

Recognition of Anomalous Motion Patterns in Urban Surveillance

Maria Andersson, Fredrik Gustafsson, Louis St-Laurent and Donald Prevost

Linköping University Post Print



N.B.: When citing this work, cite the original article.

©2013 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

Maria Andersson, Fredrik Gustafsson, Louis St-Laurent and Donald Prevost, Recognition of Anomalous Motion Patterns in Urban Surveillance, 2013, IEEE Journal on Selected Topics in Signal Processing, (7), 1, 102-110.

<http://dx.doi.org/10.1109/JSTSP.2013.2237882>

Postprint available at: Linköping University Electronic Press

<http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-93983>

Recognition of Anomalous Motion Patterns in Urban Surveillance

Maria Andersson, Fredrik Gustafsson, Louis St-Laurent and Donald Prévost

Abstract— We investigate the unsupervised K -means clustering and the semi-supervised hidden Markov model (HMM) to automatically detect anomalous motion patterns in groups of people (crowds). Anomalous motion patterns are typically people merging into a dense group, followed by disturbances or threatening situations within the group. The application of K -means clustering and HMM are illustrated with datasets from four surveillance scenarios. The results indicate that by investigating the group of people in a systematic way with different K values, analyze cluster density, cluster quality and changes in cluster shape we can automatically detect anomalous motion patterns. The results correspond well with the events in the datasets. The results also indicate that very accurate detections of the people in the dense group would not be necessary. The clustering and HMM results will be very much the same also with some increased uncertainty in the detections.

Index Terms— clustering algorithms, decision support systems, hidden Markov models, machine learning, machine vision, object segmentation, pattern recognition

I. INTRODUCTION

During the last years automatic crowd analysis has been studied for various applications including visual surveillance, crowd management and public space design. In visual surveillance crowd analysis is used for automatic detection of anomalies or threatening events. In crowd management crowd analysis is used to analyze sport events, large concerts and public demonstrations to avoid crowd related disasters. For public space design crowd analysis is used to provide guidelines for the design of shopping malls, city centers, etc [1].

With automatic crowd analysis it is possible to foresee different states of the crowd, including crowd size, crowd density, crowd flow, crowd speed and anomalous motion patterns (e.g. riots, robberies and fights). Automatic crowd analysis can improve the possibilities for an operator to detect, at an earlier stage, important events in the often very large amount of information from sensor data. The outcome of threatening and dangerous situations can then be mitigated or even avoided.

In dense environments occlusion is a problem. People will temporarily be hidden and cannot be continuously tracked. In traffic and pedestrian monitoring crowd analysis has been divided into three approaches to be able to better handle the effects from occlusion. They are microscopic approach, macroscopic approach and a combination of the two [2]. In the microscopic approach people are analyzed as discrete individuals. This information is summarized to obtain knowledge about the crowd. In the macroscopic approach the crowd is instead analyzed as a single unit. There is no information used on position estimates of individuals, which is a way of avoiding the problems with occlusion. A combination of micro- and macroscopic approaches can be made by keeping the crowd as a homogeneous mass, but at the same time considering an internal force. Another way is by keeping the characters of the persons while maintaining a general view of the entire crowd.

A combination is proposed in [3] where the aim is to understand group motion patterns in subway stations. Detection and tracking of individuals together with group tracking form a basis for group motion patterns analysis.

In [2] and [4] optical flow from the crowd movements is used to detect abnormal regions in the image. In [4] an HMM is used to interpret the optical flow. A macroscopic approach is proposed by [5] where optical flow and foreground regions are used to derive crowd features in the image. The different crowds features are fused by an HMM to obtain a final decision. In [6] optical flow, foreground regions and sound level are fused by an HMM to detect abnormal crowd events, while still regarding the crowd as a single unit.

In this paper we propose to use K -means clustering and HMM to detect anomalous motion patterns in dense crowds. The anomalous motion patterns are based on merging and splitting of groups as well as internal interactions within dense groups.

Tracking of dense groups, as well as merging and splitting, have been discussed also in another paper [7] where the authors use a dynamic Gaussian mixture model to describe the dynamics of the clusters and a measure for target concentration that is based on the probability hypothesis density (PHD) filter. Merging and splitting are described by using a point process formulation. Internal activities, such as fights, are not investigated. Group analysis (including group tracking) have been investigated in several papers, and also for other types of applications, see for example [8]-[11].

The objective of this paper is to study unsupervised K -means clustering and semi-supervised HMM for detection of anomalous motion patterns in crowds. For crowd surveillance in urban environments the context is important, e.g. time of day, time of week, time of year and weather conditions. However, if the anomaly detection algorithm is conditioned on the context to a large extent, the algorithm can often be used only in that specific context. The less prior information we

Manuscript received August 1, 2012. This work was supported in part by Vinnova (Swedish Governmental Agency for Innovation Systems) under the VINNMER programme.

Maria Andersson is with the Swedish Defence Research Agency, SE-581 11 Linköping, Sweden (phone: +46 13 378407; fax: +46 13 378287; e-mail: maria.andersson@foi.se).

Fredrik Gustafsson is with the Electrical Engineering Department, Linköping University, SE-581 83 Linköping, Sweden (e-mail: fredrik@isy.liu.se).

Louis St-Laurent and Donald Prévost are with INO, Quebec, Canada (e-mails: louis.st-laurent@ino.ca, donald.prévost@ino.ca).

can use the more generic can the algorithm become. With K -means clustering and HMM the aim is to minimize the amount of prior information, while still getting enough information from the sensor data.

We will assume that a reliable detection algorithm of people in raw images already exists in this work, and focus on anomalies in groups. The detection of people can be done with for example face detection [12] or head detection [13], using standard video surveillance cameras. An alternative is to use a combination of thermal infrared and visual cameras and fuse the results from the respective detection algorithms [14].

In [15] we made an initial study on the use of K -means clustering and HMM. In that paper we assumed only one group in the scene. In this paper we do not have any prior assumptions of the number of groups. Instead we have developed a procedure where we can estimate the most relevant number of groups and can select the most relevant input data to the motion patterns analysis. We evaluate the algorithms more rigorously on four datasets, including three different sensitivity analyses.

The paper is organized as follows. Section II presents K -means clustering and Section III presents HMM. In Section IV we discuss the application of the two algorithms on the detection of anomalous motion patterns in crowds. In Section V the algorithms are tested on data from four recorded scenarios that include smaller crowds (here denoted groups of people to distinguish them from large crowds). Section VI presents a sensitivity analysis. Section VII finally presents some conclusions.

