We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS BOOK CITATION INDEX INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

17

Automatic Detection of Unexpected Events in Dense Areas for Videosurveillance Applications

Bertrand Luvison[12], Thierry Chateau[1], Jean-Thierry Lapreste[1], Patrick Sayd[2] and Quoc Cuong Pham[2]
[1] *LASMEA, Blaise Pascal University*
[2] *CEA, LIST, LVIC*
France

1. Introduction

Intelligent videosurveillance is largely developping due to both the increasing population, especially in cities, and the exploding number of videosurveillance cameras deployed. When interesting to dense areas, mainly two kinds of scenes come to mind : crowd scenes and traffic ones. A usual treatment on these videos, usually done by security officers, is to monitore several video streams looking for anomalities. A survey of Dee & Velastin (2008) report a camera to screen ratio between 1:4 in best cases and 1:78 in worst ones. As a consequence, the chances to react quickly to an event are very low. This is the reason why this task need to be assisted. Nevertheless automatically detecting anomalies in these kinds of video is particularly difficult because of the large amount of information to be processed simultaneously and the complexity of the scenes.
Most of computer vision methods perform well in visual surveillance applications where the number of objects is low. Individuals can be successfully detected and tracked in scenarios where they appear in images with a sufficient resolution, and in the case of very limited and/or temporary occlusions. However, in crowded scenes, such as in public areas (for example, airports, stations, shopping malls), the video analysis task becomes much more complex. Abnormal behaviour definition is very scene and context dependent. Objects of interest may be small with respect of the global view, and only partially visible thus very difficult to model. Moreover, permanent interaction between individuals in a crowd even complicates the analysis.

1.1 State of the art

Crowd analysis methods can be divided in two main categories Zhan et al. (2008).
Local (or microscopic) approaches which try to segment individuals and track them. Tracking people can be performed in the monocamera case (Zhao & Nevatia (2004), Bardet et al. (2009) and Yu & Medioni (2009)), with stereo sensor (Tyagi et al. (2007)), or in the multicamera setup (Wang et al. (2010)). Learning paths enables the detection of abnormal trajectories (Junejo & Foroosh (2007), Hu et al. (2006), Saleemi et al. (2008)), or infering people interaction (Blunsden et al. (2007), Oliver et al. (2000)). The analysis of trajectories is also used in intrusion detection applications where crossing a virtual line raises an alarm or increase a counter

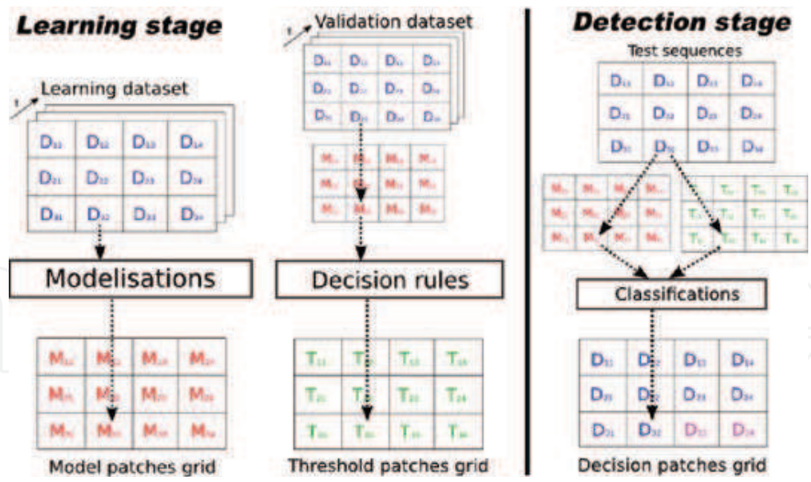


Fig. 1. System outline.

(Rabaud & Belongie (2006) or Sidla & Lypetskyy (2006)). Local approaches also tackles the problem of posture recognition in crowded area (Zhao & Nevatia (2004), Pham et al. (2007)).
Global (or macroscopic) approaches that treat crowd as a single object without segmenting persons. Most of global methods are based on motion analysis. Depending on the context, abnormal motion may be absence of movement, or unexpected movement direction in monocamera (Kratz & Nishino (2009) and Zhong et al. (2007)) or with multiple camera (Adam et al. (2008)). The problem of event detection in a crowd can consist in characterizing small perturbations, such as a person lying down (Andrade & Blunsden (2006)), in a global view of the scene, like in aerial images (Saad & Shah (2007)) or with a scene saliency measure (Mahadevan et al. (2010)). In Mehran et al. (2009) and Wu et al. (2010) the authors propose a way to detect bursting crowd. Varadarajan & Odobez (2009) and others (Wang et al. (2009)) detects pedestrians crossing streets in forbidden areas, cars stopping in unauthorized zones, wrong way displacements, etc. In Breitenstein et al. (2009), a method is proposed to detect all scenes that differ from a learned corpus of observed situations. Küttel et al. (2010) implement a framework for correlating vehicle and pedestrian typical trajectories. The recognition of a person particular movement in a crowd is also an addressed issue in crowd analysis (Shechtman & Irani (2005)) or tracking a particular person in very crowded scenes (Kratz & Nishimo (2010)).

1.1.1 Crowd features

1.1.1.1 Microscopic approches

The basic idea here is to segment individuals in order to recover their trajectory by tracking them. Methods differ in both the human appearance models (descriptors) used for segmenting people and the way datas are associated.
In Zhao & Nevatia (2004), Sidla & Lypetskyy (2006) and Pham et al. (2007), individuals are segmented using several ellipses or rectangles to represent body parts or the omega shape to model both the head and the shoulders. Yu & Medioni (2009) reinforce people tracking with occlusion by adding information on the appearance of the persons before they are occluded and an assumption on the speed continuity of tracked blobs. Kratz & Nishimo (2010) distinguish people inside a crowd using both color histogram and global mouvement model.

In tracking, the assumption is that the shape of persons does not vary much at the scale of and individual in a crowd, and that physical points lying on a person move in the same way (same trajectory and same speed) (Brostow & Cipolla (2006), Rabaud & Belongie (2006), Sidla & Lypetskyy (2006) and Hu et al. (2006)). Tracking algorithms are widely used to recover people trajectories. Kalman filtering (Stauffer & Grimson (2000), Oliver et al. (2000) and Zhao & Nevatia (2004)) and particle filtering (Bardet et al. (2009) and Yu & Medioni (2009)) are the most popular tracking algorithms. These filters can also integrate classification data and a priori information on objects.

The main drawback of tracking methods is the complexity which grows linearly with the number of targets which becomes untractable in the case of dense crowd. The second drawback is the occlusion handling, difficult to take into account in a crowd.

1.1.1.2 Macroscopics approaches

Global approaches require less assumption than local methods. They are based on global information on the crowd which can be more or less locally studied. As pedestrians are not precisely segmented, the detection of unusual motion provides unclassified information, i.e. the detection does not necessarily originate from a human, but for instance from objects in the background such as trees or shadows.

Motion is the most direct feature that can be analysed in a crowd. Motion is generally measured by computing the optical flow in the image. The Lucas-Kanade algorithm is employed in Adam et al. (2008) where the result is filtered using a block median filter (Varadarajan & Odobez (2009), Wang et al. (2009)). In Andrade & Blunsden (2006), the robust piecewise affine method of Black and Anandan is used. Saad & Shah (2007) or Wu et al. (2010) analyze the motion of a huge crowd by building an analogy with fluid dynamics. Spatiotemporal structures ar used in Shechtman & Irani (2005), in Kratz & Nishino (2009) and (2010). Zhong et al. (2007) model a movement energy and search for abnormal discontinuities of this function.

In contrast to the previous methods, some approaches are based on the modelling of the interaction forces between people inside the crowd (Mehran et al. (2009)). Dynamic textures proposed by Chan & Vasconcelos (2008) in crowd analysis context (Mahadevan et al. (2010)), it enables the detection of non pedestrian entities (bikers, skaters, etc.) in walkways or usual motion patterns. Beyond the scope of crowd analysis, Breitenstein et al. (2009) present an approach to store in an efficient way all past scenes and detect new ones. One of the applications of the method is the detection of non moving vehicles in a dense area.

1.1.2 Learning methods

Two different techniques are commonly employed to classify detected events in a crowd. The first one is based on ad-hoc rules that are defined thanks to prior depending on the context, or the application (Junejo & Foroosh (2007)). The second relies on a learning process. The assumption in the learning approach is that “normal” observations are the most frequently observed ones, whereas “abnormal” situations come from rare or unseen observations. Classification methods based on learning focus on specific features or descriptors extracted from the crowd analysis.

Data clustering approaches aim at subdividing datas in homogenous groups. One can mention K-means used in Hu et al. (2006) for finding blob centroid or Wu et al. (2010) to gather similar trajectories with a special method automatically finding the number of cluster.

In the Bayesian framework, the learning approach is expressed as the estimation of the maximum posterior density function. Several algorithms are proposed, the most commonly used are the expectation-maximization algorithm (EM) (Mahadevan et al. (2010)) and the Kernel Density Estimation (KDE) (Saleemi et al. (2008)). Markov Chain Monte Carlo Methods are also proposed in several approaches (Pham et al. (2007), Saleemi et al. (2008), Yu & Medioni (2009)). Breitenstein et al. (2009) propose an ad-hoc method for updating the maximum posterior density function once a day.
Zhao & Nevatia (2004), Andrade & Blunsden (2006) and Kratz & Nishino (2009) exploit the temporal consistency by computing spatiotemporal patterns using Hidden Markov Models (HMM). Moreover, Kratz & Nishino (2009) introduce a spatial consistency between local movement patterns by modelling them with coupled HMM, also used in Oliver et al. (2000) for trajectories interaction analysis. Küttel et al. (2010) combine HMM with natural language processing approaches for creating behaviour dependency networks.
Approaches inspired by natural language processing try to analyse the relationship between documents and the words they contain, by building topics. Varadarajan & Odobez (2009) use a probabilistic Latent Semantic Analysis (pLSA) to learn position, size and motion features. Mehran et al. (2009) use the Latent Dirichlet Allocation (LDA) algorithm with words based on the social forces computation whereas Wang et al. (2010) use motion-drawn words. Küttel et al. (2010) rely on Hierarchical Dirichlet Process (HDP) which automatically find the number of topics as opposed to LDA. Wang et al. (2009) compare an extension of HDP, the dual-HDP with LDA, showing outperforming results.

