Peaceful Monitoring of Crowds

Pierre BERNAS, Guillaume NEE, Philippe DRABCZUK[1]

[1]EVITECH – www.evitech.com – +33.820.2008.39

1. Introduction

« Peaceful Management of Crowds » is probably the closest English translation of the French expression[1] which relates to techniques for checking and maintaining Public Order in situations with large crowds, and/or particularly dense and mobile crowds such as those observed in large transport infrastructures, events or parks, in demonstrations, or in large cities centers (all such places are termed "sites" in this paper).

Here, we propose a subdivision of this "***Management***" activity of "***Crowd controllers***[2]" into a "*Preparation*" step, consisting in prior actions taken before the crowd arrives (cf. practices, organization in [25]) ; a "*Monitoring*" step, which consists in observing and establishing a diagnosis of the situations at different spots of the sites once the crowd is there ; some "*Action*" steps, which concern "live" interactions with the crowd (via such infrastructure means, as sounds, lights, billboards, or by people, such as order and rescue) ; and finally a "*Feedback*" step, which capitalizes on all the crowd controllers' knowledge once the crowd situation is terminated (e. g. in [25], a review of recorded videos is proposed as a good practice to assess and improve crowd control procedures).

In this field, it is crucial to ensure that (1) **everybody arrives safely at his/her target place**, and (2) **following the authorized paths**. Individual questions (such as identity, behavior)[3] are not relevant for peaceful crowd management.

However, many technical or social science studies (e. g. [22]) have addressed the "*Monitoring*" step with a behaviorist point of view, considering that information about observed behavior is a result that should simply be obtained or deduced from observation, and hence that this behavior would reveal any possibly dangerous situations. Fortunately or not, this approach seems a little too restrictive.

Hence, we would like here to generalize this "*Monitoring*" concept, firstly because we consider that "behavior" relates to privacy, and is not immediately related to Public Order Maintenance, and further because:

- **Crowd observation contributes to the detection of invisible events that those observed perceive** (even though they are not taking part in these events), e. g. by stopping suddenly and staring at some point, or by fleeing suddenly,
- major accidents in Crowds can be the consequence of successive situations which may begin by non accidental events or benign people motion and arrangements. Thus the observation of the latter could help prevent similar accidents.

The first concept above consists in using crowd itself as a detector of incidents in the neighborhood.
The second concept above was already referred to under the name of "***weak signals***" in the ESRAB report [1] in 2006. However, although it is obvious that sometimes "*weak signals*" precede an accident (e. g. in a

---

[1] In the French Security « doctrine » : la *Gestion pacifique des foules*.
[2] According to terminology used in the Victorian crowd control safety guide [25].
In France, crowd controllers can be relevant from Police/Army forces, Emergency forces (Firemen), or Private security organization.
[3] Questions such as "*Who was there ?*", "*Who spoke with whom ?*" or "*What origin/sex/age/skin colour was this group ?*" or "*Did somebody have a strange behavior ?*"

situation where flows of people take stairways downwards to a place where similar flows of people cannot get out or escape at the same velocity), it is neither obvious nor demonstrated that "*weak signals*" always precede accidents (e. g. someone pushed on rail tracks or terrorist attacks). However, considering the variety of accidents that may occur, it is clear that some "weak signals" are often available. For example, in a situation where many people are choking because of a small gas leak, firstly identified as "a smell", this "smell" would precede a bigger gas leak, leading everybody to faint and fall on the ground. In this case, choking of a fraction of the crowd when passing a given spot would be a weak signal.

This paper stems from work done in the framework of the RAPID DGA CrowdChecker project, led by the Evitech SME and the Willow research lab (ENS/CNRS/INRIA) between October 2010 and August 2012.

We present a tool for **peaceful monitoring of crowds**. We will first detail our aims in terms of "*Monitoring*", and then show how we developed video analytics monitoring tools, in the framework of this project.

2. Crowd monitoring aims

2.1. Accident criticalities

When we examine worldwide calls for tenders concerning crowd monitoring systems in the last 5 years (e. g. for transport sites, museums, cities, religious sites, shopping centers…), we note that **most requirements** finally **address people counting**. Even if the terms of security, density, and danger are mentioned, the ultimate requirements lead to people counting: detect groups, count people, detect if an attendance level is reached, etc.

Now crowd observation cannot be reduced to counting. Counting people, of course, is very useful for detecting overcrowding of a site, or for checking the business done in a shop, or for adapting the rent of a shop surface to the attendance in a shopping center. It is also somehow related to interactions between people in the streets, which might at some point cause "troubles", because of an unexpected number of people. Hence, counting is useful, but **should not be considered the ultimate destiny of crowd monitoring**.