II. K -MEANS CLUSTERING

Cluster analysis is used for segmenting a collection of objects into clusters, based on information found in the data. The data describes the objects as well as their relationships. Within each cluster the members are more closely related to each other than to cluster members who belong to other clusters [16]. A similarity measure is used to estimate the closeness of cluster members in each cluster. In this case we use the squared Euclidean distance.

The basic steps in the algorithm are: 1) select K points as initial centroids; 2) form K clusters by assigning each point to its closest centroid; 3) re-compute the centroid of each cluster; and 4) repeat 2 to 3 until the centroids do not change anymore.

The purpose is to obtain as well-separated clusters as possible. A measure that indicates how well-separated the clusters are, is the so-called silhouette s_i [17]. The silhouette is based on distances between a certain cluster member to its own cluster as well as to other cluster. s_i indicates what cluster members that lie well within their cluster, and what cluster members that lie somewhere in between clusters. The average silhouette representing all clusters for a given K provides a measure of the clustering quality that can be used to select the appropriate number of clusters. The silhouette ranges from $-1 \leq s_i \leq 1$. If $s_i = 1$ the cluster member has most likely been associated to the right cluster. If $s_i = 0$ the cluster member could as well belong to another cluster. If $s_i = -1$ the cluster member has most likely been associated to the wrong cluster.

We denote the average silhouette of cluster l at time t as $s_{l,t,A}$ and use its standard deviation $\sigma_{l,t}$ to illustrate the robustness of the silhouette.

Other output data include the centroid coordinates, the number of cluster members and the sum of the Euclidean distances between each cluster member and the corresponding centroid. The sum of Euclidean distances is used as a basis for describing the cluster density and this is described in more detail Section IV.B.

III. HMM

The HMM [18] is a machine learning algorithm that has been used for pattern recognition in many different applications, e.g. speech recognition, text recognition and motion recognition. The algorithm consists of two stochastic processes. The underlying (hidden) process can not be observed directly, but indirectly through a second stochastic process which produces sequences of observations. The states represent some unobservable condition of a system. The HMM (λ) is defined by the parameters $\lambda = (A, B, \pi, N, M)$.

The number of hidden states in the system is N . The individual states are denoted $S = \{S_1, S_2, \dots, S_N\}$ and the state at time t is denoted q_t . The number of distinct observation symbols in each state is M . The observation symbols represent a physical output from the system. The individual observation symbols are denoted $V = \{v_1, v_2, \dots, v_M\}$. The state transition probability distribution is $A = \{a_{ij}\}$, where

$$a_{ij} = P[q_{t+1} = S_j \mid q_t = S_i] \quad (1)$$

and $1 \leq i, j \leq N$. The observation symbol probability distribution in state j is $B = \{b_j(k)\}$ where

$$b_j(k) = P[v_{k,t} \mid q_t = S_j] \quad (2)$$

and $1 \leq j \leq N, 1 \leq k \leq M$. The initial state distribution is denoted $\pi = \{\pi_i\}$ and

$$\pi_i = P[q_1 = S_i]. \quad (3)$$

The observation sequence O , representing some physical output from the system, is denoted

$$O = (O_1, O_2, \dots, O_T). \quad (4)$$

Each observation symbol O_t is one of the symbols from V , and T is the number of observations in the sequence. Given an observation sequence O and a model $\lambda = (A, B, \pi, N, M)$ we can compute the likelihood of O given the model, i.e. $P(O|\lambda)$. The likelihood is calculated using the forward-backward procedure [18], where the forward variable $\alpha_t(i)$ is defined as:

$$\alpha_t(i) = P(O_1, O_2 \dots O_t = S_i \mid \lambda), \quad (5)$$

and describes the probability of the partial observation sequence (O_1, O_2, \dots, O_t) and state S_i at time t , given the model λ . To solve $\alpha_t(i)$ we use the following steps:

$$\alpha_1(i) = \pi_i b_i(O_1), \quad (6)$$

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^N \alpha_t(i) a_{ij} \right] b_j(O_{t+1}), \quad (7)$$

$$P(O | \lambda) = \sum_{i=1}^N \alpha_T(i), \quad (8)$$

and $1 < t < T - 1$. To be able to compute $\alpha_t(i)$ for very small numbers a scaling factor C_t is introduced which finally leads to the expressions [18]:

$$C_t = \frac{1}{\sum_{j=1}^N \alpha_t(j)}, \quad (9)$$

$$\log[P(O|\lambda)] = -\sum_{t=1}^T \log C_t. \quad (10)$$

The unknown parameters A , B and π are obtained by using the Baum-Welch algorithm [16]. This algorithm uses an iterative expectation-maximization (EM) procedure, given initial parameters for A , B and π and a set of training data.

Based on the data for normal motion patterns we define a threshold D_{HMM} that distinguishes approximately normal from anomalous motion patterns. D_{HMM} is based on the mean log-likelihood value $\log[P(O|\lambda)]_{mean}$ for normal motion patterns and its standard deviation σ , such as :

$$D_{HMM} = \log[P(O|\lambda)]_{mean} - 3 \times \sigma. \quad (11)$$

The HMM can be used for both supervised and semi-supervised anomaly detection. For supervised anomaly detection the HMM is trained on data from both normal situations and different abnormal situations. For semi-supervised anomaly detection the HMM is trained only on data from normal situations. An abnormal situation is recognized as a deviation from the expected normal situation, but it is not possible to recognize (classify) the type of abnormal event.

Input data to the HMM are based on results from the K -means clustering. In this way we use an unsupervised algorithm to create input data to the HMM, which will make the HMM less dependent on the context.

IV. TRACKING PERSONS AND GROUPS AND ESTIMATING GROUP BEHAVIOR

The overall procedure for detection of anomalous motion patterns is illustrated in Fig. 1. To start with the people in the

scene are detected. The detections are input data to the clustering. The clustering decides at every time step the current number of clusters, where each cluster can be a single person or a group of people.

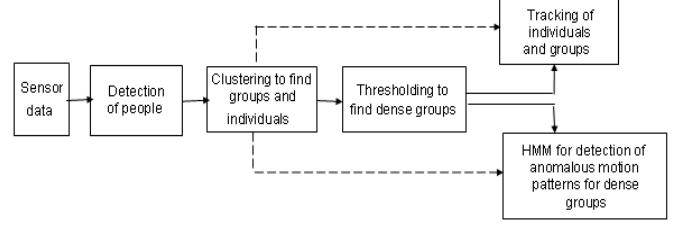


Fig. 1. The overall procedure for detection of anomalous motion patterns.