1.2 Our approach

In order to anwser to the problem of automatic crowded area analysis, several choices has been done:

- A system without calibration step to avoid complex deployement process.
- A global approch, using motion, to be independent of the number of targets in the scene and to be more persistant. Indeed, motion is estimated with few frames whereas a trajectory is issued from a long term process and can be hardly recovered if failed. The motion has the advantage to work on intensity gradient, so these kind of features are very robust various weather and illumination condition changes.
- A learning approach to be as generic as possible, working at the same time on traffic or crowd scenes.
- A supervised approach because no labeled dataset can be made when dealing with anomalies which are by definition infrequent.

Giving an video stream from a fixed camera, the proposed system is able to generate, in an offline process, a statistical model of frequently observed (considered as normal) motion. The scene is divided into blocs from a regular grid. The motion is characterized by a new spatio-temporal descriptor computed on each blocks. The detection stage consists in searching motion patterns that deviate from the model and considered as unexpected events. The decision rule is given thanks to a confidence criteria. An overview of the system is given on figure 1. This method has the asset to be completely automatic: no camera calibration is needed, no labelling task has to be done on the learning database. Moreover, the approach is independent of the number of targets and runs in real-time.
This paper is organized as follows. Section 2 introduces a new characterisation of the mouvement using a spatio-temporal structure as a feature. Section 3 presents the classification

framework. It relies on a new density estimation method competing with classical algorithm such as KDE or EM. Final section 4, compares improvements obtained using our motion features compared to classical optical flow movement estimation with our classification framework and both quantitative and qualitative results concerning unexpected event detections.

2. Movement characterisation

2.1 Opical Flow

Global movement on a scene is generally determined using optical flow estimation algorithms. These algorithms rely on the gradient constraint which suppose a constant illumination of object between two frames. This poor assumption combined with spatial constraint still manages to estimate on each pixel a displacement from one frame to another. Different optical flow techniques have been tested, such as Lucas & Kanade (1981) and its variants, Horn & Schunck (1981) or "Block matching method" (Barron et al. (1992)). From all the different techniques, the Black & Anandan (1996) has been chosen for its robustness and the cleanliness of its result compared to others methods, but also for its relative fast computation. This method is based on a piecewise affine motion assumption which is generally satisfied for our type of scene. Moreover, its computation time remains sustainable for real-time analysis.

When using displacement flow as a descriptor for our system, some special cares need to be taken. The movement magnitude for example, is not as meaningful as the orientation because of the gradient constraint. As a consequence, only the movement direction is studied. To compare two directions an angular distance can be used :

$$d_\theta(\theta_1, \theta_2) = min(\ |\theta_2 - \theta_1|, |\theta_2 - (\theta_1 + 2\pi)|\ )\ with\ \theta_1 < \theta_2 \tag{1}$$

2.2 Spatio-temporal descriptors

The more the movement characterisation is continuous over time, the better it is. Indeed, optical flow usually estimates the movement between two frames and can sometimes be biased by ponctual perturbations. Spatio-temporal structures are convinient to filter such phenomenons. Kratz & Nishino (2009) use this kind of structure in a crowd analysis context, modeling gradients along x,y and t computed on a greyscale cuboid extracted from a sub area of the video through several frames, with a 3D gaussian. To compare two cuboids, Kratz & Nishino (2009) use the symmetric Kullback-Leibler divergence.

Shechtman & Irani (2005) work also tackles the problem of spatio-temporal movement characterisation. We will describe the theory of this method because the new descriptor proposed in this paper relies on the same theory. When considering a uniform mouvement inside a cuboid, constant grey level pixels are all aligned following the same direction through the cuboid. This direction $[u\ v\ w]^T$ is perpendicular to the space-time gradients $\boldsymbol{\nabla \mathbf{I_i}} = [I_{i,x} I_{i,y} I_{i,t}]^T = [\frac{\partial I(i)}{\partial x} \frac{\partial I(i)}{\partial y} \frac{\partial I(i)}{\partial t}]^T$. Figure 2 represents this linear relationship. Let G be the matrix gathering $\boldsymbol{\nabla \mathbf{I_i}}$ gradients of all the N pixel of the cuboid, $G = [\boldsymbol{\nabla \mathbf{I_1}} \dots \boldsymbol{\nabla \mathbf{I_N}}]^T$. We obtain $G[u\ v\ w]^T = [0\ 0\ 0]^T$ which can be reformulated using the Gram matrix :

$$G^T G \begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \tag{2}$$

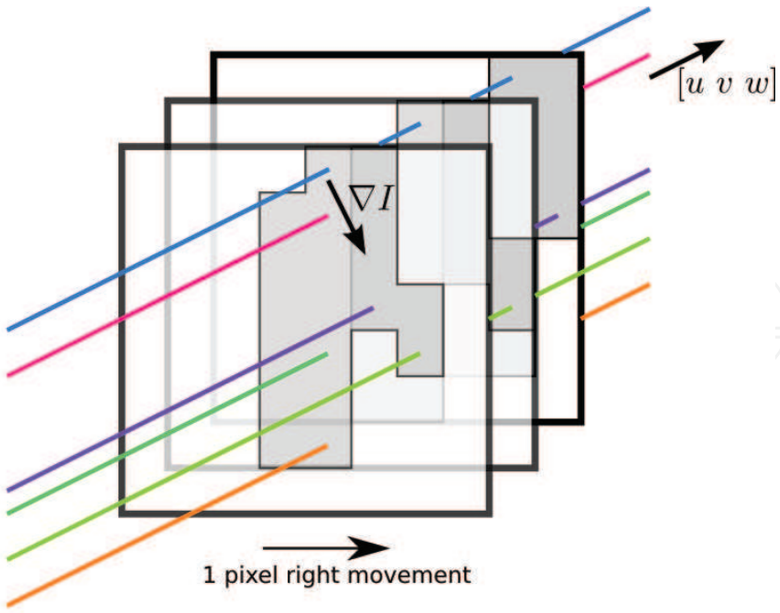


Fig. 2. Spatio-temporal structures in a translation movement case. The constant greyscale lines are all parrallel.

Let M be the Gram matrix $G^T G$ associated :

$$M = G^T G = \begin{bmatrix} \sum_i I_{i,x}^2 & \sum_i I_{i,x} I_{i,y} & \sum_i I_{i,x} I_{i,t} \\ \sum_i I_{i,y} I_{i,x} & \sum_i I_{i,y}^2 & \sum_i I_{i,y} I_{i,t} \\ \sum_i I_{i,t} I_{i,x} & \sum_i I_{i,t} I_{i,y} & \sum_i I_{i,t}^2 \end{bmatrix} \tag{3}$$

M can be considered as a extension of the Harris matrix (Harris & Stephens (1988)), whose definition is :

$$M^\diamond = \begin{bmatrix} \sum_i I_{i,x}^2 & \sum_i I_{i,x} I_{i,y} \\ \sum_i I_{i,y} I_{i,x} & \sum_i I_{i,y}^2 \end{bmatrix} \tag{4}$$

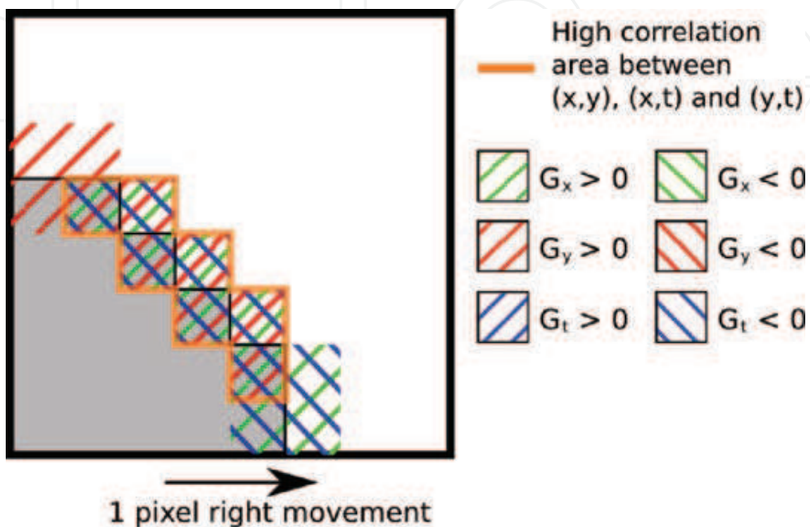
Matrix M contains all information needed for spatio-temporal corner detection.
Note that equation (2) has a solution only if matrix M is rank-deficient ($rg(G) = rg(M) \neq 3$). Otherwise, the movement inside the cuboid is not uniform, it is a spatio-temporal corner considering intensity lines. As a consequence, no increase in rank between the upper left minor $M^\diamond$ of M define on equation (4) and matrix M notices a uniform motion in the cuboid.
Two cuboids are motion consistent if appending the two cuboids along the temporal dimension still verifies the rank criteria cited above. However, this criteria provides a binary answer. As a consequence, Shechtman & Irani (2005) define a continous rank-increase measure to take into account the natural image noise and to give a graduated answer. This measure is defined by :

$$\Delta\hat{r} = \frac{\det(M)}{\det(M^\diamond).\|M\|_F} \tag{5}$$

where $\|M\|_F$ is the Frobenius norm of matrix M. Note that $\Delta\hat{r_{ii}} = \Delta\hat{r_1}$ is not necessarily equal to zero. Shechtman & Irani (2005) define another measure, m_{ij}, to ensure that m_{ii} is minimal. m_{ij} which captures the degree of local inconsistency between two cuboids, is equal to :

$$m_{12} = \frac{\Delta\hat{r}_{12}}{min(\Delta\hat{r}_1, \Delta\hat{r}_2) + \epsilon} \tag{6}$$

These spatio-temporal structures can model smoother movements or even more complex movements. In order to have the best classification results possible, we proposed a new spatio-temporal descriptor that rely on the same assumption than Shechtman & Irani (2005) descriptor.



(a)

Fig. 3. Shape influence on linear relationship estimation for translation movement.

2.3 Our descriptor

Shechtman & Irani (2005) based their descriptor on studying the linear dependency between spatial gradients and the temporal gradient. Instead of using the rank of the matrix M, we propose to look for a possible linear dependency using a correlation measure. The correlation between two random variables $X = (x_1, ..., y_n)$ and $Y = (y_1, ..., y_n)$ is given by the Pearson formula :

$$\rho_{XY} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2}} \tag{7}$$

with $E(.)$ the expected value. Other measures of dependence exist, such as mutual information but according to the application context, the Pearson correlation is the right mesure for searching linear relationship.