Examining the criticity of aeronautical equipment in civil aircraft has led to normative documents (DO178 [2]) that characterize criticity as **the loss that a failure of that equipment would produce.** Events leading to loss of the aircraft and death of all passengers is most critical (level A), while inconveniences to the cabin crew is at a lower level (D), and what does not affect the flight is level E. Failure of equipment leading to a death or the injury of several people would be intermediate (levels B/C). The mention of a risk matrix also appears in [25] as a tool for crowd incidents evaluation.

Likewise, in Crowd monitoring, the most critical situations (level A), that are to be detected as fast as possible, is the collective accident where we would observe a large number of people "on the ground" : due to the fall of everyone in all or part of the site, or in part of the site (whatever the cause : individuals crushed in panic, gas effects, temperature, explosion, earthquake, etc). At somewhat lower levels (B, C), the most important situation to be detected is the **precursor** of such collective accidents, or a situation where several people are injured (e. g. a fast car in crowd), etc. On a still lower level, there are cases that should alert the security team since they may be of interest, possibly endangering one of several people, and later, cases inducing crowd flows efficiency problems (level D).

2.2. System logical paradigm

In order to ensure *that everybody will safely and legally follow his/her path in the collective motion*, there are a few "control" parameters of interest, regarding the progression of people:

- ***directions of motion***, and their spatial distribution,
- ***speeds***, and their spatial distribution,
- ***crowd density***, and spatial distribution.

This is of course our **first** target.

Secondly, we are interested in a system that is essentially **determinist**. A determinist system is a system governed by causality. We define a condition that implies a rule, and we want this rule to be activated when the condition occurs. This is inherent to law and security : no room left for randomness.

A *non-determinist* system would typically be a system fully based on a “training” or a “learning” phase occurring in some large and unknown or uncontrolled crowd characteristics. Later, because of the data collected and organized inside the system, some learnt rules would trigger an alarm when a situation has characteristics that conform to the learnt danger configurations. Such systems need very large learning data, and their convergence is hard to obtain (particularly when targeting 100% precision), they can raise many unexpected alarms, and they can miss important events because these events were never seen before, and thus not learnt before. This is hopefully the case with crowd accidents: they are rare, and thus there is a very small learning corpus about crowd accidents.

Hence, if a learning strategy seems necessary in order to improve the system as the time goes by and new weak signals are found, we propose that the crowd monitoring system be based on deterministic principles, with supervised evolution capacities from learning, limited to weak signals, and after manual supervision.

Following this approach, we have identified several generic rules in terms of crowd monitoring :

- **Immediate detection of a "dangerous situation"** which should lead to a warning to crowd controllers,
- **Immediate detection of identified and known accident precursor situations**, which have been established on experience, either on site or from other site observations,
- Collection of **statistics** based on the aforementioned parameters in order :
  - to warn of a dangerous situation (specifically, the repartition of crowds between different parts of the site can be of interest, if not balanced),
  - to compare with other similar “days” or “events” (in order to be aware of unexpected conditions, especially **if the number or the equipment of crowd controllers is not appropriate** or if assumptions made in the *preparation step*[4] were inappropriate),
  - and to compare the current aforementioned parameters with those preceding known accidents in order to provide a weak signal detection, and possibly also to identify a situation which would be candidate to become a known precursor situation.

2.3. Dangerous situations

We have thus identified the following needs, which will be complemented as experience accumulates in the future:

- Detection of several/many people who fall,
- Detection of threshold crossing (expressed in speed, density, possibly compounded with direction conditions, such as detection of high density -e. g. crushing risk-, or high speeds in several directions -e. g. collision risk-, or people running in a given direction -e. g. risk of falling from a train platform-).
- Detection of a car/truck or a big object entering the crowd (possibly quickly, endangering people in the crowd),
- Detection of smoke growing from the crowd or reaching the crowd (fire/choking danger),
- Detection of sudden dispersion of a dense immobile group in sparse mobile crowd (e. g. possibility of a left victim, or manifestation of sudden fear...).
- Detection of one or several people walking in non-authorized direction (e. g. climbing in an aircraft by the exit),
- Detection of somebody crossing a crowd flow, possibly with a speed condition (e. g. entering by the exit, or suicide-bomber progressing quickly to control),
- Detection of somebody stopping (or walking in reverse direction of) the crowd flow it formerly belonged to (trouble in overall people progression),
- Detection of a dense immobile group formed in sparse mobile crowd (trouble in the overall people progress),

2.4. Events to detect, using crowd as a detector

As noted above, there are also situations where the observed crowd is itself a detector of events that are not seen by the cameras. We should hence detect situations such as:

- Many people suddenly run, or stop (fear),

[4] Cf. the 4 steps proposed in §1.

- Crowd flow suddenly changes (the place is suddenly empty, or, contrarily, there is a sudden rush of people, …),
- Detection of some “hole” formed in the mobile crowd, and crowd flows passing on both sides of the hole (possible fall or incident not visible on the ground because of people around),
- Likewise, detection of the formation of a standing group inside a mobile crowd, perhaps revealing an invisible fall in the middle, or a fight.