With a thresholding procedure (based on a crowd density measure) we can detect dense groups. In the next step we analyse the motion patterns of dense groups, using an HMM with input data from the thresholding and clustering. The motion patterns are described by changes in the cluster shape.

Tracking is then applied for all the clusters. If the cluster represents a group the cluster centroid will be the basis for the group tracking. If the cluster represents one person the corresponding detection will be the basis for the tracking. The influence of increased activities on group tracking, and the ability of the tracking to perform enough tracking accuracy, is briefly discussed in [15]. In this paper we focus on the steps for clustering, thresholding and motion patterns analysis.

A. Detection in raw images

For the detection of people in raw images foreground-background segmentation is used, which provides a set of detected people represented with ellipsoids. Each ellipsoid is represented with a center coordinate z_r^p in the image plane, and with ‘‘covariance’’ Σ_r^p , for $r = 1, 2, \dots, R$, with R detections. Here, superindex ‘ p ’ stands for people.

In this work we will, as mentioned earlier, assume that a reliable detection algorithm of people already exists.

B. Merging persons to groups

We apply K -means clustering to find candidates for clusters of people that can be treated as a dense group. The output from the clustering algorithm is a cluster center z_l^c in the image plane. Here, superindex ‘ c ’ stands for cluster to distinguish it from people. The cluster l consists of the people $r \in R_l$, where R_l defines a set of indices of people. The cluster center and its covariance (representing an area covering the people) are computed using standard merging formulas, which are:

$$z_l^c = \frac{1}{|R_l|} \sum_{r \in R_l} z_r^p, \quad (12)$$

$$\Sigma_l^c = \frac{1}{|R_l|} \sum_{r \in R_l} \Sigma_r^p + (z_r^p - z_l^c)(z_r^p - z_l^c)^T. \quad (13)$$

Here, $|R_l|$ denotes the cardinality of the set R_l that is the number of people in the set. To decide how dense a cluster of people is we propose to use the property that $\det(\Sigma)$ is proportional to the area of the ellipsoid. The area of the cluster normalized with the total area of all people included in the cluster is a good indicator of the density of a group. Therefore an appropriate dense-group measure is:

$$d_l = \frac{\det(\Sigma_l^c)}{\sum_{r \in R_l} \det(\Sigma_r^p)}. \quad (14)$$

For a cluster that consists of only one person, we get $d_l = 1$. For a cluster that consists of $|R_l|$ people on the same spot, we get $d_l = 1/|R_l|$. When one or more people leave the cluster, d_l starts to increase and at some point of time the clustering algorithm may detect two clusters instead of one. Consequently, a small d_l (much less than one) indicates a dense group of people. For the detection of dense groups we introduce a thresholding procedure, i.e.

$$d_l < D, \quad (15)$$

where in this application $D = 1$ to allow for some uncertainties in sensor data and position estimates. The output from the clustering algorithm is a set of validated groups z_l^c , $l = 1, 2, \dots, L$ and a set of people z_r^p , $r = 1, \dots, R$ that do not belong to a group.

Information about the motion patterns within the dense group can be obtained from d_l . If there exists a dense group for a certain time period, and d_l temporarily fluctuates around D (without giving rise to a new object/cluster according to the clustering algorithm), this can be used as an indication of increased motion activities (or increased interactions).

The fluctuations of d_l are input data to an HMM that describes normal motion patterns, which in this case reflects people that are together (socializing). Section IV.D discusses in more detail the use of the HMM in this application.

C. Tracking of persons and groups

Tracking is basically done in the same way for people and groups. The state consists of at least position and velocity in standard motion models. We use a motion model with only position and velocity in the state vector. The total model is the linear, and the Kalman filter applies for tracking. The Kalman filter relies on a correct association of people z_r^p and group centers z_l^c at each time. Association is performed using a nearest neighbor approach, where the predicted center z_r^c and covariance Σ_r^c from each filter are compared to the outputs from the clustering algorithm.

There are in total $L + R$ Kalman filters running in parallel for all clusters and unclustered people. The number of clusters ($L + R$) is given by the clustering with different K values and by the analysis of the cluster qualities $s_{l,t,A}$. The K value that has the highest $s_{l,t,A}$ indicates the most likely number of clusters.

The clustering also gives information on the number of detections (cluster members) that exists in each cluster. For a dense cluster (with more than one person) the cluster can be

regarded as an extended object [10] and it is not necessary to know the exact number of people in the dense cluster.

D. Estimation of group behavior

The HMM is used to model the expected motion patterns of a dense and calm group. We compute the likelihood that the observation sequence O represents normal motion patterns. A low likelihood would indicate that O does not represent normal motion patterns and we have instead detected some anomalous motion pattern. We decide to model only normal motion patterns since that is many times easier than to model different abnormal events (and to try to classify the different abnormal events). Consequently we use a semi-supervised anomaly detection approach [1].

Changes of the cluster shape $d_{l,t}$, as described by changes of d_l (14) over time t , is assumed to reflect the degree of interactions between people in the dense group. Few changes would indicate normal motion patterns, i.e. that people are together. Intense changes would instead indicate intense activities, e.g. fights. The observation symbol O_t , can take the values 1, 2 or 3 according to (14) and the following:

$$\text{If } d_{l,t} \leq 1 \text{ and } d_{l,t-1} \leq 1 \rightarrow O_t = 1, \quad (16)$$

$$\text{If } d_{l,t} \text{ has increased and } d_{l,t-1} > 1 \rightarrow O_t = 2, \quad (17)$$

$$\text{If } d_{l,t} \text{ has decreased and } d_{l,t-1} > 1 \rightarrow O_t = 3. \quad (18)$$

Equation (16) implies that the group is still dense and calm compared to $t - 1$. Equation (17) implies that the density has decreased compared to $t - 1$, and (18) implies that the density has increased compared to $t - 1$.

The HMM parameters A , B , π are obtained by training a specific model topology (defined by N and M) on a set of training data. The only parameter that is known in advance in this case is $M = 3$ (according to (16)–(18)). The computation of $P(O|\lambda)$ will then answer the question: given the observed changes of the cluster shape, over some time period, what is the likelihood that the changes represent normal motion patterns for a group of people?

