, compared with the situation when the good label is used for the same purpose. More exactly, the main idea is first to fix the history layer to an arbitrary instance from the test data set. Predictions of the present frame using all possible classes are then performed. Finally, these predictions are compared with the true value TV_t of the present frame for that specific instance. To find the prediction closest to the *real*-true values, a similarity or distance measure is then adopted. The class which made the closest prediction is chosen to be the class for that instance. Formally, this can be written as:

$$\mathcal{Y} = \arg\min_{y \in \mathbf{Y}} \Big[\mathrm{d}\Big(\mathrm{PV}(v_{<t}, y), \mathrm{TV}_t \Big) \Big] \tag{4}$$

where, $d(\cdot,\cdot)$ is a distance measure, $\mathcal{Y}$ is the foretold class, $\mathbf{Y}$ the set of all classes, $\mathrm{PV}(v_{<t}, y)$ represents the predicted value of the present frame, based on the history $v_{<t}$ and the class label y and, TV_t denotes the true value of the present frame.

IV. Experiments & Results

We have assessed our approaches in two sets of experiments. In the first one, artificial data generated from relatively complex trigonometric functions was used to assess the performance of eFCRBM. The goal in the second set of experiments was the detection and localization of humans through data gathered from real sensory measurements. In both cases eFCRBMs were evaluated and compared to four widely known classification methods. Namely, Support Vector Machines (SVMs), Naive Bayes (NB), AdaBoost (AB), and Gaussian Mixture Models (GMMs) were used as comparison benchmarks. In all experiments, the methods were tested using 10 fold cross validation, and the aggregated data was separated to training and test datasets.

A. Artificial data

In this experiment the goal was to classify data points arriving from either $f_1 = t\sin(t^2)$, or $f_2 = \frac{1}{2}t\cos(t)$, as shown in Figure 2. The data set was generated by evaluating each of the function in $t = \{1, \ldots, 750\}$. eFCRBMs included 5 hidden neurons, 5 factors, 3 history frame neurons, and 1 visible neuron. The collected dataset was split into 66% for training and 33% for testing. Namely, 500 instances were used for training and the remaining 250 were used for testing. Classification results reported in Table I clearly manifest that eFCRBM outperforms state-of-the-art techniques including SVM with radial basis function kernels.

TABLE I. Results showing Accuracy (A), Precision (P), Recall (R) and Specificity (S) for each of SVM, Naive Bayes, AdaBoost, GMM and eFCRBM.

	SVM	NB	AB	GMM	eFCRBM
A	58.20%	61.40%	59.42%	61.20%	**80.52%**
P	55.46%	54.66%	58.76%	57.78%	**100%**
R	59.42%	84.4%	76.40%	87.60%	61.13%
S	33.20%	30%	46.40%	36%	**100%**

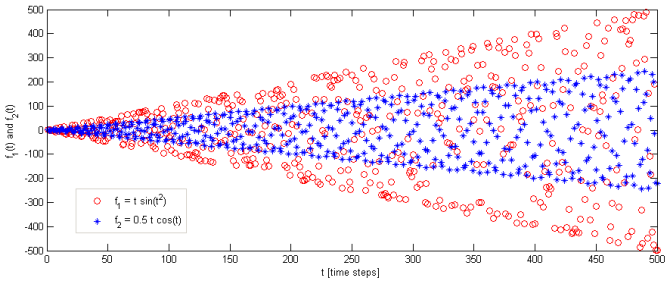

Fig. 2. The first five hundred values generated using the two divergent functions f_1 and f_2.