According to equation (7), standard deviations of each random variable need to be different to zero. In our case, the natural noise in the image is usually enough to ensure this property. Singular remaining cases represent either a perfect gradient color or a uniform image area. Both cases which are not interesting situations, can be filtered by thresholding the gradient magnitude.

The proposed descriptor is thus constructed looking on the linear correlation between both x and t and y and t. We obtain the movement characterisation $\mathbf{C} = [\rho_{xt}\ \rho_{yt}]^T$. The distance between two descriptors is defined by equation (8). Values are taken in $[0, 2]$.

$$d_{corr}(\mathbf{C_1}, \mathbf{C_2}) = 1 - \frac{\mathbf{C_1}.\mathbf{C_2}}{\|\mathbf{C_1}\|\|\mathbf{C_2}\|} \tag{8}$$

This feature is not an estimation of the movement since no movement magnitude is defined. Only a confidence measure on linear dependency existence is expressed with this feature. Since magnitude gradients are not taken into account, separation between diagonal mouvement in the same quartile is theorically impossible. For example, considering two translation movements $(2,1)$ and $(1,2)$, in both cases the correlation vector should be $\mathbf{C} = [1\ 1]^T$. In practice, correlations on real image gradients are never perfect. The more a movement is well defined in a direction, the more the correlation is high. This is the reason why two different diagonal movements from the same quartile will give different features $\mathbf{C}$. Nevertheless, vector $\mathbf{C}$ magnitude gives a confidence criteria on the characterisation. If $\mathbf{C}$ magnitude is too low, one can consider that no main movement exists in the cuboid. This piece of information is analog to the consistency criteria defined by Shechtman & Irani (2005). Moreover, normalizing data through correlation computation makes the descriptor invariant to affine illumination changes of type $\widehat{I} = aI + b$ where I is the greyscale cuboid. Indeed, such a change, modify gradients such as $\widehat{G} = aG$ but does not change the linear relationship, so $C_{\widehat{G}} = C_G$.

However, the descriptor suffers, like optical flow estimation from the aperture problem. This problem appears along straight edge where only the normal component of the movement can be estimated. Here, the problem is similar, since gradients are computed in a given base (the image orthonormal basis x,y), a movement can be fully defined if part of edges are aligned along both axis in the cuboid. A gradient aligned along an axis can only give information on this axis component of the movement. Diagonal edges gives diagonal spatial gradients which may bias the movement characterisation. The positive correlation relationship is theorically not transitive except under some conditions. This transitivity relationship is discussed by Langford et al. (2001) who show that for three random variables A, B et C such as $\rho_{AB} > 0$ et $\rho_{BC} > 0$, the correlation between A et C is bounded by :

$$\rho_{AB}\rho_{BC} - \sqrt{(1-\rho_{AB}^2)(1-\rho_{BC}^2)} \leq \rho_{AC} \leq \rho_{AB}\rho_{BC} + \sqrt{(1-\rho_{AB}^2)(1-\rho_{BC}^2)} \tag{9}$$

For diagonal edges, components x and y are correlated. When the correlation is strong and if component x for example is correlated to component t, then component y will be correlated to. As a consequence, if a cuboid contains mainly a diagonal edge, the caracterisation will tend to be $\mathbf{C} = [\alpha\ \beta]$ with $|\alpha| \approx |\beta|$ whatever the true movement is, as shown with the orange area on figure 3.

To avoid such problem, only the thrustful information contained in the cuboid can be kept for linear relationship estimations. This subset is made from gradients aligned along x and y axis. Let S_x et S_{xt} be respectively gradient sets $I_{i,x}$ and $I_{i,t}$ for points with spatial gradient aligned along x axis. Such a filtering makes movement characterisation more precise and thus more discriminative but considering only subset of gradients can lead to singular cases. These cases occur when not enough gradients are aligned along one of the two axis. To avoid such a phenomenon, the alignement constraint is relaxed to accept gradients in an angular interval of $\frac{\pi}{4}$ around axis. In the same way,S_y and S_{yt} are defined with gradient aligned around axis y. Subset S_x, S_{xt}, S_y and S_{yt} are defined such as :

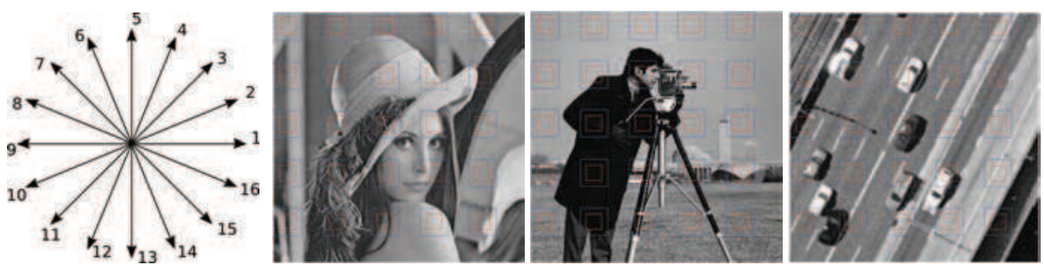
$$
\begin{aligned}
S_x &= \{I_{i,x}| - \frac{\pi}{8} \leq \theta \leq \frac{\pi}{8}\} \ \ and \ \ S_{xt} = \{I_{i,t}| - \frac{\pi}{8} \leq \theta \leq \frac{\pi}{8}\} \\
S_y &= \{I_{i,y}| - \frac{3\pi}{8} \leq \theta \leq \frac{5\pi}{8}\} \ \ and \ \ S_{yt} = \{I_{i,t}| - \frac{3\pi}{8} \leq \theta \leq \frac{5\pi}{8}\}
\end{aligned}
\tag{10}
$$

where $\theta = \arg(I_{i,x}, I_{i,y})\ [\pi]$. Finally, vector $\mathbf{C}$ is equal to $\mathbf{C} = [\rho_{S_x S_{xt}}$ $corr S_y S_{yt}]^T$. For the remaining singular cases where there are still not enough gradients along axis, typically very low frequencies image areas, instead of giving a wrong movement characterisation, an invalidate state for the feature is set.
In the rest of this paper this new descriptor is named "Separated Selected Correlation"(SSC).

2.4 Experimental results

2.4.1 Movement separation

In order to validate the proposed descriptor, movement class separation of descriptor SSC has been compared to the initial version our the proposed descriptor without filtering on spatial gradient orientation, but also compared to Shechtman & Irani (2005) and Kratz & Nishino (2009) descriptors. Cuboids have been generated and compared for movement in 16 different directions (cf. figure 4(a)). Descriptors have been computed on $T = 3$ frames. The movement generated are exact translations, thus parameter T does not have a lot of influence. On the contrary, for real uniform movement, parameter T smoothes and reinforces the movement characterisation. Spatio-temporal gradients have been computed with Canny method with gaussian standard deviation equal to 1 and filter size of 5 pixels.



(a) Movement indexes. (b) Interest regions for descriptor comparison.

Fig. 4. Synthetic movement generation.

Results are represented as a distance matrix $\mathcal{M}$ of size $(16, 16)$ for a given descriptor and the associated distance. This matrix is assumed to be symetric, with minimal diagonal and a sub-diagonal corresponding to the distance between a mouvement and the opposed one. Cuboids have been generated from real images in differents regions of interest $r_i \in R$ represented in red on figure 4(b) in translation along the 16 directions of figure 4(a). The blue boxes on figure 4(b) correspond to area possibly seen through the displacement. Movement characterisation has to be independant to the shape contained in the cuboid, as a consequence, distances between two direction i et j are computed for all the couples of regions of interest and then averaged.

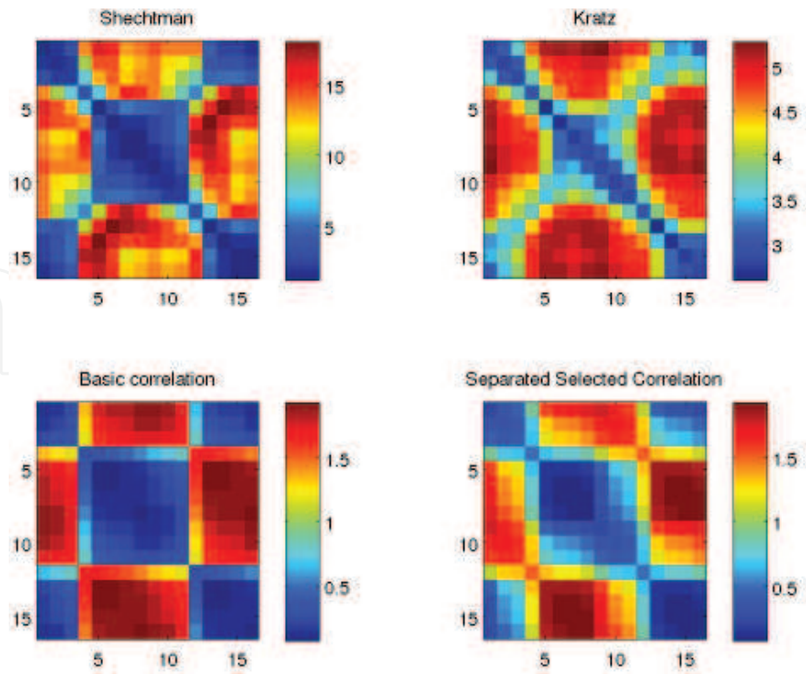


Fig. 5. Distance matrices for real images translation movement using different spatio-temporal descriptors.

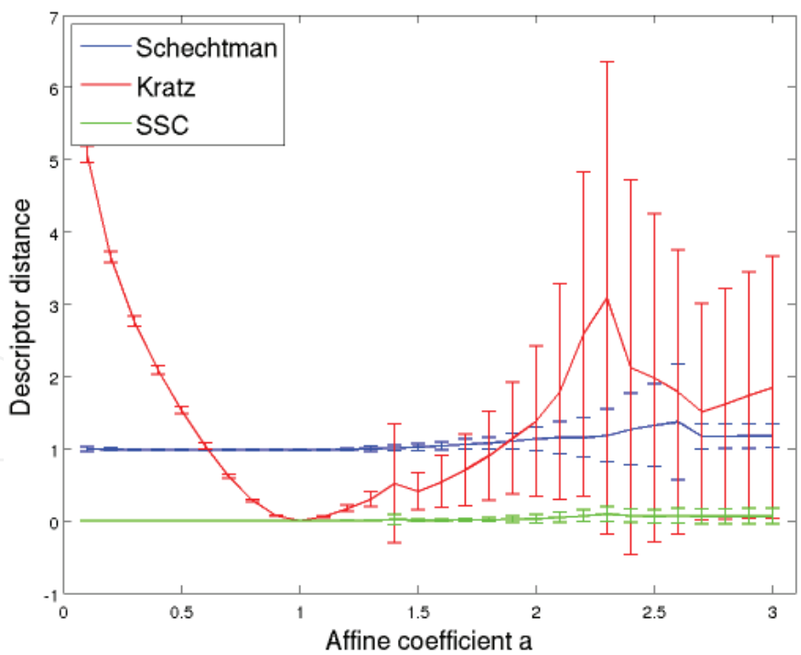


Fig. 6. Evolution of distance mean and standard deviation between descriptors with affine illumination change.