We have used artificial training data to represent normal motion patterns. Normal motion patterns is here characterized by mostly $O_t = 1$. But $O_t = 2$ and $O_t = 3$ may occur one at a time (in a random manner). It is less common that there are several consecutive time steps with $O_t = 2$ and $O_t = 3$. The training data consists of 760 observation sequences O , where each O consists of four observation symbols, i.e. $O = (O_t, O_{t+1}, O_{t+2}, O_{t+3})$. The time difference between each O_t is 1 second. An example of consecutive observation sequences in the training data is: $O = (1,1,1,1)$, $O = (1,1,2,3)$, $O = (1,2,3,1)$, $O = (1,1,3,2)$ and $O = (1,1,1,1)$.

Input data to the Baum-Welch algorithm are the 760 observation sequences, an assumed model topology (N and M) and initial guesses of the parameters for A , B and π . The initial guesses are based on randomly selected probabilities. The Baum-Welch algorithm adjusts the parameters so that the likelihood for obtaining the training data will be maximized.

We perform the training for different N to find a suitable model topology. For the final HMM parameters we select $N = 3$ and obtain for A, B and π :

$$A = \begin{pmatrix} & S_1 & S_2 & S_3 \\ S_1 & 0.62 & 0.27 & 0.11 \\ S_2 & 0.45 & 0.31 & 0.24 \\ S_3 & 0.60 & 0.18 & 0.22 \end{pmatrix}, \quad (19)$$

$$B = \begin{pmatrix} & v_1 & v_2 & v_3 \\ S_1 & 0.54 & 0.16 & 0.30 \\ S_2 & 0.84 & 0.13 & 0.03 \\ S_3 & 0.86 & 0.09 & 0.05 \end{pmatrix}, \quad (20)$$

$$\pi = \begin{pmatrix} \pi_1 & 0.05 \\ \pi_2 & 0.72 \\ \pi_3 & 0.23 \end{pmatrix}. \quad (21)$$

The state transition model is represented by an ergodic model [18], which means that each state can be reached from all other states, i.e. $a_{ij} > 0$. This would be a representative transition model for normal motion patterns in this case. At least we do not know any other structure that we know would be better. To obtain a suitable ergodic model we selected N so that $a_{ij} > 0$. That is to say, with $N > 3$ we got several $a_{ij} \approx 0$.

V. EXPERIMENTS

We have selected four scenarios that describe two different environments, i.e. a parking place [19] and a scenario resembling a check-in area at an airport [20], see Fig. 2. We focus on the development of the clustering and HMM based methods. Input data are therefore annotated image coordinates for the detection of people, as seen from one camera.



Fig. 2. Four datasets are recorded from two environments: the parking place to the left and the check-in area to the right.

A. Experiment 1: people merge to a dense group

In this experiment there are nine people in a parking place. Initially they come from different directions and are well-separated from each other. They merge to a dense, calm and stationary group in the middle of the scene. After a while the dense group splits up and people leave in different directions

and they become well-separated again. Fig. 3 shows $d_{l,t}$ for $K = 1$. The curve indicates the merging of people into a dense cluster and the dashed line represents the threshold for dense group, $D = 1$.

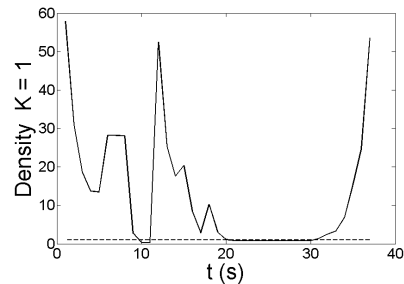


Fig. 3. The variation of $d_{l,t}$ over t , for $K = 1$. The dashed line represents $D = 1$.

A dense group seems to be formed two times. For $8 \text{ s} < t < 11 \text{ s}$ there is a cluster of three to four people. The cluster is dense since $d_{l,t} \leq 1$, which means that people stand close to each other and the cluster area is equal to, or smaller than, the minimum area of the number of people (according (14)). At $t > 11 \text{ s}$ more people enter the scene and eventually merge to the cluster. Just as the people enter there is a strong increase of $d_{l,t}$ (and a reduction in density). This happens since $K = 1$ and no further cluster can be formed. If $K > 1$ the people that enter the scene would have formed one or more clusters of their own. At $t = 20 \text{ s}$ there is an indication of a dense cluster again. This time the cluster lasts until $t = 30 \text{ s}$. When $t > 30 \text{ s}$ people start to leave and $d_{l,t}$ is constantly increasing for the rest of the time.

To see if there are several smaller clusters, K -means clustering is done for $K = 2$ and $K = 3$. Since the minimum number of people is three, the number of clusters that we investigate is $K \leq 3$. Fig. 4 shows $d_{l,t}$ for $K = 2$ and $K = 3$. The variations of $d_{l,t}$ show the same motion patterns as for $K = 1$, i.e. there are dense clusters two times, one for a shorter time, $9 \text{ s} < t < 12 \text{ s}$, and another for a longer time, $17 \text{ s} < t < 33 \text{ s}$. Since we have detected a dense cluster already for $K = 1$, the clusterings with $K = 2$ and $K = 3$ may give clusters that are not so well-separated. If we want to use $d_{l,t}$ for estimating group motion patterns it is important to have as well-separated clusters as possible. In the next section we will discuss how to select data for motion patterns analysis for $K > 1$.

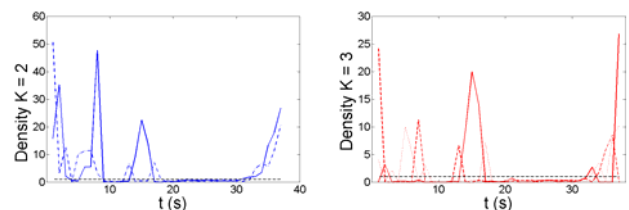


Fig. 4. The densities $d_{l,t}$ for the clusters from $K = 2$ (blue) and $K = 3$ (red). Several clusters show higher densities compared to the cluster from $K = 1$.

B. Experiment 2: people merge to small, dense and active groups

In this scenario people merge into a group and after a while two fights start, each involving two people. The group is split up during the fights to two smaller and denser groups. There

are at the most seven people in the scene. When the fights are over the people leave the scene.