B. People Detection and Localization

1) User tracking using MIST1431: In this set of experiments real-world data collected using the low-cost and low-energy multi-sensor, MIST1431, has been used. The goal was the detection and localization of humans in various positions in an environment using just one device. All the calculation needed and the eFCRBM implementation were performed using an external computer after the data were acquired from the sensor. For these experiments, the settings for eFCRBM were: 30 hidden neurons, 30 factors, 30 history frames, and the number of visible neurons was set to the number of MIST1431 outputs. The number of labels neurons y_t was set

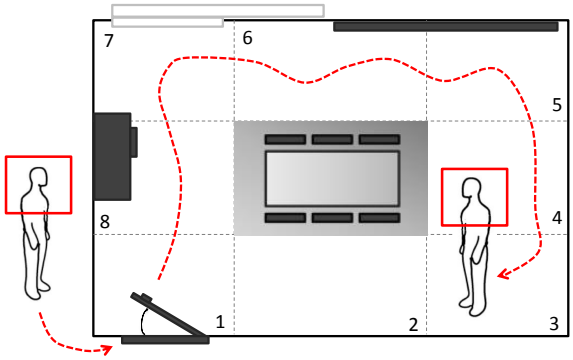

Fig. 3. Experimental design: A room was split into eight locations. The sensor was placed either at a table or at the ceiling in the middle of the room.

to the number of classes. For each class just 1 neuron was active and the others were set to 0. The distance measure used in the "Classification Procedure" was the Euclidean distance. All MIST1431 measurements were acquired with a sampling rate of T_s=2 seconds. Each gathered sample contained 12 outputs, where the first six signals corresponded to un-filtered ultra violet color, and the second six to the near-infra red part of the spectrum. In what comes next, each set of six signals is denoted by AL_1 and AL_2. MIST1431 also includes sensors allowing for the detection of humidity and temperature. In our experiments, these additional outputs were also taken into account to improve the accuracy of the classifier. To achieve the goal of human detection and localization, a general experimentation protocol was designed. A room of 4.8×3 m, was split into eight possible positions, corresponding to four corner and wall mid-points (see Fig. 3). Three situations were of major interest: 1) a person moving (i.e., M_i with $i = \{1, \ldots, 8\}$) in one of the eight positions, 2) a person standing still (i.e., S_i with $i = \{1, \ldots, 8\}$) in one of the

eight positions, or 3) the room is empty (i.e., E). Three scenarios were then recorded. Major difference between these three recording scenarios, are detailed next and summarized in Table II.

TABLE II. MAJOR DIFFERENCES BETWEEN THE THREE SCENARIOS CONSIDERED FOR HUMAN DETECTION AND LOCALIZATION.

	S1	S2	S3
No. of data	3878	4891	4124
Lamp	On	On	Of
MIST1431 position	ceil	ceil	table
MIST1431 with pinhole	yes	yes	no
No. of crossing the room	1	2	2
Time spent in one position	5 min	3 min	$\leq$3 min

Scenario 1: In this scenario, the room starts in an empty state. A person enters the room, where he/she moves for 5 minutes in one of the eight positions. He/she then stands still for 5 minutes in that same position. After that a transition to an unvisited neighboring position is performed. The procedure is repeated until all eight positions in the room were visited. Fig. 4, summarises the time response of the ambient light sensors (AL_2) in this scenario.

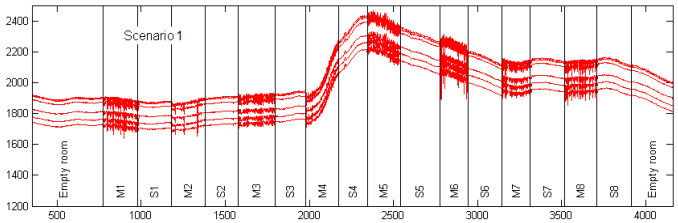

Fig. 4. Time response of the ambient light sensors (AL_2) integrate in MIST1431 for Scenario1

Scenario 2: In this scenario, a person crosses all locations in the room twice with a break of 10 minutes. Firstly, all these are crossed in an anti-clockwise direction, while in the second they are crossed in a clockwise fashion. Fig. 5, shows all outputs of MIST1431 sensor, including ambient light sensor, temperature ($^{\circ}C$) and relative humidity ($\%$).