Movement separation results are shown on figure 5. All the three descriptors with their own distance roughly distinguish movements. However the SSC descriptor is more constant

and precise. In order to measure distance matrix quality, the mean of the maximum relative position is computed. For a movement in direction i, the maximum distance is expected for opposed movement, that is to say movement with index $i + 8$ [16]. Table 1 shows that SSC descriptor is the nearest to the theoritical index 8 than other methods. Moreover, standard deviations on these maximum positions show that the separation is more stable whatever the shape contained in the cuboid is.

	Shechtman	Kratz	Simple correlation	SSC
$\overline{i_{max}}$	6	8.5	7.625	7.875
$\sigma_{\overline{i_{max}}}$	2.7809	2.3094	0.8851	0.6191

Table 1. Average shift and standard deviation between two movement extremum.

One may note that concerning the simple correlation method, movement is distinguished in roughly two classes illustrating the shape influence phenomenon. The spatial correlation biases **C** computation to make it constant for a mouvement class whatever the true movement. SSC version of the algorithm decreases this effect in a significant manner.

2.4.2 Affine illumination change invariance

To validate this property, the distance between a cuboid without illumination change and one with it has been computed for a given direction on all regions of interest $r_i \in R$ represented in red on figure 4(b). The different curves on figure 6 represent distance mean and standard deviation on all region of interest of R function of coefficient a. Descriptors are characterizing the same direction, so distance between them should be minimal (0 for Kratz & Nishino (2009) and SSC descriptor and 1 for Shechtman & Irani (2005)). Except for Kratz & Nishino (2009) descriptor, other ones have a very low distance mean and standard deviation whatever the value of a until reasonable values. Indeed, a for very high value of a, pixels saturate to white which leads to a false descriptor characterization. On the contrary, Kratz & Nishino (2009) descriptor for low value of a does not return low distance as excepted. An affine illumination change modifies the 3D gaussian from $\mathcal{N}(\mu, \Sigma)$ to $\mathcal{N}(a\mu, a^2\Sigma)$ which are two different distributions according to Kullback-Leibler divergence. This deficiency is quite important when dealing with outdoor videos.

2.4.3 Computation efficiency

Because of the real-time constraint, motion characterisation computation time is important. We compared spatio-temporal SSC and Shechtman & Irani (2005) descriptors computation time with Black & Anandan (1996) optical flow method. The implementation was done in C++ with optimised code. Spatio-temporal cuboid have 16x16x5 size. Note that spatio-temporal descriptors gives blocks information whereas optical flow returns a dense information. Thus, performances are not really comparable, times are given for information.
Most of spatio-temporal computation time is caused by the gradients estimation as seen on table 2. Concerning correlation computation for the SSC descriptor, it can be optimised to calculate the correlation in one pass instead of two, dividing computation time by two as shown on figure 2. To do so, the following formula is used for the correlation computation :

$$\rho_{XY} = \frac{N\sum_{i=1}^{N}(x_i y_i) - \sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i}{\sqrt{N\sum_{i=1}^{N} x_i^2 - (\sum_{i=1}^{N} x_i)^2}\sqrt{N\sum_{i=1}^{N} y_i^2 - (\sum_{i=1}^{N} y_i)^2}} \tag{11}$$

Image size	Gradients	unoptimised SSC	optimised SSC	Shechtman	Black & Anandan
320x240	67.92	45.28	19.81	25.47	152.82
640x480	297.15	155.65	76.41	104.71	761.27

Table 2. Number of average clock cycle (million of cycles).

This formulation has to be taken with care because it is not numerically stable. Safegards need to be taken to avoid these cases, such as thresholding gradient magnitudes to consider only significant variations.
Table 2 show that SSC and Shechtman & Irani (2005) descriptors have the same complexity. Moreover, computation time is linear with image size as shown with times fourfold between image resolution 320x240 and 640x480.

3. Classification frameworks

Our application framework imposes us the use of unsupervised learning machines (no labelled databases with abnormal behaviours can be made). The main problem is to drawn a decision function from a set of features representing the "normal" behaviour. We will focus on probabilist approaches which aim at estimating a likelihood function and thresholding it to decide new sample class.
Likelihood functions are widely used into computed vision algorithms like recognition, detection or tracking. However, estimating such fonction from observations is still a challenging task because: 1) in the general case, no prior on the shape of the likelihood can be used to define a simple parametric function and 2) methods have to deal with high dimensionnal features and huge training sets.
For approximating the unknown likelihood distribution of the model, given observations (the learning features) drawn from this model, non parametric or parametric approaches can be used. For the non parametric one, Kernel Density Estimation (named KDE or Parzen windows model Duda et al. (2001)) relies on the choice of a kernel function. This method converges to the true distribution with the number of learning features but with a heavy computational cost which is generally not acceptable as we will see later. K-Nearest Neighbour estimation (KNN) is also a non parametric method that does not assume a window with a given size like KDE. Contrarily, this method defines a cell volume as a function of the training data Duda et al. (2001).
Other methods for approximating unknown distribution are parametric and generally assume that this distribution is a gaussian mixture (GMM). In Dempster et al. (1977) the authors propose an algorithm to estimate the parameters of a mixture of gaussians, using a prior on the number of gaussians. This well known algorithm called Expectation Maximisation has been improved. In Figueiredo & Jain (2002) the constraint on the number of gaussians which is usually unknown in practice, has been removed. Recently, Han et al. (2008) proposed a sequential approach, named SKDA, to approximate a given distribution with GMMs, adding gaussian one by one and mixing it in the previous gaussian mixture if needed. The main drawback of these parametric methods is to suppose a model which may not always fit to the real model. For example, EM or SKDA algorithms use intrinsic Mahalanobis distance to compare features. This may be complete out of sense for use of spatio-temporal features seen in section 2 which have their own comparison distance.
As a consequence, a parametric estimation using adhoc features distance like KDE or KNN, without computational cost constraint is proposed. The decision function needs for

classification context associated with this proposed estimation can be very simple with a fix threshold or more subtle. We choose to use a confidence criteria that will be presented with the proposed estimation method.

3.1 An hybride method

We propose to approximate the KDE with a sparse model composed by a weighted sum of kernel functions in order to withdraw the computational burden associated to the KDE while keeping its precision. Our method will be called SKDE for Sparse Kernel Density Estimation. It aims at selecting the most important features and weighted the kernel functions associated to it, as shown on figure 7. The weight of a feature defines its amplitude and thus its range.

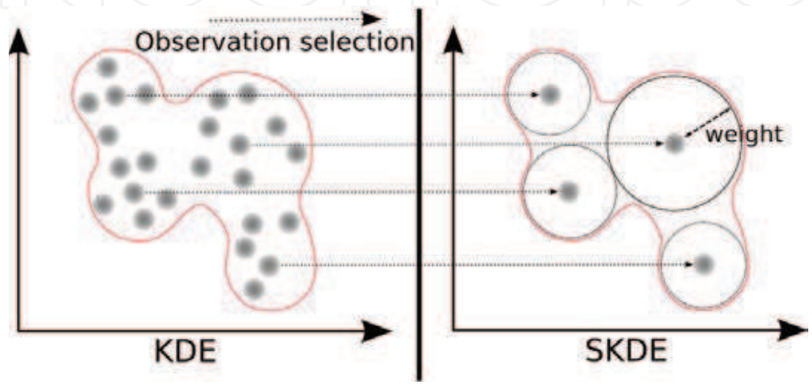


Fig. 7. Feature selection process for KDE approximation.

3.1.1 Likelihood Non-Parametric Approximation

Let $\mathbf{Z} \doteq (z_1, z_2, ..., z_K)^T$ denotes the features belonging to a given model. We choose to represent the likelihood $P(z)$ with a non-parametric model using KDE:

$$P_{\text{KDE}}(z) \approx K^{-1} \sum_{k'=1}^{K} \phi_{k'}(z) \tag{12}$$

where $\phi_{k'}(.)$ is a kernel function (not necessarily gaussian). With such an approach, no assumption needs to be done over the shape of the distribution. However, one of the drawbacks of this approach is that the estimation of the probability is proportional to the number of samples used. We propose a solution wherein a sparse model is obtained by approximating equation (12) by a weighted sum of basis functions.

3.1.2 A Sparse Kernel Density Estimation

Equation 12 can also be expressed as:

$$P_{\text{KDE}}(z) \approx \mathbf{w}^T (\boldsymbol{\phi}(z)) \tag{13}$$

with $\mathbf{w}^T = (1, ...1)^T / K$ is a vector of size K and $\boldsymbol{\phi}$ is a vector function defined by $\boldsymbol{\phi}(z) = (\phi_1(z), \phi_2(z), ..., \phi_K(z))$.
We propose a sparse model formulation of equation (13) by fixing most of the coefficients of $\mathbf{w}$ to zero as it is classically done. This new vector will be called $\tilde{\mathbf{w}}$ and the reduced one, that

is to say the vector of non zero coefficient, $\tilde{\mathbf{w}}$. As a consequence the new estimator expression is:

$$\widehat{P_{\mathrm{KDE}}}(z) \approx \tilde{\mathbf{w}}^T \tilde{\phi}(z) \tag{14}$$

with $\tilde{\phi}$ a vector function extracted from ϕ with the kernel function associated to non-zero weight kept in $\tilde{\mathbf{w}}$. To obtain equation (14), we solve the following least square problem:

$$\widetilde{\mathbf{w_{LS}}} = \arg\min_{\tilde{\mathbf{w}}} \left(\sum_{k=1}^{K} (\mathbf{w}^T \phi(z_k)) - \tilde{\mathbf{w}}^T \tilde{\phi}(z_k))^2 \right) \tag{15}$$