As in experiment 1 clustering is first done with $K = 1$ to get an overview of the situation. The density $d_{l,t}$ shows that a group is formed, but the group is sparse since $1 < d_{l,t} < 4$ for the whole scenario. Clustering with $K = 2$ shows that there exists two dense groups and that there are strong fluctuations of $d_{l,t}$ around $D = 1$ for $13 \text{ s} < t < 17 \text{ s}$. This can be seen in Fig. 5 in the left diagram. During this time period the two fights occur. Clustering with $K = 3$ indicates three dense groups but there are no fluctuations around $D = 1$.

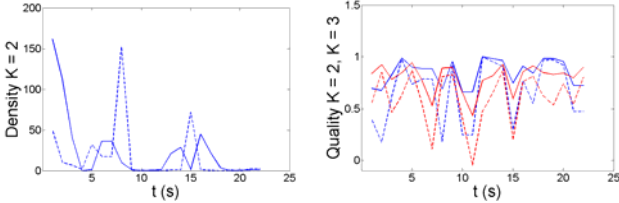


Fig. 5. The diagram to the left shows the densities $d_{l,t}$ for the two clusters from $K = 2$. The diagram to the right shows the cluster qualities $s_{2,t,A}$ for $K = 2$ (blue) and $s_{3,t,A}$ for $K = 3$ (red). The diagram to the right also shows the cluster qualities including the standard deviation, i.e. $s_{2,t,A} - \sigma_{2,t}$ for $K = 2$ (blue dashed line) and $s_{3,t,A} - \sigma_{3,t}$ for $K = 3$ (red dashed line).

Consequently, we have somewhat different recommendations from the clustering with $K = 2$ and $K = 3$, i.e. increased activities for $K = 2$ and no increased activities for $K = 3$. Which data should be used for the estimation of group motion patterns? In the right diagram in Fig. 5 the cluster qualities $s_{2,t,A}$ (for $K = 2$) and $s_{3,t,A}$ (for $K = 3$) are presented. The clustering with the highest quality should be used to represent the situation, since the highest quality describes the most well-separated clusters at the time. In this case we are most interested in the cluster quality for $13 \text{ s} < t < 17 \text{ s}$ when the strong fluctuations occur for $K = 2$. The quality including the standard deviation $s_{2,t,A} - \sigma_{2,t}$ shows that clustering with $K = 2$ is robust for $13 \text{ s} < t < 17 \text{ s}$. For $K = 3$, $s_{3,t,A} - \sigma_{3,t}$ does not show the same robustness. Since $s_{2,t,A} - \sigma_{2,t} > s_{3,t,A} - \sigma_{3,t}$ input data to the HMM should be taken from $K = 2$. We extract for the specific time period the following observation symbols O_t from the left diagram in Fig. 5 (and based on (16)–(18)):

- 1) $9 \text{ s} \leq t \leq 12 \text{ s} \rightarrow O_9 = 1, \dots, O_{12} = 1.$
- 2) $13 \text{ s} \leq t \leq 17 \text{ s} \rightarrow$
 $O_{13} = 2, O_{14} = 2, O_{15} = 3, O_{16} = 2, O_{17} = 3$ (first cluster).
 $O_{13} = 1, O_{14} = 2, O_{15} = 2, O_{16} = 3, O_{17} = 3$ (second cluster).
- 3) $18 \text{ s} \leq t \leq 22 \text{ s} \rightarrow O_{18} = 1, \dots, O_{22} = 1.$

O_t are introduced in O in a sequential procedure. Each second there is a new O_t introduced while the oldest O_{t+3} is taken away. In this experiment we obtain for example the following consecutive observation sequences for the first cluster: $O = (1,1,1,1)$, $O = (2,1,1,1)$, $O = (2,2,1,1)$, $O = (3,2,2,1)$ and $O = (2,3,2,2)$.

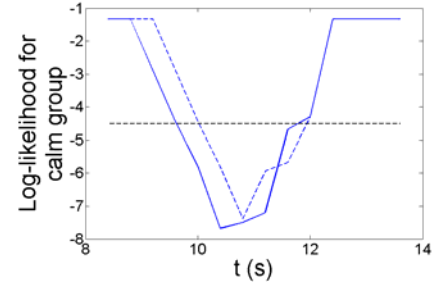


Fig. 6. Log-likelihood ($\log[P(O|\lambda)]$) for calm group for the two clusters. The threshold between approximately normal and anomalous motion patterns is $D_{HMM} = -4.5$.

The log-likelihood $\log[P(O|\lambda)]$ for calm group, computed according to (5)–(10), is presented in Fig. 6. The estimation of group motion patterns describes correctly the two fights which occur for $13 \text{ s} < t < 17 \text{ s}$. At this time the log-likelihood is reduced and below the threshold for normal motion patterns $D_{HMM} = -4.5$ (11).

C. Experiment 3: dense group with increased activities

In this experiment it is assumed that we have a check-in area at an airport (see Fig. 2). There is a check-in desk to the right in the scene. The people enter the check-in area, move around, queue in front of the desk and then leave the area. There are at the most nine people in the scene. The threatening situation takes place during $20 \text{ s} < t < 60 \text{ s}$. A fight between two people takes place during $40 \text{ s} < t < 60 \text{ s}$.

Clustering with $K = 1$ indicates that there is a dense group present all the time. There are minor fluctuations of $d_{l,t}$ around $D = 1$, but no major internal activities are identified. Clustering with $K = 2$ and $K = 3$ indicate on the other hand major fluctuations around D . Fig. 7 presents in the diagram to the left $d_{l,t}$ for $K = 1$ (black), $K = 2$ (blue) and $K = 3$ (red). The diagram to the right presents the qualities $s_{2,t,A}$ for $K = 2$ (blue) and $s_{3,t,A}$ for $K = 3$ (red).

For the motion pattern analysis we extract changes of $d_{l,t}$ according to (16)–(18) for the K with the highest quality $s_{l,t,A}$. In Fig. 7 we can see that there are strong fluctuations of $d_{l,t}$ around D for $6 \text{ s} < t < 10 \text{ s}$ with $K = 2$.

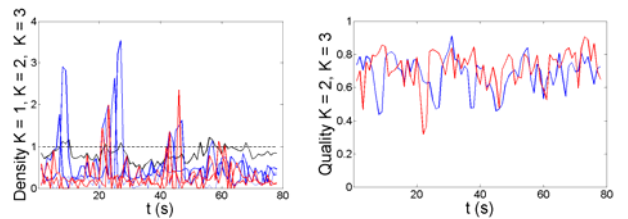


Fig. 7. The densities $d_{l,t}$ for $K = 1$ (black), $K = 2$ (blue) and $K = 3$ (red) to the left. The cluster qualities $s_{2,t,A}$ for $K = 2$ (blue) and $s_{3,t,A}$ for $K = 3$ (red) to the right.