2.5. Weak signals

Lastly, we need a statistical system in order to collect and compile observation data, for the various needs expressed above, and to allow the definition of accident precursors. Such a statistical system is useful, when several successive accidents occur, in order to facilitate the "*Feedback*" step mentioned in §1, e. g. for place layout, crowd flow indicators. In that system, supervised learning algorithms may help collecting remarkable similarities preceding several different accidents (candidates to become a weak signal).

3. Crowd monitoring by video analytics

In the CrowdChecker project, we wished to propose and demonstrate a tool for automatically supporting such monitoring.

3.1. Conditions of observation of crowds

We chose conditions as close as possible to real conditions in transport sites, cities, museums, etc, and used previously installed indoor and outdoor CCTV color cameras. We developed a video analytics solution for real time monitoring of crowds. We assumed that the crowd participants would not massively carry large Mexican hats, or umbrellas –this is is almost always the case indoors, and outdoors in decent weather for dense crowds (*Note that bad weather conditions -very hot, or very rainy- have often been observed to reduce crowd density…*).

Most of these cameras are installed in ceilings and observe the crowds from above. However, the limited visibility of people in a crowd is an issue, as we will detail here.

Suppose a camera observes a crowd from above in a flat area. If the camera has an angle α with the flat ground (in typical installations receiving crowds, $10° < \alpha < 45°$), and if somewhere in the crowd two persons are following each other in the direction of the camera at a head top distance of *D*[5], if the second one is smaller by $D*tan(\alpha)$ or more, he or she is completely hidden by the first one from the camera.

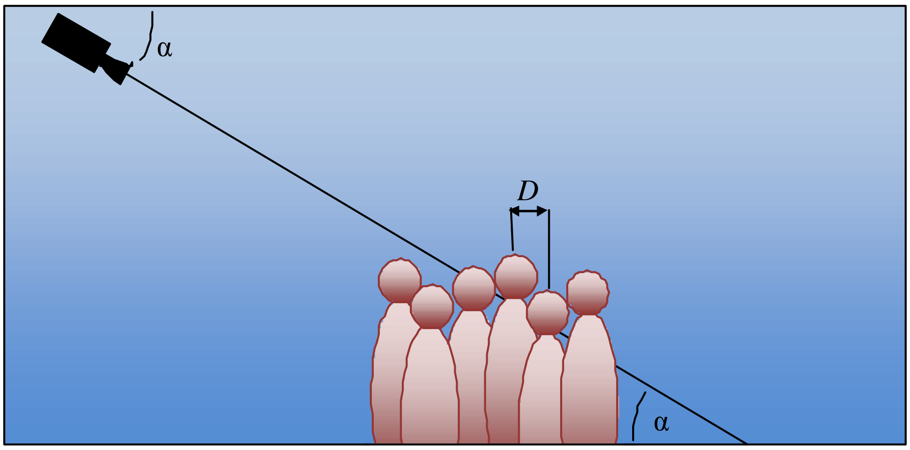

Figure 1 : crowd observation

Assume that people's sizes are distributed according to a Gaussian law with mean *m* and standard variation *s*. The distribution of size differences between two neighbors is then a Gaussian law of mean *0* and standard variation $s*\sqrt{2}$. The probability that a head top is hidden by the head just before it in the camera field is the Gaussian tail beyond $D * \tan(\alpha) / (s * \sqrt{2})$ for a Gaussian law, centered at, and standard *N (0, 1)*. According to values of *D* and α, probability values are given by Gaussian law tables.

---

[5] *D* is the distance between two successive head tops, typically 50 cm in a dense crowd, 1 meter or more in a sparse crowd.

If we assume that the standard deviation in a human population is e. g. *15 cm* (which sounds reasonable for men, women, children), we have the following *probabilities* of hidden head tops, at the center of the camera image :

Average angle of view	Dense crowd (*D=50 cm*)	Sparse crowd (*D=1 m*)
α=45°	1%	-
α=30°	8,7%	0,33%
α=20°	19,5%	4,4%
α=10°	33%	20,4%

Table 1 : camera hidden head tops in crowd

If the camera is far enough from the crowd (many people seen) this probability can be interpreted as the *average proportion* of head tops at the center of the image that are hidden.

If we have a camera with a lean angle of 20° and a vertical field of view of 20° or more, there will be almost no hidden head top at the bottom of the image, 4,4% at 20°, and more than 20% that are just completely hidden at 10° or above.

Now, if we specify that in order to detect a head, we need for example at least 5 cm[6] of height of the top of this head to be visible, these figures are dramatically increased (e. g. for *α=20°*, and *D=50 cm*, **27%** of heads are not visible over 5 cm height (for 19,5% of the heads that are not visible at all), and **7%** (instead of 4,4%) in case of sparse crowd, when *α=20°* and *D=1 m*).

In conclusion, it is important to notice that the best observations (counting, event detection...) using existing installed CCTV cameras can be done at the bottom of most camera scenes (at angles of view between 45° and 90° off the horizon). It soon becomes uncertain or completely impossible at angles between 20° and 0° off the horizon. Without hats or umbrellas, people's heads which are very close to the camera are thus all quite easy to observe (large in pixels, good angle of view, good visibility), while further from the camera they become very quickly far more difficult to observe (small in pixels, angle of view leading to head overlap and hiding, and thus bad visibility).

As a general rule of thumb, it is important to notice that counting in dense crowd by a camera leads to underestimates of order of 10% to 40% or more when the angle varies from 30° to 10° above the ground.

As a result of this task, we integrated a 3D calibration model of the image in order to be able to estimate the real size of a head at a given position in the image.

3.2. Subtracting background

Many video analytics systems based on background subtraction techniques [20]. These techniques consist in establishing a stability model of the "background" of the scene (visible ceiling, ground, walls, ...), over which several targets are moving. The basic assumption of these techniques is the assumption that target frequency and density are lower than background observability (for one pixel in the image, most of the time this pixel concerns the empty background). The background model is updated periodically in order to manage/reflect light changes (daylight, shades, day/night, etc).
When a stable change happens (e. g. when a car parks in a parking lot), this change is integrated in the background (or in a middle term background assumption) after having been immobile for a while, so the detection of new targets is enabled when passing in the foreground of this change.

However, these assumptions are no more valid in long periods of dense crowd situations. In a background model, dense crowd is a perpetually moving large blob over a rarely seen background (Furthermore, this background is heavily changed by crowd shades). Background model updating is meaningless over long periods because of this crowd blob. Furthermore, the crowd blob has fixed coordinates, and thus analysis of people's positions (expected as independent targets) is not possible with these techniques.

6 This will depend on the detection algorithm. Invisible heads cannot be counted, but there is also a minimum head size to be represented on the image to enable detection. We will see later that for example, in order to detect the Ω shape of the head-shoulders, 25 to 35 cm would be requested instead of the 5 cm assumption here.

3.3. Putting a cross on each "head"

Locating and tracking properly each person in a crowd would be a very good solution for all the questions we want to address (cf. §2.2) about density, speeds, directions of people. Knowing each person's position would be the solution for detecting almost all the situations we specified in §2.

Several approaches have been pursued in order to identify and track different combination of body-part (head, shoulders, torso, arms, …) shapes in the crowd [3,4,5,6,7,8,9,10]. Some of them have addressed the search of ellipsoidal shapes, others have addressed head shape learning. Noticing that the shape of the head over the shoulders of an individual resembles the Ω Greek character, some work has also addressed specific searches of the Ω shape [4].

In the Crowdchecker project, the project team has tried to reproduce and improve these algorithms as a general approach for head detection and tracking [11], in order to measure their performance and their ability to run at real time over one core of a good PC hardware, which would be a reasonable cost target for such a software[7].

In [11] the use of a regression-based person density estimation [24] was studied to improve person detection and tracking in crowded scenes. In this approach, a density map was used to predict head localizations.

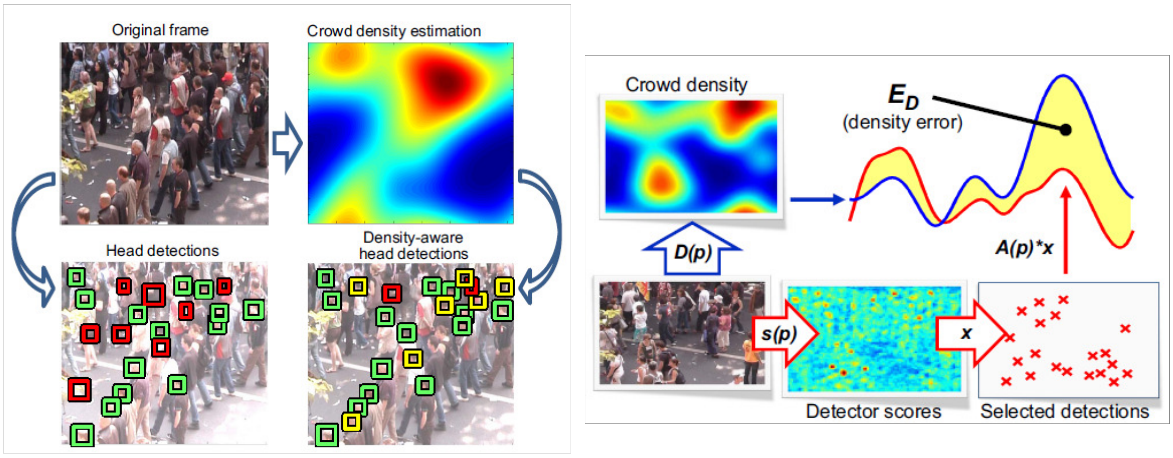

Figure 2 : overview of the person detection model

For doing this, the person detection task was formulated in a global energy minimization framework (Eq. 3.1), that involves a joint energy function incorporating scores of individual detections E_S, pair-wise non-overlap constraints E_P, and constraints imposed by the estimated person density in the scene E_D (cf. Figure 2).

$$\underset{x\in\{0,1\}^N}{\operatorname{argmin}} \left(\overbrace{-s^T x}^{E_S} + \overbrace{x^T W x}^{E_P} + \overbrace{\alpha||D - Ax||_2^2}^{E_D}\right) \qquad \text{(Equation 3.1)}$$

where x is a single *N-vector* representing all detections in the entire image ($x_i = 1$ if detection at a location p_i is valid), s is a confidence score N-vector corresponding to each location p_i, W is a $N \times N$ matrix where $W_{i,j} = \infty$ if detections at locations p_i and p_j have a significant area overlap ratio and $W_{i,j} = 0$ otherwise, D is a density regression based estimator, Ax is a matrix multiplication representing an evaluation of the density of the active detections. The $N \times N$ matrix A is designed so that each row A_i is a Gaussian window of size σ centered at p_i.

In simple words, Eq. 3.1 helps detecting heads where the crowd is the denser, and helps avoiding false head detection where the crowd is sparse.

Significant gains were obtained in detection and tracking performance on challenging videos of crowded scenes with varying density (see Fig. 3), compared to the baseline detector [23] itself (*curve a*) to the latter combined with geometric filtering involving precedence given to detection size, or to the baseline detector

[7] The cost of a performing quadcore PC is something between 1000 and 2000 €. Analyzing 4 cameras over such a hardware would thus represent a hardware cost of 250 to 500 € (to which software license, installation, configuration must be added), while costs of cameras used in these situations are of 100-1000 €.

combined with time consistency constraints using agglomerative clustering (*curve b, c*). The density-aware detector (*red curve d*) definitely outperforms all three other detectors. Detection results for the detector using *ground truth density estimates* obtained by smoothing ground truth person detections with Gaussians have also been drawn (green curves e to h) and reveal the benefits of the density-aware detector, when jointly used with a competitive density estimation method (but obviously only as a validation technique).

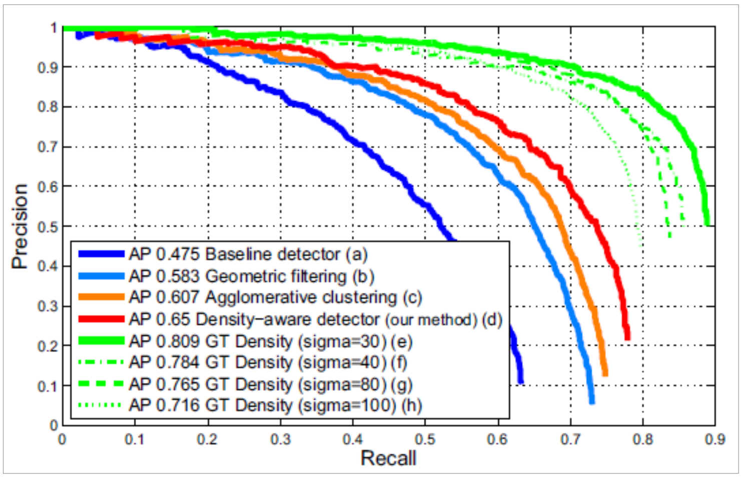

Figure 3 : precision/recall curves

Despite several improvements that were proposed by the project's research team, we found that we could not detect more than 50% of true heads, while 15% of false heads were detected ; alternatively, the numbers were 70% true heads and 40% false (red curve). Moreover, these unsatisfactory numbers required unreasonable computation time with cost effective hardware.

However, this approach can be of interest when used in a small detection area, and thus the principle was retained for local uses, with the visibility limits exposed in §3.1, since observing a full head Ω shape requires something about 30 cm, thus significantly reducing the fraction of individuals detected, especially when the distance to the camera increases.

3.4. Learning crowd situation patches

During the Crowdchecker project, the project team has also investigated a completely different approach for learning crowd situation patches (small pieces of videos from a fixed camera above the crowd, in size and in duration, that would constitute a typical situation). Here [12], these patches are aggregated via a learning strategy, and described all videos as aggregates of multiple patches, in image space and time.

This approach was found to be very useful for predicting individual paths in crowds, but unreliable when considering the entire crowd. Furthermore, it was inordinately time consuming on reasonable hardware.

3.5. Analyzing motion

Finally, we investigated the motion estimation approaches [13, 14, 19], in order to identify moving crowd areas (and thus speeds and directions) by their motion. Motion estimation has been investigated at length, but its use in crowds is complex because of the rather "brownian" motion of people in crowds. Although it has been shown [15] that very simple rules can account for people’s motion in crowds, a crowd is not made up of solid groups with homogeneous speed and direction (except in military parades!).

Different motion estimation methods were evaluated for this task :

- Optical flow methods : differential methods based on the assumption that the brightness (*I*) of a moving point (*x, y*) is constant along the time *t* : $I(x, y, t) = I(x + \Delta_x, y + \Delta_y, t + \Delta_t)$. In complement :
    - Horn and Shunck [26] assume smoothness of the global flow,
    - Lucas and Kanade [27] assume that the displacement of a pixel p is small and related to its neighboring pixels (inside a predefined window). In addition, Shi and Tomasi [28, 29] corner detection algorithm can be applied prior to motion estimation in order to select the points of

interest at which motion estimation will be applied. This method is widely used in crowd flow estimates [21].

- Block matching consists in dividing the image into small overlapping blocks and matching them in successive frames according to a similarity criterion (generally the Sum of Absolute Differences). In [22], an example of block matching techniques in a more general method for crowd monitoring is presented.

When comparing these methods, a major drawback of the Horn and Shunck method is its inability to extract motion discontinuities due to its global smoothness aspect. This is not the case for the two other previously mentioned methods, at a cost of some false motion estimation inside homogenous regions.
Ideally, we wanted to detect large crowd streams (green), as well as "long" individual tracks (red), which didn't follow these streams.

Thus, time integration of individual motions helped to identify the main crowd flows in the image, as well as smaller singularities, using a tracking complement.

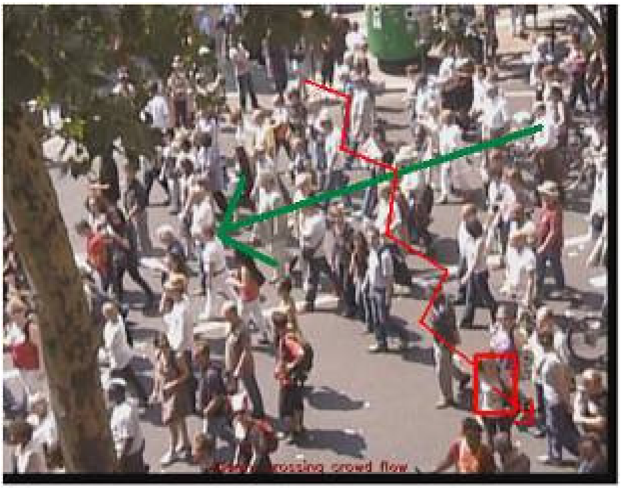

Figure 4 : crowd motion tracking

This approach was found costly in computation time, but performing at real time. Some encoding features included in the digital compressed image flows (e. g. H.264 …) could help to reduce this cost.

When motion stops, the method must be complemented by detection of immobile people. When trying to detect an immobile group inside a mobile crowd, we had to discriminate immobility because of emptiness from immobility because of stopped crowd. Thus, a **long term tracking model** was designed to store information about immobile people.

3.6. Estimating people density

Then we completed this approach with a **density** model, in order to have full information on the number of moving people, their direction and speed. We studied different models in order to estimate granularity, as in Refs [16, 17, 18, 24].

The authors of [24] propose a supervised learning general framework for density estimation. They assume a set of training images with dense feature map $\phi(p) \in \mathrm{R}^m$ at each pixel p and ground-truth annotation of head positions ξ. The density functions in this approach are real-valued functions over pixels, whose integrals over image regions should match the object count. For each training image I, the ground truth density function is defined as a kernel density estimate based on the provided points:

$$\forall p \in I, \qquad F^0(p) = \sum_{P \in \xi} \mathcal{N}(p; P; \sigma^2 \mathbf{1}_{2\times 2})$$

where $\mathcal{N}(p; P; \sigma^2 \mathbf{1}_{2\times 2})$ denotes a normalized 2D Gaussian kernel evaluated at p, with the mean at the user-placed dot P, and an isotropic covariance matrix. Given this set of training images together with the ground truth density functions, the linear transformation of a given feature representation that approximates the density function at each pixel is learned:

$$\forall p \in I, \qquad F(p|w) = w^T \phi(p)$$

where $w \in \mathrm{R}^m$ is a parameter vector of the linear transform that will be obtained from the k training data by minimizing the regularized MESA distance:

$$\hat{w} = \underset{w \in \mathbf{R}^m}{\operatorname{argmin}} \quad (w^T w + \lambda \sum_k \mathcal{D}(F_k^0(.), F_k(.|w)))$$

Once trained, an estimate for object counts can be obtained at every pixel or in a given region by integrating across the area of interest.
This density model allowed us to develop flow counting functions, as well as area counting for a standing crowd.

3.7. Project results

Figure 5 : typical crowd situations studied in previous works

This led us to a demonstrator which was tested with video records and live videos. Many previous studies about crowd analysis have addressed sparse crowd, as examples shown in Figure 5. In our work, we wanted to address also dense moving crowd situations. We have tested our algorithms on videos of dense moving crowd. We provided several videos with ground truth [11, 12] that are available for further research works. Our system accurately estimates crowd density in moving crowd datasets where average density is 2.1 person/m^2 (a factor of 1.5 to 3 denser than crowd datasets often used by state of the art methods). Higher densities can be encountered in real situations (up to 7 or even 9 persons per m^2), but at this stage the move is considerably slow, no more than several meters per minute: contacts between people limit motion capacity [30].

Moreover, our system automatically detects, in real-time, a singular behavior such as a reverse walking without any assumption about shape (even if it is partly occluded and if it overlaps other crowd components). This detection operates on targets as small as a square of 10-20 pixels side (size depending on contrast). The demonstrator gives very good results on implemented functions such as counting, detecting individual reverse walk, speed measurement of crowd flows, the detection of above threshold speed or density, etc. A few examples of detection images for such situations are shown below:

Figure 6 : (a) density map, (b) man crossing the crowd flow, (c) sudden crowd acceleration

Several features of the demonstrator are shown here (Fig. 7). Angles of view w. r. t. horizon range from 43,5° at the bottom of the image to 17,3° at the top (cf. §3.1). 3D calibration provides the ground surface of the drawn area (24.1 m^2). Counting and density estimates in this area are displayed (1.0 pers/m^2). Crowd flow counting across drawn lines (on the left and on the right) is displayed. An alarm is raised about a person crossing the crowd stream (blue square). The image on the right corresponds to the direction vectors and speed estimations of the main crowd flow (purple) and of the alarming target (green), computed at a reduced scale. Ground truth density in the drawn area is 1.04 pers/m^2 (error = 4%). Speed of the main crowd stream is measured instantly at 2.10 km/h, when long term ground truth speed is rather at 2.24 km/h (error = 6.6%). Use of particle advection is under study for improving the accuracy.

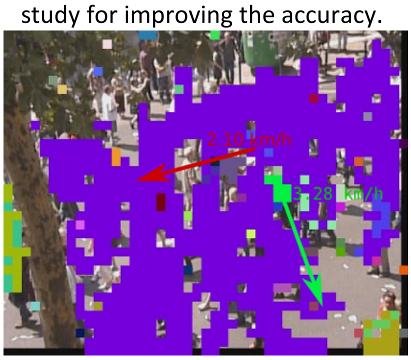

Figure 7 : (a) simultaneous density estimation, counting, and detection, (b) speed and direction

Future works will address complementing the model in order to detect smoke in a crowd, vehicle in a crowd, and other applications mentioned above in §2.3 and §2.4.

4. Perspectives

Proposing tools for detecting dangers in crowds in the public space is a very ambitious target.

As mentioned above, the functional needs for detection presented in this paper are not included so far in the calls for tenders. It will probably take some time before they are introduced.

Furthermore, because of the intrinsic nature of a crowd, each crowd situation can lead to an unlimited number of accidents. Considering all possible risks is staggering. Such a tool changes the responsibilities of the crowd surveillance operators as well as their organizations: they can no longer say that they didn't see an accident, if the system warned them of it. Their responsibility for help assistance is enhanced, and they take full responsibility in case of non-assistance in case of unmanaged alarm(s).

It is likely that once reluctance to acceptance is overcome, such systems will probably first enhance the *Feedback* step (cf. §1) of peaceful crowd management, then the *Preparation* step, and will ultimately be complemented by evolutions of crowd surveillance operators work organization, in order to adapt the *Action* steps.

We are confident that global security shall be improved.

5. About

We would like to thank here Jacques Blanc-Talon and Veronique Serfaty of DGA who supported our work, as well as the contributors of the SNCF and the Prefecture de Police de Paris, for their contribution to the specifications. We also thank the Willow research team for its participation and research work, and Y-O Renault for his help with the math.

The French Willow research team, is a joint team of CNRS, INRIA, and ENS, including about 40 people, led by Jean Ponce.

EVITECH is a SME member of the **SYSTEM@TIC Paris Region** cluster. It studies and develops innovative image processing systems for global security. P. Bernas is Ingénieur Civil des Mines and PhD in computing science. P. Drabczuk is graduate of the Optics Institute Graduate school. G. Née is currently completing his PhD in image processing at the GREYC / University of Caen.

6. Bibliography

[1] European communities, *Meeting the challenge, the European Security Research Agenda, a report from the security research advisory board*, sept. 2006.

[2] RTCA DO-178B, *Software Considerations in Airborne Systems and Equipment Certification,* RTCA Inc., Washington D.C, 1992 / ED 12B, EUROCAE, Paris, 1992.

[3] Ben Benfold, Ian Reid, Stable, *Multi-Target Tracking in Real-Time Surveillance Video*, CVPR 2011

[4] Tao Zhao, Ram Nevatia, *Stochastic Human Segmentation from a Static Camera*, CVPR 2004

[5] Junliang Xing, Haizhou Ai and Shihong Lao , *Multi-Object Tracking through Occlusions by Local Tracklets Filtering and Global Tracklets Association with Detection Responses* , CVPR 2009

[6] Chang Huang, Bo Wu, and Ramakant Nevatia , *Robust Object Tracking by Hierarchical Association of Detection Responses* , ECCV 2008

[7] Vivek Kumar Singh, Bo Wu, Ramakant Nevatia , *Pedestrian Tracking by Associating Tracklets using Detection Residuals* , WMVC 2008

[8] Bo Wu and Ram Nevatia , *Detection and Tracking of Multiple, Partially Occluded Humans by Bayesian* , IJCV 2007

[9] Bo Wu, Ram Nevatia and Yuan Li, *Segmentation of multiple, partially occluded objects by grouping, merging, assigning part detection responses*, CVPR 2008

[10] Zhao T. and Nevatia R., *Tracking multiple humans in crowded environment*, CVPR 2004

[11] M. Rodriguez, I. Laptev, J. Sivic and J.Y. Audibert, *Density-aware person detection and tracking in crowds*, ICCV 2011

[12] M. Rodriguez, J. Sivic, I. Laptev and J.Y. Audibert, *Data driven crowd analysis in videos*, ICCV 2011

[13] Saad Ali and Mubarak Shah, *A Lagrangian Particle Dynamics Approach for Crowd Flow Segmentation and Stability Analysis*, CVPR 2007

[14] Min Hu, Saad Ali and Mubarak Shah , *Learning Motion Patterns in Crowded Scenes Using Motion Flow Field*, International conference on pattern recognition 2008

[15] Mehdi Moussaıd, Niriaska Perozo, Simon Garnier, Dirk Helbing, Guy Theraulaz, *The Walking Behaviour of Pedestrian Social Groups and Its Impact on Crowd Dynamics*, PLoS ONE 2010

[16] Siu-Yeung Cho, T.W.S. Chow and Chi-Tat Leung, *A neural-based crowd estimation by hybrid global learning algorithm*, IEEE Transactions on Systems, man, and cybernetics 2002

[17] A.B. Chan, Z.S.J. Liang and N. Vasconcelos, *Privacy preserving crowd monitoring: counting people without people models or tracking*, CVPR 2008

[18] D. Ryan, S. Denman, C. Fookes and S. Shridharan, *Crowd counting using multiple local features*, Digital image computing 2009

[19] Ovgu Ozturk, Toshihiko Yamasaki, Kiyoharu Aizawa, *Detecting Dominant Motion Flows In Unstructured/structured Crowd Scenes*, , ICPR 2010

[20] T. Bouwmans, F. E. Baf, and B. Vachon. *Statistical background modeling for foreground detection: A survey*. Handbook of Pattern Recognition and Computer Vision World Scientific Publishing, 4:181–199, Jan. 2010.

[21] Anil M. Cheriyadat and Richard J. Radke, *Automatically determining dominant motions in crowded scenes by clustering partial feature trajectories*, International Conference on Distributed Smart Cameras 2007

[22] Sergio A. Velastin, Boghos A. Boghossian b, Maria Alicia Vicencio-Silva, *A motion-based image processing system for detecting potentially dangerous situations in underground railway stations*, Transportation Research Part C: Emerging Technologies 2006

[23] P. Felzenszwalb, R. Girshick, D. McAllester, and D. Ramanan, *Object detection with discriminatively trained part based models*, IEEE PAMI 2010.

[24] V. Lempitsky and A. Zisserman. *Learning to count objects in images*, NIPS 2010.

[25] WorkSafe Victoria (Australia), *Crowd control at venues and events, a guide to support and assist crowd control agencies*, available at http://www.worksafe.vic.gov.au/

[26] B.K.P. Horn and B.G. Schunck, *Determining optical flow,* Artificial Intelligence 1981

[27] B. D. Lucas and T. Kanade, *An iterative image registration technique with an application to stereo vision,* Imaging Understanding Workshop 1981

[28] J. Shi and C. Tomasi, *Good Features to Track,* CVPR, 1994

[29] C. Tomasi and T. Kanade, *Detection and Tracking of Point Features,* Pattern Recognition 2004

[30] Dirk Helbing, Anders Johansson, HE Habib Z. Al-Abideen, *Crowd turbulence: the physics of crowd disasters,* 5th International conf. on non-linear mechanics (ICNM-V), Shanghai, June 2007.