The remaining question to solve problem (15) is how to choose $\tilde{\phi}$. Let Φ denote, a matrix of size $K \times K$ and built such as the element of the line i and column j is given by $\Phi_{i,j} = \phi_i(z_j)$. Φ is a square and symetric matrix, from which, an estimator of the likelihood associated to the sample z_k of the training set is given by the sum of elements of the line or the column k of Φ, that is to say:

$$P_{\mathrm{KDE}}(z_k) = K^{-1} \sum_{k'=1}^{K} \Phi_{k,k'} \tag{16}$$

A likelihood vector $\boldsymbol{\varphi}$ related to the training set is built:

$$\boldsymbol{\varphi} = \Phi \times (\mathbf{1}_K)/K = (P_{\mathrm{KDE}}(z_1), ..., P_{\mathrm{KDE}}(z_K))^T \tag{17}$$

with $(\mathbf{1}_K)$ is a vector of one of size K. With these new notations, problem (15) can be rewritten:

$$\widetilde{\mathbf{w_{LS}}} = \arg\min_{\tilde{\mathbf{w}}} (\|\Phi_v \, \tilde{\mathbf{w}} - \boldsymbol{\varphi}\|) \tag{18}$$

with Φ_v the reduced matrix where only columns with index in set v are taken from Φ. To find this set v, we choose to keep iteratively, indexes of vectors with the maximum residual likelihood. Algorithm 1 is fully described by a two step recursive process:

Algorithm 1 Non parametric estimator approximation algorithm

Require: matrix Φ, stopping criterium Q_l
  Likelihood vector computation: $\boldsymbol{\varphi}_1 = K^{-1}\Phi \times \mathbf{1}_K$
  Initialisation: $m = 0$
  repeat
    Maximum likelihood index extraction: $v(m) = \underset{i}{argmax}\, \varphi_{m,i}$
    Computation of weight vector $\tilde{\mathbf{w}}_m$ solution of problem 18
    Likelihood vector update $\boldsymbol{\varphi}_{m+1} = \boldsymbol{\varphi}_m - \Phi_v \, \tilde{\mathbf{w}}_m$
    $m = m + 1$
  until $\max \varphi_{m+1,i} > h(Q_l)$
  return Weight vector $\tilde{\mathbf{w}}_M$ and the selected feature indexes: $\mathbf{v} = (v(1), v(2), ..., v(M))$

Steps one and two are repeated until $\max \varphi_{m+1,i} > h(Q_l)$. The parameter Q_l represents the precision of the likelihood approximation and h is the confidence criterium for the KDE distribution described in section 3.1.3. For a coarse approximation Q_l can be decreased. In this case the number of used vectors decreases. Illustrations of the effect of this parameter are given in section 3.2. This approach enables to give a good approximation of the likelihood with few vectors. Initially, the non parametric model set $\mathbf{Z}$ contained K elements whereas the sparse vector machine model $\widetilde{\mathbf{Z}} = \mathbf{Z}_v$ contains only M elements with $M \ll K$. Let note :

$$\widetilde{\mathbf{Z}} \doteq (\widetilde{\mathbf{z_1}}, \widetilde{\mathbf{z_2}}, ..., \widetilde{\mathbf{z_M}})^T \tag{19}$$

This reduction in size is mandatory since it makes the real-time classification possible.
In practice, the problem solved on equation (15) can be simplified in our case. Instead of solving the least square problem on all observations z_k, we use the same problem only at control points, that is to say on the selected features, $z_{v(k)}$. In this condition equation (15) can be rewritten:

$$\widetilde{\mathbf{w_{LS}}} = \arg\min_{\tilde{\mathbf{w}}} \left(\sum_{k=1}^{M} (\mathbf{w}^T \phi(z_{v(k)})) - \tilde{\mathbf{w}}^T \tilde{\phi}(z_{v(k)})) \right) \tag{20}$$

And with our notation :

$$\widetilde{\mathbf{w_{LS}}} = \arg\min_{\tilde{\mathbf{w}}} (\|\Phi_{v,v}\ \tilde{\mathbf{w}} - \varphi_v\|) \tag{21}$$

with $\Phi_{v,v}$ the reduced matrix where only rows columns with index in set v are taken from Φ. It amounts to solve a square linear system of reduced dimensionality.

3.1.3 Confidence criteria

Intuitively when approximating a likelihood distribution, regions with high probability are expected to be well approximated whereas regions with almost zero probability can be neglected. The problem is to define the threshold from which probabilities can be neglected (cf. figure 8). This problem is easy to solve for simple distribution such as gaussian density. But for more intricate density such as GMM, the problem could not have exact solution anymore. Nevertheless, approximated methods can be used. One solution is to use a confidence criteria (Paalanen et al. (2006)).

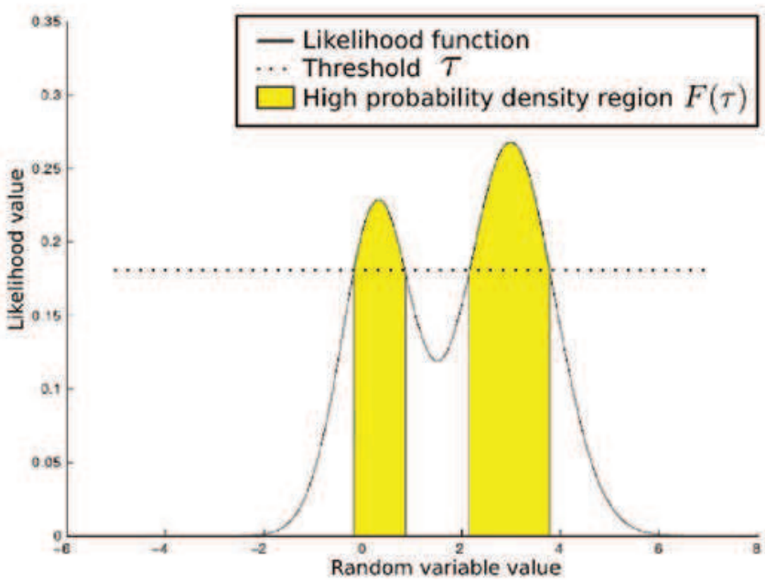


Fig. 8. How threshold τ should be chosen to keep only $F(\tau)$% of the highest probability of a given distribution (represented in yellow) ?

Let $F(\tau)$ be the density quantile for a given probability density value τ,

$$F(\tau) = \int_{p(\mathbf{x}) \geq \tau} p(\mathbf{x}) d\mathbf{x} \tag{22}$$

This density quantile corresponds to the highest density region for density value above τ. A reverse mapping $h(F) = \tau$ can be chosen such as $F \in [0, 1]$ is the density quantile needed (0.9 for 90% of the probability density for example).
The approximating method to solve this problem rely on a Monte Carlo algorithm. Let $X = (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N)$ be N points randomly chosen following distribution p and $p_i = p(\mathbf{x}_i)\ \forall i \in [1, N]$. p_i are then sorted in an ascending order $\mathbf{Y} = (y_1, y_2, ..., y_N)$. This set is used to estimate $F(\tau)$ and $h(F)$ using linear interpolation. Let $i = \underset{i}{argmax}\ \{y_i | y_i \leq \tau\}$, we get :

$$F(\tau) \approx \begin{cases} 1 - \frac{l(0,\tau)}{N} & if\ \tau < y_1 \\ 0 & if\ \tau \geq y_N \\ 1 - \frac{i + l(i,\tau)}{N} & otherwise \end{cases} \tag{23}$$

with

$$l(i, \tau) = \begin{cases} \frac{\tau}{y_1} & if\ i = 0 \\ 0.5 & if\ y_{i+1} - y_i = 0 \\ \frac{\tau - y_i}{y_{i+1} - y_i} & otherwise \end{cases} \tag{24}$$

The inverse transform $h(F)$ can be deduced by:

$$\tau = h(F) \approx \begin{cases} y_N & if\ i = N \\ (N(1-F))y_1 & if\ i = 0 \\ y_i + (N(1-F) - i)(y_{i+1} - y_i) & otherwise \end{cases} \tag{25}$$

with $i = \lfloor N \times (1 - F) \rfloor$.
This confidence criteria is used as a stopping criteria during the iterative approximation algorithm presented on algorithm 1 but it can also be used to determine the decision function for classification purpose thanks to equation (25). Indeed, in classification context once likelihood density is estimated thanks to SKDE, new observations $\mathbf{Z}$ can be considered as random points drawn from the estimated model, that is to say set X. Quantil parameter F will make detections more or less strict, considering F% of observation belonging to the estimated model.

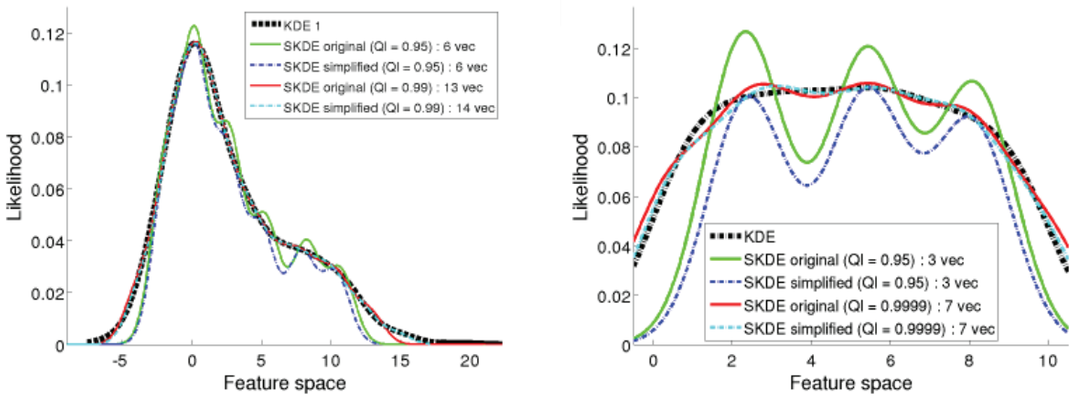
3.2 Experiments

In this section, we present result of our algorithm with other classical method. Several density estimation approximation algorithm have been tested: the KDE, the SKDA and Figueiredo-Jain EM algorithm. The tests have been done on both synthetic and real datas. For synthetic ones, given a known gaussian mixture, a learning set Z of points are randomly drawn from the known distribution. On the way back, we compare the gaussian mixture retrieved from this learning set Z with the different methods. As a consequence, the kernel chosen for the KDE and all more reason for our method, will be gaussian one.