We can also see that at the same time $K = 3$ is more stable than $K = 2$, i.e. $s_{3,t,A} > s_{2,t,A}$. Therefore we should not consider the fluctuations of $d_{l,t}$ for $6 \text{ s} < t < 10 \text{ s}$. (This is correct since the disturbances do not start until $t = 20$.) For the whole time period the following observation symbols O_t for $K = 2$ can be extracted for the motion pattern analysis:

- 1) $1 \text{ s} \leq t \leq 20 \text{ s} \rightarrow O_1 = 1, \dots, O_{20} = 1.$

- 2) $21 \text{ s} \leq t \leq 23 \text{ s} \rightarrow$
 $O_{21} = 2, O_{22} = 2, O_{23} = 3$ (first cluster).
 $O_{21} = 2, O_{22} = 2, O_{23} = 3$ (second cluster).
- 3) $24 \text{ s} \leq t \leq 41 \text{ s} \rightarrow O_{24} = 1, \dots, O_{41} = 1.$
- 4) $42 \text{ s} \leq t \leq 44 \text{ s} \rightarrow O_{42} = 2, O_{43} = 3, O_{44} = 3$ (first cluster).
- 5) $45 \text{ s} < t < 78 \text{ s} \rightarrow O_{45} = 1, \dots, O_{78} = 1.$

The observations for $K = 3$ are:

- 1) $1 \text{ s} \leq t \leq 44 \text{ s} \rightarrow O_1 = 1, \dots, O_{44} = 1.$
- 2) $45 \text{ s} \leq t \leq 47 \text{ s} \rightarrow O_{45} = 2, O_{46} = 2, O_{47} = 3$ (first cluster).

The results from the motion pattern analysis are presented in Fig. 8. We can see that $\log[P(O|\lambda)]$ is low for $20 \text{ s} < t < 60 \text{ s}$ which indicates the disturbances and quick movements during the fight.

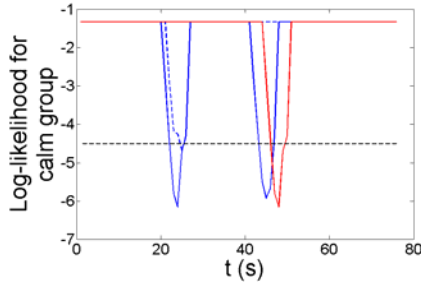


Fig. 8 The log-likelihood for calm group for $K = 2$ (blue) and for one cluster from $K = 3$ (red).

D. Experiment 4: calm group

In this scenario there is a normal situation at the check-in area. People enter, stand in queue and then leave the check-in area. There are no fights or other threatening situations. At the most there are six people in the scene. From the clustering with $K = 1$ it can be seen that there is no single dense group. Clustering with $K = 2$ and $K = 3$ indicate that there are dense groups with some increased activities. Fig. 9 shows to the left $d_{l,t}$ for $K = 2$ and $K = 3$ and to the right the corresponding cluster qualities $s_{2,t,A}$ and $s_{3,t,A}$.

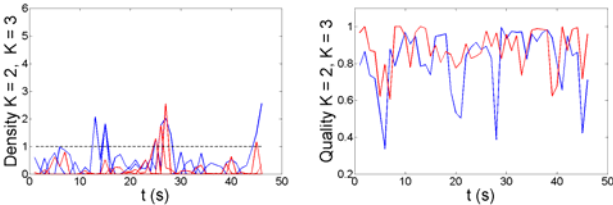


Fig. 9. To the left the densities $d_{l,t}$ for $K = 2$ (blue) and $K = 3$ (red). To the right the cluster qualities $s_{2,t,A}$ for $K = 2$ (blue) and $s_{3,t,A}$ for $K = 3$ (red).

By comparing the two diagrams one can see that the increased activities during $12 \text{ s} < t < 16 \text{ s}$ for $K = 2$ should not be considered since at that time $s_{3,t,A} > s_{2,t,A}$. The increased activities for $25 \text{ s} < t < 28 \text{ s}$ should, on the other hand, be considered for $K = 3$, since at that time $s_{3,t,A} > s_{2,t,A}$. The following observation symbols O_l for $K = 3$ are obtained:

- 1) $1 \text{ s} \leq t \leq 25 \text{ s} \rightarrow O_1 = 1, \dots, O_{25} = 1.$

- 2) $26 \text{ s} \leq t \leq 28 \text{ s} \rightarrow$
 $O_{26} = 3, O_{27} = 2, O_{28} = 3$ (first cluster).
 $O_{26} = 1, O_{27} = 2, O_{28} = 3$ (second cluster).
- 3) $29 \text{ s} \leq t \leq 46 \text{ s} \rightarrow O_{29} = 1, \dots, O_{46} = 1.$

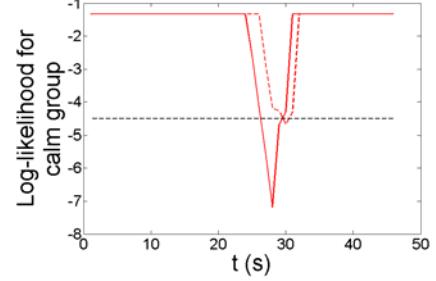


Fig. 10 The log-likelihood for calm group for two clusters from $K = 3$.

The results from the motion pattern analysis are presented in Fig. 10. One of the clusters has a distinct reduction of $\log[P(O|\lambda)]$ for calm group at $t = 28 \text{ s}$. In this case the reduction is not a result from increased activities. It is instead an effect of occlusion, when a person temporarily is hidden behind a tree and then appears again after some time. This experiment shows how occlusion can cause changes in the cluster compositions, and thereby cause false alarms.