Scenario 3: This scenario is similar to the previous in which a person is crossing all the locations in the room twice. Here, however, crossing the locations was in the anti-clockwise direction. It is worth highlighting, that in this scenario the lamp was off in the room, and the sun light had a bigger influence on the MIST1431 outputs. More exactly, as it can be seen in Fig. 6, the sensor's outputs have a descending trend. This is naturally explainable because the recording took place in the late afternoon and it lasted for about two hours. Motion detection and human localization was carefully analyzed in all previous scenarios. The first step was to investigate whether SVMs, Naive Bayes, AdaBoost, Gaussian Mixture Models and eFCRBM are suitable enough for the purposes of human detection and localization based on the gathered data. Corresponding to each of the eight positions a person can perform two types of activities, moving or standing still. Results obtained by applying the aforementioned techniques are shown in the "No preprocessing" column of Table III. Given that all models did not achieve a high accuracy, a preprocessing step was performed. Namely, the absolute differences between the values acquired from the light sensors of MIST1431 were

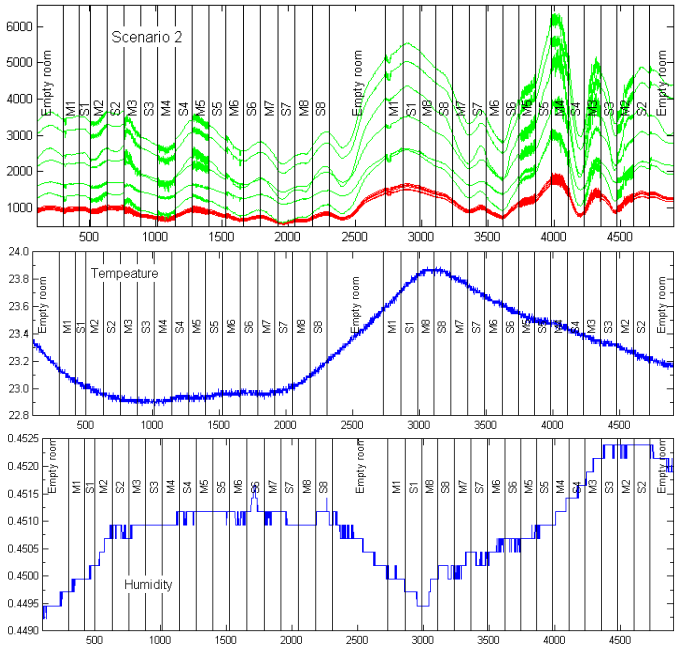

Fig. 5. Time response of all sensor integreted in MIST1431, AL_1 (green) and AL_2 (red), Humidity[RH] and Temperature [T]) for Scenario2

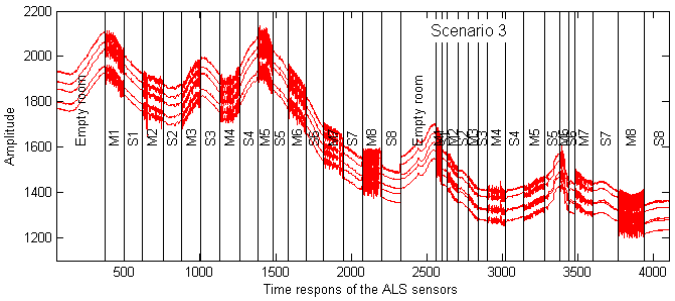

Fig. 6. Time response of the ambient light sensors (AL_2) integrate in MIST1431 for Scenario3

used. Results presented, in the "Preprocessing" column of Table III, show an increase in the classification accuracy for all techniques used.

2) User tracking using MIST1431 and PIR: Although successful, the previous localization and detection techniques still suffered from the following two problems: 1) inhibition of detection in the absence of ambient light, and 2) inaccuracies when it comes to people close to the display. Aiming at enhancing the quality of such estimates as well as at increasing the accuracy of localization, a method of fusing information from thermal and visible light sensors has been developed. The reason to add a thermal sensor is because the thermal images are not affected by lighting or shadowing and are not overtly affected by smoke, dust or unstable backgrounds. Moreover, the thermal sensor allows to obtain results overnight. The method relies on motion detection in both signals. Namely, the ambient light sensor, MIST1431, is only considered after a passive infrared sensor (PIR) detects motion. Fig.7 illustrates the spatial response in time of the ambient light sensors (AL_2). We recorded data using MIST1431 sensor, mounted on the ceiling and a PIR with Fresnel lens. The first step,

TABLE III. COMPARISON BETWEEN SVM, NAIVE BAYES, ADABOOST, GAUSSIAN MIXTURE MODELS, AND EFCRBM IN TERMS OF ACCURACY.

	Localization- 17 clases (Empty room, 8 moving position and 8 sitting position)									
	No preprocessing					**Preprocessing**				
	SVM	NB	AdaB.	GMM	eFCRBM	SVM	NB	AdaB.	GMM	eFCRBM
Scenario 1	49.77%	41.29%	24.70%	52.26%	56.27%	70.45%	70.98%	30.15%	74.92%	76.37%
Scenario 2	39.90%	18.57%	20.43%	42.39%	53.65%	48.71%	44.28%	25.04%	57.04%	60.09%
Scenario 3	53.24%	22.29%	31.10%	58.92%	61.23%	64.77%	42.21%	28.54%	65.34%	68.36%

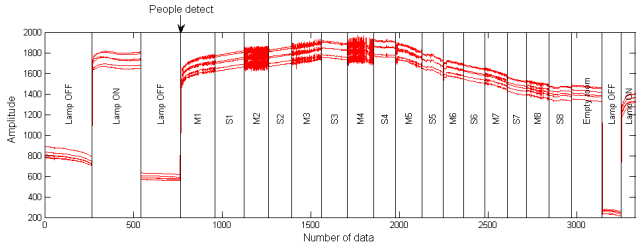

Fig. 7. Measured MIST1431 output after it is triggered by PIR.

PIR detect movement, is accomplished by using a threshold $PIR > 3 \cdot \sigma$, where σ is the standard deviation, calculated using data recordings when the room is empty. Having detected human presence, the second step consisted of two phases: i) localize, and ii) classify the activity being done. Table IV shows the accuracy of the methods considering 16 classes, corresponding to: one person moving, M_i, and one person sitting in the room, S_i, observed in all positions $i = 1, ..., 8$. More specifically, in this experiment we classified 2197 data points in 16 classes. Notable, the proposed approach is capable

TABLE IV. COMPARISON BETWEEN SVM, NAIVE BAYES, ADABOOST, GMM AND EFCRBM IN TERMS OF ACCURACY.

Localization-16 clases (8 moving position and 8 sitting position)				
SVM	NB	AdaB.	GMM	eFCRBM
71.56%	64.81%	40.21%	74.07%	**88.72%**

of distinguishing not only when a person is situated in a specific area of the room, with a surface around 1.2 m^2, but also to detect the level of motion for that person. The highest accuracy achieved is by using the combination between MIST1431, PIR and eFCRBM. More exactly, this combination has an accuracy of 88.72%, being at least 10% higher than in any other combination.

V. CONCLUSION.

In this paper we are proposing a novel framework capable to accurately detect and localize people in a room, including their level of motion. The framework contributes on two main directions. The first is a technological one and consists in using a combination of MIST1431 and PIR, two low-cost and low-energy sensors. The second direction is theoretical one, consisting in the introduction of a novel classification method for time series, namely Extended Factored Conditional Restricted Boltzmann Machines. This new technique builds on FCRBMs by incorporating a label layer and a classification procedure. Artificial as well as real-world experiments clearly demonstrate the effectiveness of the proposed technique. Namely, eFCRBMs were capable of outperforming each of SVMs, GMMs, AdaBoost, and Naive Bayes classifiers, in all the tested scenarios. As further research directions, we are intending to investigate other combinations of low-cost and low-energy sensors together with eFCRBM to be able to distinguish not just the level of motion in a room specific location, but also to differentiate between different types of human activities.

REFERENCES

[1] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," *ACM Comput. Surv.*, vol. 38, no. 4, Dec. 2006.

[2] T. Teixeira, G. Dublon, and A. Savvides, "A survey of human-sensing: Methods for detecting presence, count, location, track, and identity," *ACM Computing Surveys*, vol. 5, 2010.

[3] K. O. Arras, . M. Mozos, and W. Burgard, "Using boosted features for the detection of people in 2d range data." in *ICRA*. IEEE, 2007, pp. 3402–3407.

[4] O. Mozos, C. Stachniss, and W. Burgard, "Supervised learning of places from range data using adaboost," in *Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on*, April 2005, pp. 1730–1735.

[5] Z. Zivkovic and B. Krose, "Part based people detection using 2D range data and images," in *Proceedings of the IEEERSJ International Conference on Intelligent Robots and Systems*, San Diego, USA, 2007.

[6] D. Lymberopoulos, A. Bamis, T. Teixeira, and A. Savvides, "Behaviorscope: Real-time remote human monitoring using sensor networks," in *Information Processing in Sensor Networks, 2008. IPSN '08. International Conference on*, April 2008, pp. 533–534.

[7] A. Kokkinis, M. Raspopoulos, L. Kanaris, A. Liotta, and S. Stavrou, "Map-aided fingerprint-based indoor positioning," in *Personal Indoor and Mobile Radio Communications (PIMRC), 2013 IEEE 24th International Symposium on*, Sept 2013, pp. 270–274.

[8] H. Deng, G. Runger, E. Tuv, and M. Vladimir, "A time series forest for classification and feature extraction," *Information Sciences*, vol. 239, no. 0, pp. 142 – 153, 2013.

[9] X. Xi, E. Keogh, C. Shelton, L. Wei, and C. A. Ratanamahatana, "Fast time series classification using numerosity reduction," in *Proceedings of the 23rd International Conference on Machine Learning*, ser. ICML '06, 2006, pp. 1033–1040.

[10] V. Mnih, H. Larochelle, and G. Hinton, "Conditional restricted boltzmann machines for structured output prediction," in *Proceedings of the International Conference on Uncertainty in Artificial Intelligence*, 2011.

[11] G. W. Taylor, G. E. Hinton, and S. T. Roweis, "Two distributed-state models for generating high-dimensional time series," *Journal of Machine Learning Research*, vol. 12, pp. 1025–1068, 2011.

[12] P. Smolensky, "Information processing in dynamical systems: Foundations of harmony theory," in *Parallel Distributed Processing: Volume 1: Foundations*, D. E. Rumelhart, J. L. McClelland *et al.*, Eds. Cambridge: MIT Press, 1987, pp. 194–281.

[13] H. Larochelle and Y. Bengio, "Classification using discriminative restricted Boltzmann machines," 2008, pp. 536–543.

[14] R. Salakhutdinov, A. Mnih, and G. Hinton, "Restricted boltzmann machines for collaborative filtering," in *In Machine Learning, Proceedings of the Twenty-fourth International Conference (ICML 2004). ACM*. AAAI Press, 2007, pp. 791–798.

[15] P. V. Gehler, A. D. Holub, and M. Welling, "The rate adapting poisson model for information retrieval and object recognition," in *In Proceedings of 23rd International Conference on Machine Learning (ICML06*. ACM Press, 2006, p. 2006.

[16] Y. Bengio, "Learning deep architectures for ai," *Found. Trends Mach. Learn.*, vol. 2, no. 1, pp. 1–127, Jan. 2009.

[17] G. E. Hinton, "Training Products of Experts by Minimizing Contrastive Divergence," *Neural Computation*, vol. 14, no. 8, pp. 1771–1800, Aug. 2002.