3.2.1 SKDE parameters influence

First of all, some results concerning simplified approximation of SKDE method expressed with equation (21) and parameters influence, realized on monodimensional synthetic data, can be seen on figures 9. The KDE distribution which is the ground truth of our method has

been represented by the black dashed curve. The sparse probability $\widetilde{\mathbf{Z}}$ has been drawn for the original problem approximation of equation (15) and for the simplified one of equation (20). With equal Q_l, the original problem tends to converge oscillating around the true distribution whereas the simplified one, converges toward the KDE distribution without overestimating it. The convergence speed is also shown on figure 10 which represents the Mean Integrated Square Error (MISE) between each approximation and the KDE. We can see on this semilog curves that both methods roughly converge at the same speed. For the rest of the tests the simplified version of the approximation will be used.



(a) Two gaussians distribution (b) Uniform distribution

Fig. 9. Approximation of the kernel based non parametric density estimation with the original SKDE and simplified one.

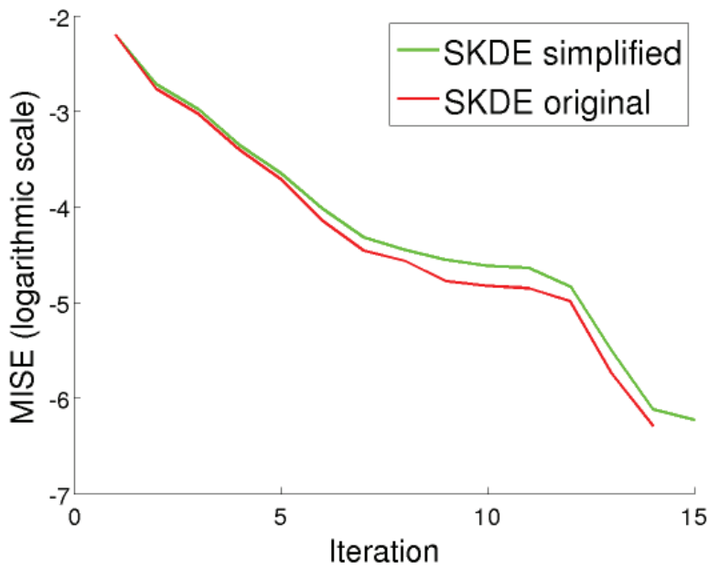


Fig. 10. MISE evolution through iteration process for original SKDE and simplified one.

Concerning the Q_l parameter influence, we can see figures 9 that with low Q_l values the approximation underestimates the distribution for some features (blue curves) whereas cyan curves obtained with a high Q_l fairly well approximate KDE distribution.

3.2.2 Estimation comparative results

We compare the four algorithms in term of precision, sparseness and computation time. The precision is computed thanks to MISE between the true distribution if known (TD) or the KDE distribution. The error is computed only at learning sample locations. Different databases have been used for the comparison, from the monodimensional ones used before to real databases:

- B_1 is drawn from a two gaussian mixtures ($N(0, 2^2, 0.6)$ and $N(7, 4^2, 0.4)$). It contains 3000 monodimensional features.
- B_2 is drawn from a uniform distribution between 1 and 10, it also contains 3000 monodimensional features.
- B_{ripley} containing 2 different classes with 125 2D learning features for each class. Result criteria have been computed on each class separately and averaged.

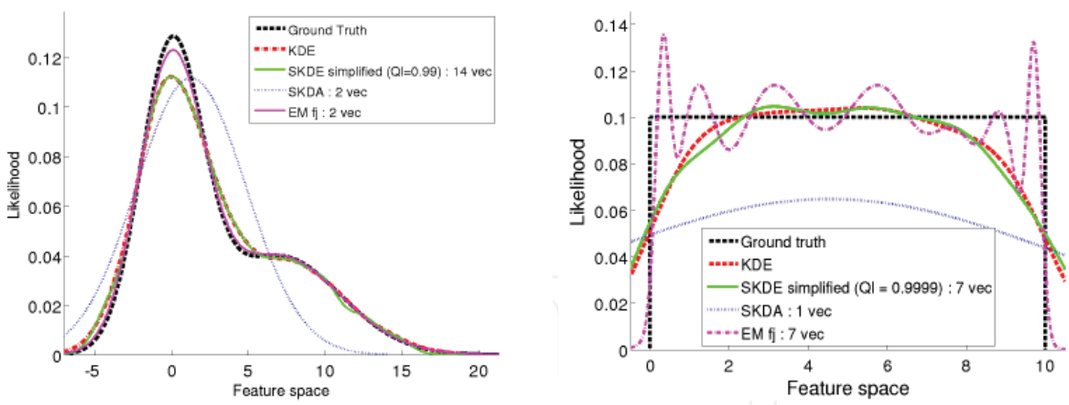
In order to conveniently compare the different methods, we assume that observations follow a gaussian mixture distribution. As a consequence, the kernel chosen for the KDE and SKDE, will be gaussian one. The comparison results are summed up in table 3. The parameter set for each method are the same than those used in figure 11 for databases B_1 and B_2. They have been chosen in order to have the closest results to the true distribution. For B_{ripley}, the parameter of each method have been chosen in order to have the best classification result as shown in section 3.2.3. Compared to other method, SKDE gives similar results in term of precision. As expected, we are very close to KDE distribution since it is the refered distribution. Concerning the number of support vectors kept, we largely reduced the KDE model, but we generally keep more vectors than SKDA or EM method. The reason is the kernel fixed bandwidth, that may need several gaussians for approximating a unique one with larger bandwith whereas SKDA or EM method will just adapt the bandwidth. On more difficult distribution such as uniform one which are not easily approximated by gaussian mixture, we see that our method fits quite well to the true distribution. For the learning computation time, the time given in number of cycles, should be taken with care. All the algorithm have been run under Matlab, the times presented are given for information only since algorithms coding are not necessarily optimized and EM algorithm complexity is unknown. The SKDA has a linear time complexity and it is clearly the fastest method but also the less accurate which is the exact opposite of EM algorithm. SKDE method is balanced between the two. Most of SKDE computation time is due to the Φ computation which is $O(K^2)$. Concerning B_{ripley} database, no true distribution is known. As a consequence, the comparison is done with KDE distribution. The very large MISE of EM algorithm are not due to wrong gaussian means but to overestimation. Moreover, our method with a coarse approximation (only 2 vectors kept) still gives comparable results with other methods.

A graphical representation is given on figure 11. The true distribution, that is to say the original gaussian mixture from which the learning observation have been drawn, is represented with the black dashed curve. Note that, the KDE distribution does not necessarily fit perfectly the true distribution. Theoretically the KDE converges to the true distribution for an infinite number of observations, whatever kernel bandwidth. Here the learning set is 3000 features long. As a consequence, the bandwidth selection is very important. For the moment this bandwidth is experimentally chosen. It should be large enough to avoid the KDE

Databases	Method s	MISE / TD	MISE / KDE	Support vectors	Computation time
B_1	KDE	$7.02e^{-5}$		3000	
	SKDE	$6.93e^{-5}$	$\mathbf{6.71e^{-7}}$	14	4.23
	SKDA	$4.31e^{-4}$	$2.76e^{-4}$	**2**	**0.13**
	EM	$\mathbf{8.13e^{-6}}$	$3.1e^{-5}$	**2**	163.06
B_2	KDE	$2.64e^{-4}$		3000	
	SKDE	$2.8e^{-4}$	$\mathbf{6.6e^{-6}}$	**7**	4.3
	SKDA	$1.76e^{-3}$	$1.22e^{-3}$	1	**0.13**
	EM	$\mathbf{1.8e^{-4}}$	$3.65e^{-4}$	**7**	180.12
B_{ripley}	KDE			150	
	SKDE		$\mathbf{2.2e^{-3}}$	2	**0.005**
	SKDA		$3.8e^{-3}$	**1**	**0.0005**
	EM		1.75	3	0.237

Table 3. Learning results. Each methods is evaluated on different criteria: MISE compared to true distribution, MISE compared to KDE, number of support vectors and learning computation time expressed in billion of clock cycle.

distribution to look like a Dirac comb, each pseudo Dirac being the gaussian of a learning feature, but also not too large in order not to melt different modes in one. We can see on the two mode distribution that except for SKDA, the other methods are quite similar and have roughly found the two modes. The second one is just slightly underestimated. On the uniform distribution, EM gives an oscillating approximation whereas SKDA approximate the square by a very large gaussian which is not acceptable. Our method fit quite well to KDE as expected.



(a) Two gaussians distribution (b) Uniform distribution

Fig. 11. Density estimation algorithm comparison.

3.2.3 Classification comparative results

This section propose to test our likelihood approximation in a learning machine context. The classification decision rule for all the method is the same, deduce from confidence criteria presented in section 3.1.3 to take into account likelihood distribution kurtosis. B_{ripley} database has been used for this comparison. The learning has been done on each class separately and tested on a thousand features, half from one class and half from the other one. The ROC

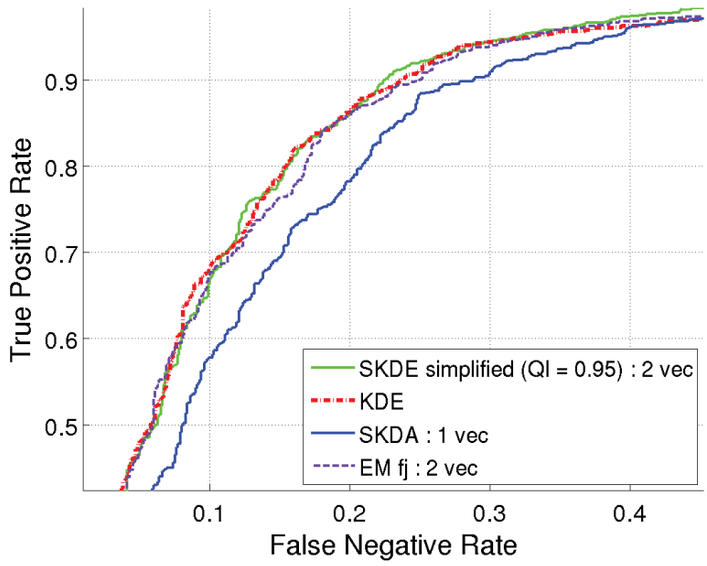


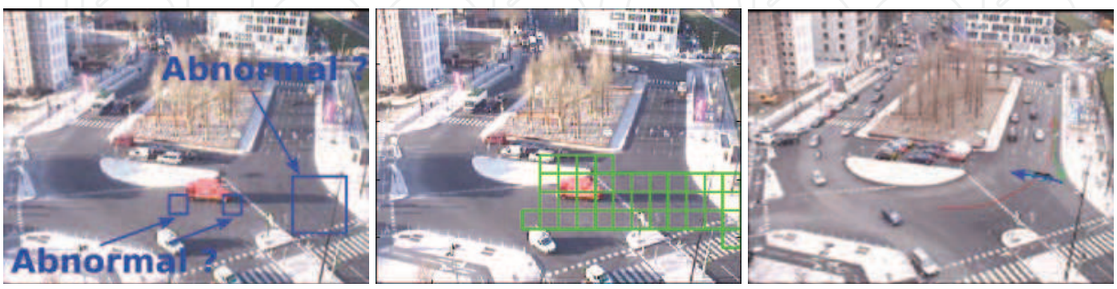
Fig. 12. ROC curves comparison.

curves on figure 12 show that the proposed method gives the best classification results with a reasonable number of control points (2 points).