VI. SENSITIVITY ANALYSIS

To study how the detection of dense groups may be influenced by increased position uncertainties for the detected people, we have made K -means clustering with different position uncertainties. A dense group can be regarded as an extended object and the sensitivity analysis can be seen as obtaining different detections on the extended object. The position uncertainties are modeled with a uniform distribution for two different intervals, i.e. $\pm 2\%$ and $\pm 4\%$ of the correct positions. In Table I, $d_{l,t}$ has been calculated specifically for the dense groups and for the different uncertainty intervals. For each case we made 10 Matlab runs and calculated the mean of $d_{l,t}$. The results show that with correct positions (column 2), $d_{l,t}$ for the dense groups is well below the threshold ($D = 1$), i.e. $d_{l,t} < 1$. With increased uncertainties (columns 3 and 4) d_l is increasing (i.e. indicating that the groups become less dense). However, $d_{l,t}$ never exceeds D and the correct decision can be taken in all the cases, i.e. that the groups are dense.

TABLE I

CLUSTER DENSITY WITH UNCERTAINTIES IN POSITION ESTIMATES

Number of clusters K and specific cluster l	Mean $d_{l,t}$ for correct position	Mean d_l for ± 2 % of correct position	Mean d_l for ± 4 % of correct position
$K = 1$	0.85	0.91	0.99
$K = 2, l = 1$	0.43	0.52	0.66
$K = 2, l = 2$	0.46	0.51	0.64
$K = 3, l = 1$	0.27	0.34	0.38
$K = 3, l = 2$	0.28	0.26	0.40
$K = 3, l = 3$	0.23	0.28	0.35

The Baum-Welch algorithm finds local maxima when searching for the maximum likelihood estimates. It is therefore interesting to see what HMM parameters that are obtained from other initial parameter settings and how these other HMM parameters would influence on $\log[P(O|\lambda)]$. Fig. 11 shows the results from 10 Matlab runs with randomly chosen initial parameter, where each run gives ten new sets of HMM parameters compared to (19)-(21). The two scenarios in Sections V.B and V.C are used. We also investigate how variations in training data would influence on the maximum likelihood estimates from the Baum-Welch algorithm, and in the next step on the HMM results. Minor random changes are introduced to the original training data and the scenarios in Sections V.B and V.C are used again. The results are presented in Fig. 12.

As can be seen in Figs. 11-12 there are variations of $\log[P(O|\lambda)]$ for the same scenario. But the variations still show the same behavior as for the original initial parameter settings as well as for the original training data. Consequently, the detection of anomalous motion patterns can be done for the scenarios in V.B and V.C also for some variations in input data.

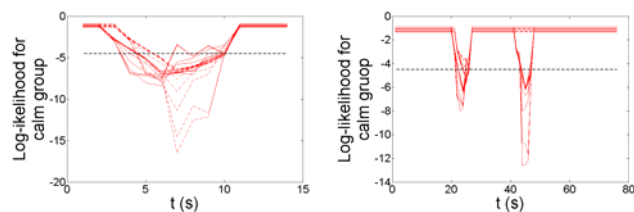


Fig. 11. The log-likelihood for calm group with different initial parameters to the Baum-Welch algorithm. The diagram to the left shows the scenario in section V.B. The diagram to the right shows the scenario in section V.C.

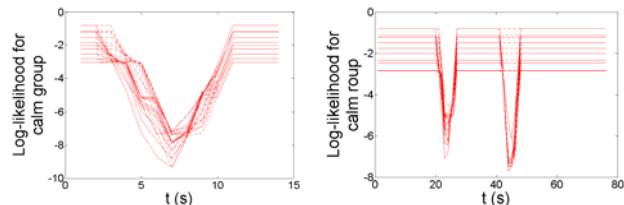


Fig. 12. The log-likelihood for calm group with different training data. The diagram to the left shows the scenario in section V.B. The diagram to the right shows the scenario in section V.C.

VII. CONCLUSIONS

With the clustering and HMM algorithms we have used a rather limited amount of prior information. For the HMM the prior information was the knowledge of the motion pattern of a calm (and normal) group, based on the dynamics of the cluster shape. For the clustering the prior information corresponded to the approximate number of pixels per person, as observed from a certain camera.

Clustering with different K values was used to estimate the number of clusters in the scene. Each cluster could consist of one or several persons. The experiments showed that dense groups can be observed with $K = 1$ as well as with $K > 1$. The same (or at least similar) motion patterns from different views

(or different K values) could be used to strengthen the detection of dense groups.

When analyzing motion patterns it was important to have well-separated clusters. With well-separated clusters the fluctuation of $d_{i,t}$ was likely to reflect the movements of the people. With less-separated clusters $d_{i,t}$ was instead likely to reflect the uncertainty of the clustering (i.e. which cluster member should belong to which cluster). Therefore, observation symbols O_t should be derived from the cluster with the highest quality $s_{i,t,A}$.

How would the approach work for a large crowd? A solution could be to divide a large crowd into smaller sub groups where each sub group corresponds approximately to the scenes used in these experiments. The sub groups are then analyzed according to the proposed procedure and the final results are obtained from fusing the decisions from the sub groups.

The results corresponded well with the real events in the scenarios. In one case we obtained however false alarms because of occlusion. The risk for occlusion would for example be reduced if there was information from several cameras.

In this approach a very accurate detection of each person would not be necessary. Instead we used coarse changes of the group dynamics as indications on anomalous motion patterns. In future work we will investigate in more detail how the signal processing steps before the clustering and HMM may influence on the results.

ACKNOWLEDGMENT

The authors would like to thank PhD student M. Raciti for highlighting the usefulness of the Matlab function for the silhouette value.

REFERENCES

- [1] B. Zhan, D. N. Monekosso, P. Remagnino, S. A. Velastin, and L.Q. Xu, "Crowd analysis: a survey," *Machine Vision and Applications*, vol. 19, 2008, pp. 354-357.
- [2] R. Mehran, A. Oyama, and M. Shah, "Abnormal crowd motion patterns detection using social force model," *IEEE Int. Conf. Computer, Vision and Pattern recognition (CVPR)*, Miami, 2009.
- [3] S. Zaidenberg, B. Boulay, C. Garate, D.P. Chau, E. Corvée, and F. Brémond, "Group interaction and group tracking for video-surveillance in underground railway stations," *Int. Workshop on Motion Patterns Analysis and Video Understanding (ICVS 2011)*, Sophia Antipolis, September 2011.
- [4] E. L. Andrade, S. Blundsen, and R. B., Fischer, "Modelling crowd scenes for event detection", *18th Int. Conf. Pattern Recognition (ICPR 2006)*, Hong Kong, 20-24 August 2006.
- [5] M. Andersson, J. Rydell, and J. Ahlberg, "Estimation of crowd motion patterns using sensor networks and sensor fusion," *12th Int. Conf. on Information Fusion*, Seattle, 6-9 July 2009, pp. 396-403
- [6] M. Andersson, S. Ntalampiras, T. Ganchev, J. Rydell, J. Ahlberg, and N. Fakotakis, "Fusion of acoustic and optical sensor data for automatic fight detection in urban environments," *13th Int. Conf. Information Fusion*, Edinburgh, 26-29 July 2010.
- [7] A. Carmi, F. Septier, and S. J. Godsill, "The Gaussian mixture MCMC particle algorithm for dynamic cluster tracking," *Automatica*, vol. 48, no. 10, October 2012, pp. 2454-2467.
- [8] W. Kone, "Tracking of aircraft groups in an operational air surveillance system", *14th Int. Conf. Information Fusion*, Chicago, 5-8 July 2011.
- [9] A. Swain, and D. Clark, "The single-group PHD filter: an analytic solution", *14th Int. Conf. Information Fusion*, Chicago, 5-8 July 2011.

[10] M. Baum, B. Noack, and U. D. Hanebeck, "Extended object and group tracking with elliptic random hyperface models, *13th Int. Conf. Information Fusion*, Edinburgh, 26–29 July 2010.

[11] K. Granström, C. Lundquist, U. Orguner, "A Gaussian mixture PHD filter for extended target tracking", *13th Int. Conf. on Information Fusion*, Edinburgh, 26–29 July 2010.

[12] E. Hjeltnäs, and B. K. Low, "Face detection: a survey", *Computer Vision and Image Understanding*, vol. 83, no. 3, September 2001, pp. 236–274.

[13] Y. Ishii, H. Hongo, K. Yamamoto, and Y. Niwa, "Face and head detection for a real-time surveillance system", *17th Int. Conf. Pattern Recognition*, vol. 3, 23–26 August 2004, pp. 298–301.

[14] S. Singh, A. Gyaourova, G. Bebis, and I. Pavlidis, "Infrared and visible image fusion for face recognition", *SPIE, Biometric Technology for Human Identification*, 25 August, 2004.

[15] M. Andersson, J. Rydell, L. Saint-Laurent, D. Prévost, and F. Gustafsson, "Crowd analysis using with target tracking, *K*-means clustering and hidden Markov models", *15th Int. Conf. Information Fusion*, Singapore, 9–12 July 2012, pp. 1903–1910.

[16] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, 2nd edition, Springer, 2009.

[17] P. J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis", *Journal of Computational and Applied Mathematics*, vol. 20, 1987, pp. 53–65.

[18] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," In Proc. of the *IEEE*, vol. 77, no. 2, 1989, pp. 257–286.

[19] www.ino.ca/Video-Analytics-Dataset.

[20] S. Ntalampiras, D. Arsic, M. Hofmann, M. Andersson, and T. Ganchev, "PROMETHEUS: heterogeneous sensor database in support of research on human motion patterns in unrestricted environments", *Signal, Image and Video Processing*, DOI: 10.1007/s11760-012-0346-9, 2012.

excellence and community commitment. In 2004, he completed his M.Sc. degree in Electrical Engineering and, in 2012, his Ph.D. degree in Computer Vision, both at Laval University, Quebec City, Canada. His thesis project, entitled "Combination of thermal and color sensors for foreground / background segmentation in outdoor environment", was performed in collaboration with INO.

Dr. Louis St-Laurent has served as a part-time researcher at INO, Quebec City, Canada, from 2007 to 2011. Since 2011, he is now a research scientist in the Computer Vision group at INO. His research interests include image enhancement, motion detection, object tracking, data fusion, image registration, machine vision and thermography.



Donald Prévost was born in 1968 in Quebec City, Canada. He received both his B. Eng. and M.Sc. degrees in physics from Laval University, Quebec City, Canada in 1990 and 1992 respectively. He completed his doctoral degree in image science at University Paris-Sud, Orsay, France. His thesis was about an optical implementation of a stochastic low-level image restoration algorithm preserving discontinuities.

Dr. Donald Prévost joined INO as a Researcher in the computer vision group in 1996. In 2000, he took the position of Group Leader of the Advanced Imaging Systems group at INO. Since 2007, he is Program Manager of the Vision Program at INO, Quebec City, Canada. His research interests include image enhancement, motion detection, object tracking, data and information fusion, 3D reconstruction and synthetic free-viewpoint rendering.



Maria Andersson is senior scientist in Sensor Informatics at the Swedish Defence Research Agency (FOI) in Linköping, Sweden. She is also guest researcher at Automatic Control, Linköping University. She received the M.Sc. degree in mechanical engineering 1989 and the Ph.D. degree in energy systems 1997, both from Linköping University. During 1997–2001 she was a system engineer at SAAB. In 2001 she joined FOI.

At FOI Dr. Maria Andersson is technical coordinator for a European FP7 project on automatic detection of anomalous behavior for mobile critical assets (including trucks and vessels). Her research interests are in machine learning, sensor fusion, anomaly detection and object tracking for security and safety applications in urban and sea environments.



Fredrik Gustafsson is professor in Sensor Informatics at Department of Electrical Engineering, Linköping University, since 2005. He received the M.Sc. degree in electrical engineering 1988 and the Ph.D. degree in Automatic Control, 1992, both from Linköping University. During 1992–1999 he held various positions in Automatic Control, and 1999–2005 he had a professorship in Communication Systems. His research interests are in stochastic signal processing, adaptive filtering and change detection, with applications to communication, vehicular, airborne, and audio systems. He is a co-founder of the companies NIRA Dynamics (automotive safety systems), Softube (audio effects) and SenionLab (indoor positioning systems).

He was an associate editor for IEEE Transactions of Signal Processing 2000–2006 and is currently associate editor for IEEE Transactions on Aerospace and Electronic Systems and EURASIP Journal on Applied Signal Processing. He was awarded the Arnberg prize by the Royal Swedish Academy of Science (KVA) 2004, elected member of the Royal Academy of Engineering Sciences (IVA) 2007, elevated to IEEE Fellow 2011 and awarded the Harry Rowe Mimno Award 2011 for the tutorial "Particle Filter Theory and Practice with Positioning Applications", which was published in the AESS Magazine in July 2010.



Louis St-Laurent was born in 1979 in Amqui, Canada. He received his B. Eng. degree in electromechanical systems engineering from Rimouski University in 2002, where he was awarded the Lieutenant-Governor's prize for academic