4. Global experiments

4.1 Quantitative results

Giving quatified performances for such a kind of system is a difficult task. If a wrong way movement is clearly an anomality, other deviating movements can be harder to classify. Anyway, in order to give quantitative results we define permissive ground truth. We say permissive because defining exactly which blocks to consider as abnormal for every frames is impossible. Is shadow part of the anomaly ? What about neighbouring blocks ? etc. (cf. figure 13(a)). As a consequence the defined ground truth is spatially blurred on purpose (cf. firefighter truck going wrong way on figure 13(b)) but also temporally because first, defining the exact frame an event begins or ends is impossible.



(a) How to define groung truth ? (b) Ground truth example. (c) Augmented reality example.

Fig. 13. Evaluation database creation and definition.

With such a ground truth, we choose to make a frame counting for good detection and false alarms. ROC curves will be used for comparing descriptors and decision functions. A true

positve is raised when at least one block in the ground truth is considered as abnormal at time t and one false positive when the block is outside the ground truth. Note that with the temporal blurring on ground truth definition, the true positive rate is decreased. **The ROC curves are not really well-shaped but since ground truth is the same for all the appoaches, comparing ROC curves is still valid**.
Roc curves have been drawn on a synthetic database with artifical events. Real sequences of a complex crossroad have been used for inserting a textured object following a user-defined trajectories (cf. figure 13(c)). The inserted object respect the scene perspective but is not photo-realistic since no 3D model of the scene was available. 15 abnormal trajectories with 9 different textures have been used, for creating a total of 135 video containing abnormal behaviours, that is to say about half an hour. Videos are 320x240 size at 12 fps. Trainings have been done on 33 real videos of 30 seconds each, with various illumination and weather conditions. Decision functions have been computed on another 24 real videos representing normal situations.
First of all, the influence of the decision function, the confidence threshold is compared with a fix threshold for all the blocks. The same descriptor (SSC) is used with SKDE as a machine learning. We can see on figure 14 that confidence threshold (red plain curve) improve classification results compared to a static threshold (blue dashed curve). Adapting detection threshold depending on the distribution shape is usefull to lower detection sensibility on area where movement is not well-defined (every direction may be seen) and to raise it in the opposite case. The improvement in classification context for the proposed descriptor is also shown on figure 15. SSC descriptor (red plain curve) has been compared with traditional optical flow features (blue dashed curve). The main orientation per block for optical flow feature is obtained with SKDE process ensuring to keep only the first found control point ($\tilde{K} = 1$). Once again, the proposed descriptor improves the classification task, decreasing ponctual false alarms and smoothing the detections. Only SSC descriptor will be used in the rest of these tests.

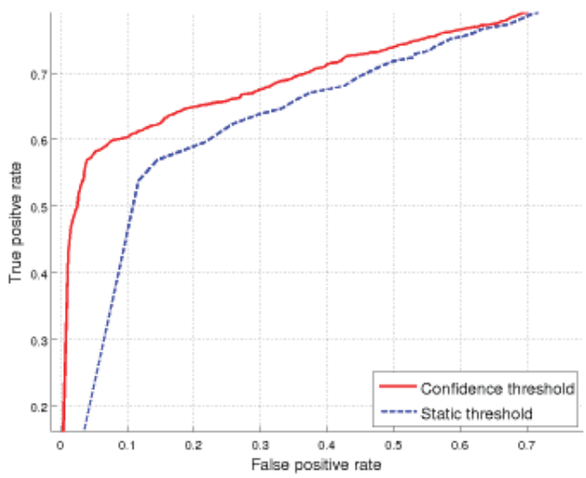


Fig. 14. ROC curves comparison between confidence threshold and static theshold decision function.

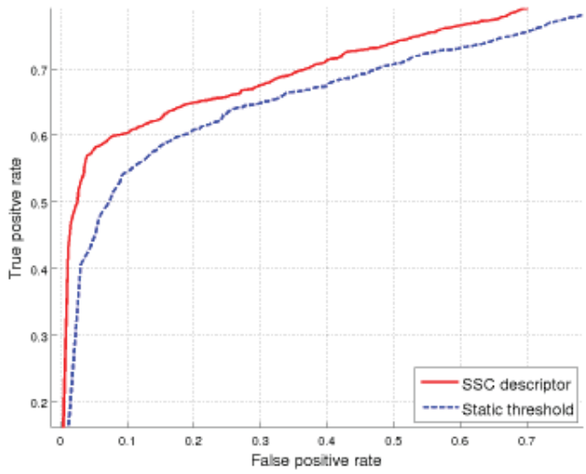


Fig. 15. ROC curves comparison between descriptors.

In order to evaluate the proposed algorithm in a more realistic context, we defined an event alarm when at least one bloc per frame is classified as abnormal on K consecutive frames, in the same neighbourhood. This filtering remove the remaining ponctual false alarms and give a more robust answer since an event usually lasts several seconds and propagates from one bloc to the adjacent ones. A diturbance rate is also defined as the coresponding false alarm rate with such a filtering.
For example, with $K = 8$ the proposed system is able to detect up to 70% of right event detection from a total of 145 events among the 135 videos analysed. With such a detection rate, the disturbance rate is less than $0,2\%$, representing less than 2 wrong alarms per hour on average. Such performances fit well with a video assistance system requirements, that is to say beeing able to detect most of the main problems while ensuring a low false alarm rate which can be very annoying for operators.

4.2 Qualitative results

To describe what kind of event can be detected thanks to the proposed application framework, different examples of detections in various illimination (indoor/outdoor sequence) and weather conditions are presented on figure 16. We can see that various events can be detected such as jaywalkers, wrong way movement, argument between people, etc. Conditions can be very different in terms of illumination with night detections in particularly hard conditions but also in term of population or traffic density with wrong way pedestrian detections in marathon crowd for example.

5. Conclusion

Crowded scenes are particularly difficult to analyse because of the large amount of information to be processed simultaneously and the complexity of the scenes. Tracking based systems cannot handle numerous targets at the same time. In this paper we consider the crowd as a whole. We propose a new framework that cut the problem in two, the movement characterisation and the learning and classifying procedure. Two main contributions can be pointed out.

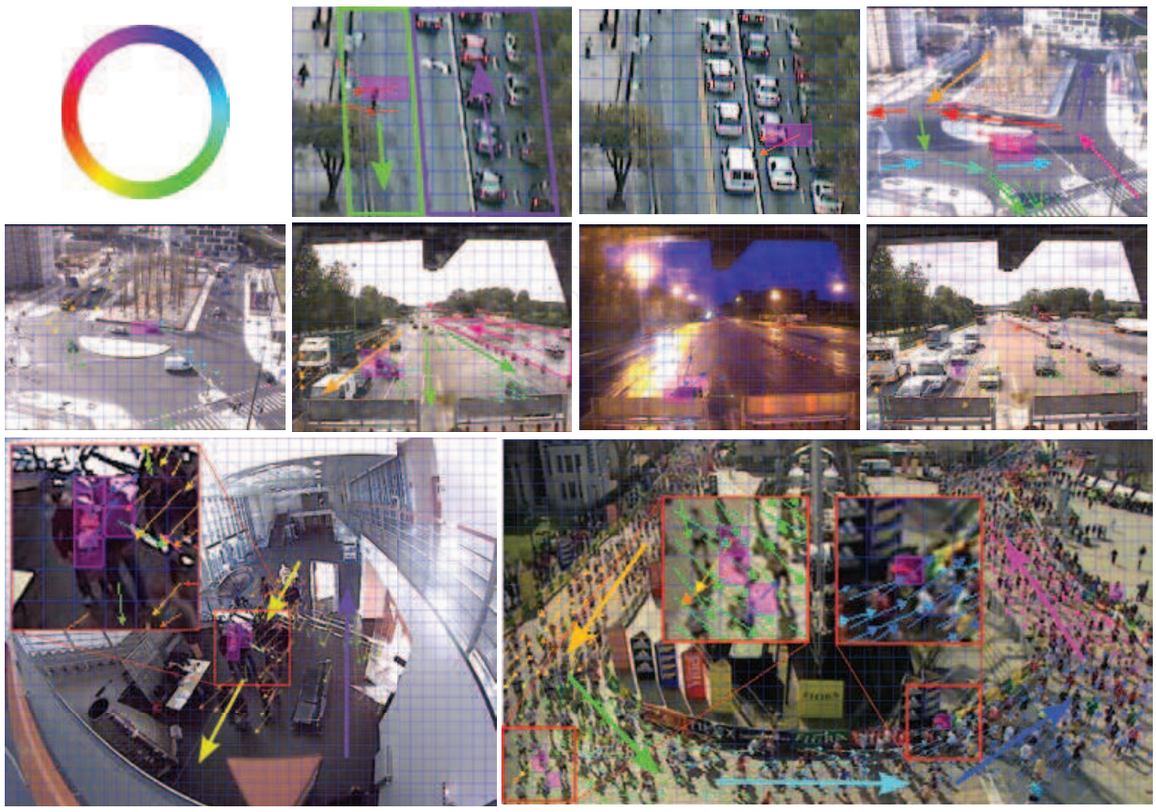


Fig. 16. Example of detection such as jaywalkers, wrong way movement, car pulling out or chaotic movement due to people argument, etc. All the arrows respect the color legend given in the first on the top left image. Ground truth for each scene is represented with large arrows.

The first one is a new method of movement characterisation. We study the global scene movement thanks to a new descriptor based on a spatio-temporal structure. This descriptor outperforms other spatio-temporal descriptors studied in terms of movement separation. Moreover, it is invariant to affine illumination changes which is particularly usefull when treating outdoor sequences.
The second main contribution of this paper is a new framework for modelling motion pattern of any scene with structured motion. The proposed framework relies on a new density estimation method which is a sparse representation of the KDE distribution, adapted to real-time evaluations. This method gives results of same quality as other classical algorithms aiming at retrieving gaussian mixture parameters, but with a better compromise between precision, sparseness of the model and time computation.
Moreover, our approach requires neither camera calibration nor any 3D scene model, the learning phase is unsupervised and thus the framework applies to a large number of scenes such as outdoor or indoor areas, traffic or crowd monitoring, etc. It works under various illumination and weather conditions but also with various population or traffic density. It can reveal subtle perturbations in the global motion, such as wrong ways or movement deviations, jaywalkers dangerous behaviours or chaotic movements due to abnormal interactions between people of a crowd when they are arguing for example.

Currently, we are investigating temporal and spatial consistency in movement propagation through more sophisticated modelisation. We are also studying block size adaptation with a multiscale approach in order to adapt automatically to both scale change due to strong perspective projections or large movements that should need a lower resolution to be perceived conveniently.

6. References

Adam, A., Rivlin, E., Shimshoni, I. & Reinitz, D. (2008). Robust real-time unusual event detection using multiple fixed-location monitors, IEEE *Transactions on Pattern Analysis and Machine Intelligence* **30**(3): 555–560.

Andrade, E. & Blunsden, S.and Fisher, R. (2006). Hidden markov models for optical flow analysis in crowds, *Proceedings of the International Conference on Pattern Recognition*, Vol. 1, pp. 460–463.

Bardet, F., Chateau, T. & Ramasasan, D. (2009). Illumination aware mcmc particle filter for long-term outdoor multi-object simultaneous tracking and classification, *Proceedings of the International Conference on Computer Vision*.

Barron, J., Fleet, D., Beauchemin, S. & Burkitt, T. (1992). Performance of optical flow techniques, *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, Vol. 92, pp. 236–242.

Black, M. & Anandan, P. (1996). The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields, *Computer Vision and Image Understanding* **63**(1): 75–104.

Blunsden, S., Andrade, E. & Fisher, R. (2007). Non parametric classification of human interaction, *Pattern Recognition and Image Analysis*, pp. 347–354.

Breitenstein, M., Grabner, H. & Van Gool, L. (2009). Hunting nessie: Real time abnormality detection from webcams, *Proceedings of the International Conference on Computer Vision - Workshop on Visual Surveillance*.

Brostow, G. & Cipolla, R. (2006). Unsupervised bayesian detection of independant motion in crowds, *Proceedings of the International Conference on Computer Vision and Pattern Recognition*.

Chan, A. & Vasconcelos, N. (2008). Modeling, clustering, and segmenting video with mixtures of dynamic textures, IEEE *Transactions on Pattern Analysis and Machine Intelligence* **30**: 909–926.

Dee, H. & Velastin, S. (2008). How close are we to solving the problem of automated visual surveillance ? : A review of real-world surveillance, scientific progress and evaluative mechanisms, *Machine Vision Applications* **19**(5-6): 329–343.

Dempster, A., Laird, N. & Rubin, D. (1977). Maximum likelihood from incomplete data via the em algorithm (with discussion), *Royal Statistical Society* **B 39**: 1–38.

Duda, R., Hart, P. & Stork, D. (2001). *Pattern Classification*, John Wiley & Sons Inc.

Figueiredo, M. & Jain, A. (2002). Unsupervised learning of finite mixture models, IEEE *Transactions on Pattern Analysis and Machine Intelligence* **24**: 381–396.

Han, B., Comaniciu, D., Zhu, Y. & Davis, L. (2008). Sequential kernel density approximation and its application to real-time visual tracking, IEEE *Transactions on Pattern Analysis and Machine Intelligence* **30**: 1186–1197.

Harris, C. & Stephens, M. (1988). A combined corner and edge detector, *Proceedings of the Alvey Vision Conference*, pp. 147–151.

Horn, B. & Schunck, B. (1981). Determining optical flow, *Artificial Intelligence* **17**: 185–203.

Hu, W., Xiao, X., Fu, Z., Xie, D., Tan, T. & Maybank, S. (2006). A system for learning statistical motion patterns, IEEE *Transactions on Pattern Analysis and Machine Intelligence* **28**: 1450–1464.

Junejo, I. & Foroosh, H. (2007). Trajectory rectification and path modeling for video surveillance, *Proceedings of the International Conference on Computer Vision*.

Kratz, L. & Nishimo, K. (2010). Tracking with local spatio-temporal motion patterns in extremely crowded scenes, *Proceedings of the International Conference on Computer Vision and Pattern Recognition*.

Kratz, L. & Nishino, K. (2009). Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models, *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, pp. 1446–1453.

Küttel, D., Breitenstein, M., Van Gool, L. & Ferrari, V. (2010). What's going on ? discovering spatio-temporal dependencies in dynamic scenes, *Proceedings of the International Conference on Computer Vision and Pattern Recognition*.

Langford, E., Schwertman, N. & Owens, M. (2001). Is the property of being positively correlated transitive ?, *The American Statistician* **55(4)**: 322–325.

Lucas, B. & Kanade, T. (1981). An iterative image registration technique with an application to stereo vision, *Proceedings of the International Joint Conference on Artificial Intelligence*, pp. 674–679.

Mahadevan, V., Li, W., Bhalodia, V. & Vasconcelos, N. (2010). Anomaly detection in crowded scenes, *Proceedings of the International Conference on Computer Vision and Pattern Recognition*.

Mehran, R., Oyama, A. & Shah, M. (2009). Abnormal crowd behavior detection using social force model, *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, Vol. 0, pp. 935–942.

Oliver, N., Rosario, B. & Pentland, A. (2000). A bayesian computer vision system for modeling human interactions, IEEE *Transactions on Pattern Analysis and Machine Intelligence* **22**: 831–843.

Paalanen, P., Kamarainen, J., Ilonen, J. & Kälviäinen, H. (2006). Feature representation and discrimination based on gaussian mixture model probability densities : Practices and algorithms, *Pattern Recognition* **39**: 1346–1358.

Pham, Q., Gond, L., Begard, J., Allezard, N. & Sayd, P. (2007). Real time posture analysis in a crowd using thermal imaging, *Proceedings of the International Conference on Computer Vision and Pattern Recognition*.

Rabaud, V. & Belongie, S. (2006). Counting crowded moving objects, *Proceedings of the International Conference on Computer Vision and Pattern Recognition*.

Saad, A. & Shah, M. (2007). A lagrangian particle dynamics approach for crowd flow segmentation and stability analysis, *Proceedings of the International Conference on Computer Vision and Pattern Recognition*.

Saleemi, I., Shafique, K. & Shah, M. (2008). Probabilistic modeling of scene dynamics for applications in visual surveillance, IEEE *Transactions on Pattern Analysis and Machine Intelligence* **PP, Issue 99**: 1–1.

Shechtman, E. & Irani, M. (2005). Space-time behaviour based correlation, *Proceedings of the International Conference on Computer Vision and Pattern Recognition*.

Sidla, O. & Lypetskyy, Y. (2006). Pedestrian detection and tracking for counting applications in crowded situations, *Proceedings of the International Conference on Advanced Video and Signal Based Surveillance*.

Stauffer, C. & Grimson, W. (2000). Learning patterns of activity using real-time tracking, IEEE *Transactions on Pattern Analysis and Machine Intelligence* **22**: 747–757.

Tyagi, A., Keck, M., Davis, J. & Potamianos, G. (2007). Kernel-based 3d tracking, *Proceedings of the International Conference on Computer Vision and Pattern Recognition*.

Varadarajan, J. & Odobez, J. (2009). Topic models for scene analysis and abnormality detection, *Proceedings of the International Conference on Computer Vision - Workshop on Visual Surveillance*, Kyoto.

Wang, X., Ma, X. & Grimson, W. (2009). Unsupervised activity perception in crowded and complicated scenes using hierarchical bayesian models, IEEE *Transactions on Pattern Analysis and Machine Intelligence* **31**(1): 539–555.

Wang, X., Tieu, K. & Grimson, W. (2010). Correspondence-free activity analysis and scene modeling in multiple camera views, IEEE *Transactions on Pattern Analysis and Machine Intelligence* **32**(1): 56–71.

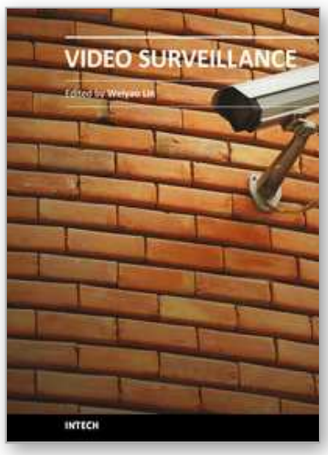
Wu, S., Moore, B. & Shah, M. (2010). Chaotic invariants of lagrangian particle trajectories for anomaly detection in crowded scenes, *Proceedings of the International Conference on Computer Vision and Pattern Recognition*.

Yu, Q. & Medioni, G. (2009). Multiple-target tracking by spatiotemporal monte carlo markov chain data association, IEEE *Transactions on Pattern Analysis and Machine Intelligence* **31**(12): 2196–2210.

Zhan, B., Monekosso, D., Remagnino, P., Velastin, S. & Xu, L. (2008). Crowd analysis: A survey, *Machine Vision Applications* **19**: 345–357.

Zhao, T. & Nevatia, R. (2004). Tracking multiple humans in complex situations, IEEE *Transactions on Pattern Analysis and Machine Intelligence* **26**: 1208–1221.

Zhong, Z., Yang, M. & Wang, S. (2007). Energy methods for crowd surveillance, *Proceedings of the International Conference on Information Acquisition*, pp. 504–510.



Video Surveillance

Edited by Prof. Weiyao Lin

ISBN 978-953-307-436-8
Hard cover, 486 pages
Publisher InTech
Published online 03, February, 2011
Published in print edition February, 2011

This book presents the latest achievements and developments in the field of video surveillance. The chapters selected for this book comprise a cross-section of topics that reflect a variety of perspectives and disciplinary backgrounds. Besides the introduction of new achievements in video surveillance, this book also presents some good overviews of the state-of-the-art technologies as well as some interesting advanced topics related to video surveillance. Summing up the wide range of issues presented in the book, it can be addressed to a quite broad audience, including both academic researchers and practitioners in halls of industries interested in scheduling theory and its applications. I believe this book can provide a clear picture of the current research status in the area of video surveillance and can also encourage the development of new achievements in this field.

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen