

DOI:10.1145/2043174.2043192

**People in high-density crowds appear to move with the flow of the crowd, like particles in a liquid.**

**BY BRIAN E. MOORE, SAAD ALI, RAMIN MEHRAN, AND MUBARAK SHAH**

# Visual Crowd Surveillance through a Hydrodynamics Lens

VIDEO CAMERAS MONITORING the activity of people in public settings are commonplace in cities worldwide. At large events, where crowds of hundreds or thousands gather, such monitoring is important for safety and security purposes but is also extremely (technically) challenging. Human operators are generally employed for the task, but even the most vigilant humans miss important information that could ultimately contribute to unfavorable consequences.

Major research efforts are under way to develop systems that cue security personnel to individuals or events of interest in crowded scenes. Essential are methods by which information can be extracted from video data in order to recognize crowd behaviors, track

individuals in crowds, and detect abnormal events.

This article explores cutting-edge techniques we have used in real-world scenarios to provide solutions to such problems.<sup>1,2,17</sup> We developed them based on the notion that people in crowds behave, in ways, like particles in fluids. Hence, we treat crowds as collections of mutually interacting particles.

Typically, the motion of a high-density crowd appears to behave like a liquid, and interaction forces tend to dominate the motion of the people. This is in contrast to crowd motion appearing in states like gases, where interactions between people are few but random motions of individuals tend to dominate the behavior. With all this in mind, we contemplate visual crowd surveillance using ideas and techniques based in hydrodynamics. Hence, we say “fluid” and “liquid” interchangeably, distinguishing our approach from aerodynamics, which considers fluids in gaseous states.

Our hydrodynamics point of view is well suited for analyzing high-density crowds,<sup>9,12</sup> with surveillance the primary concern. Though the number of people will never reach the astronomical numbers of particles in fluids, we pursue tasks in crowd analysis using a similar concept of scale. Ranging from the macroscopic view of all particles to the microscopic view of individual

## » key insights

- **Computer algorithms extract information from digital videos of people in crowds as a way to automatically track individuals, detect abnormal behavior, and segment characteristic patterns of flow in crowds.**
- **Individuals in dense crowds, like particles in a fluid, are restricted in their motion by neighboring individuals, reflecting a kind of interdependence that is pivotal for solution development.**
- **The tools of computational and applied mathematics are indispensable for visual analysis of crowds; pixel information is translated into particle trajectories used to understand crowd flow on length scales ranging from the macroscopic to the microscopic.**



Figure 1. (a) New York City Marathon; (b) political rally in Los Angeles; (c) pilgrims circling Kabba in Mecca.

particles, we address the problems of segmentation, abnormal-behavior detection, and tracking.

Techniques devised by other researchers have been used to consider similar problems. Here, we briefly review research in the area of visual crowd surveillance, referring the reader to detailed articles on tracking<sup>23</sup> and crowd-behavior analysis.<sup>24</sup>

For problems involving crowd segmentation, Chan and Vasconcelos<sup>8</sup> used dynamic texture-based representations of scenes to determine how regions differ, proposing a method<sup>7</sup> for counting pedestrians in high-density crowds. Sand et al.<sup>20</sup> implemented a particle-based framework for the purpose of estimating the motion in a scene but did not use it for interpretation of significant segments.

For problems involving behavior analysis, methodologies are available for understanding crowd behavior. The first, advanced by Marques et al.<sup>16</sup> and Tu et al.,<sup>22</sup> perceived a crowd as an assembly of individuals, using segmentation or tracking algorithms to understand their behavior. The other, promoted by Andrade et al.,<sup>3</sup> viewed a crowd as an organism, such that its behavior is studied and accepted on a global level. Reisman et al.<sup>19</sup> proposed that crowd behavior is recognized by modeling the scene, giving a description of important features within it. Kratz et al.<sup>15</sup> detected anomalies as statistical deviations from the ordinary motion patterns in space-time volumes to characterize the scene.

With regard to tracking in crowded scenes, one of the first important methods was devised by Zhao et al.<sup>25</sup> using ellipsoids to model the human shape and color differences to mark appearances. Another framework, by Brostow et al.,<sup>5</sup> assumed that points appearing to move together are prob-

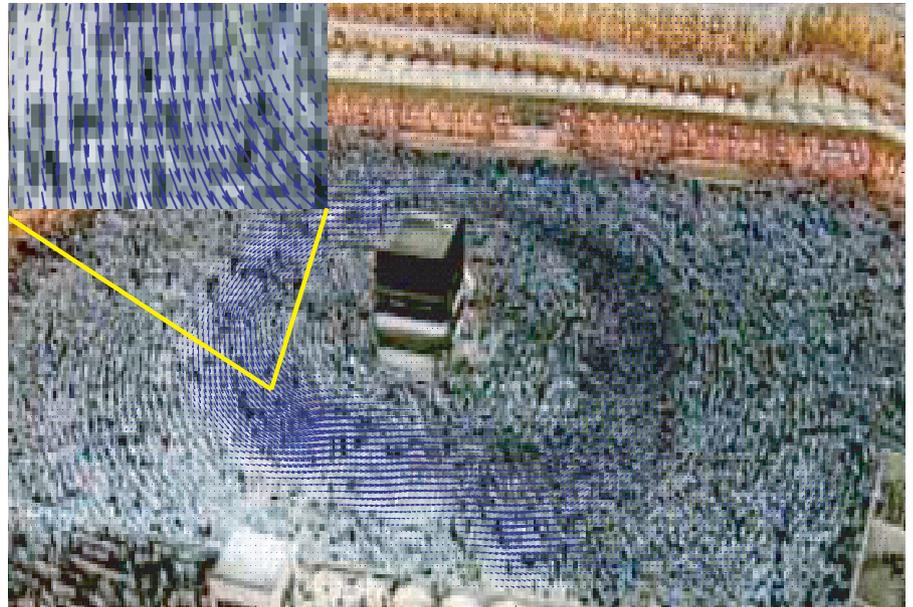


Figure 2. Flow field of a frame in the Kabba video.

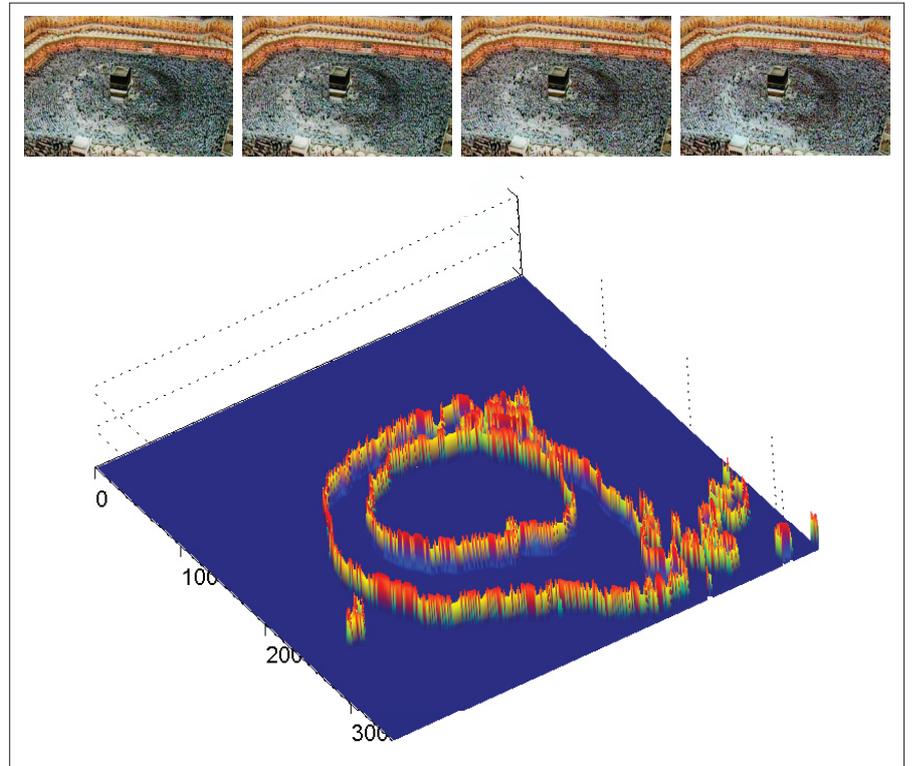


Figure 3. Four frames from the video sequence of pilgrims circling Kabba and the FTLE field.

ably part of the same object, tracking individuals based on the probability that points could be clustered together. More recently, Pellegrini et al.<sup>18</sup> expanded a social-force model to take into account destinations and desired directions of individuals, making it

suitable for tracking individuals in crowded scenes.

### Particles and People

Random actions, relationships between energy and density, and a gas/liquid/solid-state demeanor are all

characteristics of particles in a fluid and of people in a crowd. Most important, the motions of particles/people are determined by the external forces exerted on them; for example, both particles and people are affected by boundary forces (such as walls) and feel the forces of neighboring particles/people. One difference is that people are, to some extent, able to determine their own destiny, so the crowd may be viewed as a “thinking fluid,”<sup>11</sup> but there are still probabilistic similarities to particle motion regardless of this difference.

When scientists consider hydrodynamics, they often use different scales, depending on the questions being addressed.<sup>14</sup> At the microscopic, one may examine the position or velocity of a particular particle among many. On another, the macroscopic is used to scrutinize the nature of enormous collections of particles (such as a tree branch moving in water). Between them is the mesoscopic scale, which is used to analyze the interaction of “small” collections of particles, giving characteristic information (such as temperature and average density).

Considering the behavior of people in a crowd, we take a similar approach, depending on the questions we want to answer. We focus on three generally recognized key problems in visual crowd surveillance—crowd segmentation, behavior analysis, and tracking—corresponding to the scales. To be clear, some situations might necessitate tracking a particular person in a crowd, requiring a microscopic point of view. Others might call for descriptive information on when the behavior of a crowd is abnormal, meaning it is neither necessary nor feasible to track every individual in a crowd but important to understand how groups of individuals interact, for which we employ a mesoscopic point of view. Still, a macroscopic point of view is more appropriate for segmenting global patterns of flow.

Here, it is pertinent to discuss the types of scenes and spatio-temporal range of crowd behaviors that can be handled through an understanding of the hydrodynamics point of view. To begin, hydrodynamics-based techniques require that a crowd be viewed from above, thereby minimizing artifacts resulting from independent

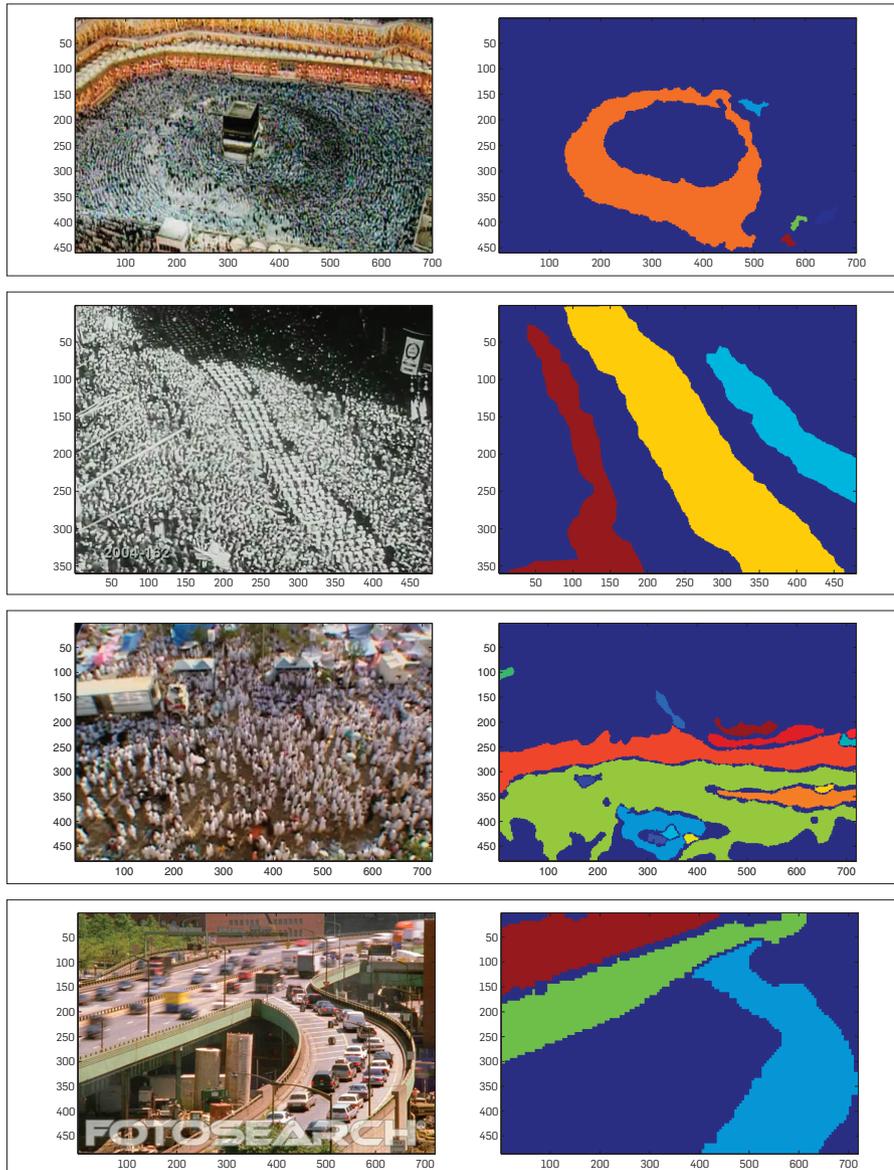


Figure 4. Video scene (left) and corresponding segmentation (right).

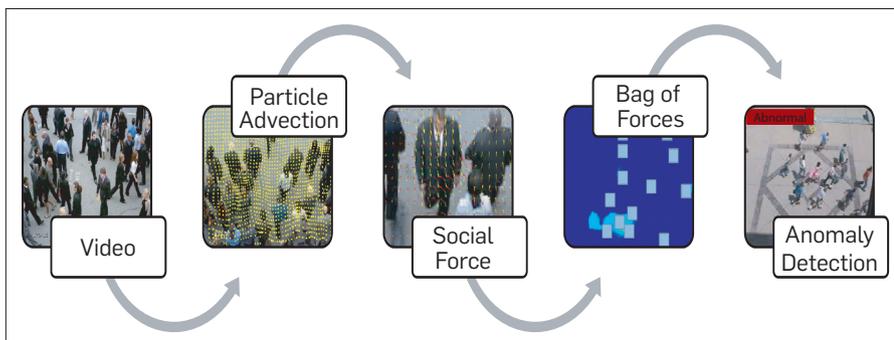


Figure 5. Our approach for detecting abnormal behavior in crowd videos.

movement of multiple body parts. Side views of the scene are least preferable within the particle-based framework; Figure 1 includes examples of such scenes and camera setup. Our algorithms next allow each pixel to represent a particle, with a minimum requirement on the spatial scale of at least one pixel per person. If two or more people are matched to a single particle, the methods may encounter problems, but allowing as many particles per person as the scene dictates is certainly permissible.

Another noteworthy requirement is that video scenes exhibit a dominant trend typical of high-density crowds, where the movement of individuals is restricted by other individuals, obligating the group to move as a whole, like a fluid. However, the dominant trend is key to the analysis, while the density of the crowd is allowed to vary. Since crowd behavior is naturally dynamic, and the flow (trend) of a crowd can change with time, any video-based analysis of crowd motion should take a sliding-window approach. It performs the analysis over a particular sequence of frames (a window in time), then “slides” the window to another sequence of frames to repeat the analysis. The size of the window may be adaptable or fixed but depends on the level of activity in the scene. Methods explored here follow this principal.

*Macroscopic scale (crowd segmentation).* The macroscopic scale suggests a focus on global crowd behavior, requiring a comprehensive point of view; Figure 1 includes examples of the types of scenes we consider, with thousands of people in view. In such settings, we are primarily interested in the overall movement of the crowd, meaning we are able to find segments of common motion within it.<sup>1</sup>

A key ingredient for our solution to the problem of segmentation is called “particle advection” and used in each of our three problem synopses. The approach itself mimics a common mathematical formulation of fluid mechanics, or Lagrangian specification, characterized by following particular particles as they move with the flow.<sup>4</sup> The first step in applying this idea to a video sequence is to compute the optical flow, or apparent visual motion of objects in the scene



## Random actions, relationships between energy and density, and a gas/liquid/solid-state demeanor are all characteristics of particles in fluids and of people in a crowd.



(see Figure 2). Every pixel has position  $\mathbf{x} = (x, y)$ , and the optical flow provides velocities  $(u, v)$  at each position, so objects are related to their velocities by the system of equations

$$\frac{dx}{dt} = u(x, y, t), \quad \frac{dy}{dt} = v(x, y, t)$$

Particle advection is performed by overlaying the scene with a grid of particles that serves as the initial conditions for this system of equations; particles are then transported to new coordinates in subsequent frames using a time-stepping technique for integrating the system of equations, as in

$$\begin{aligned} x(t+1) &= x(t) + u(x(t), y(t), t), \\ y(t+1) &= y(t) + v(x(t), y(t), t) \end{aligned}$$

Thus, the flow of the crowd in the scene is given by particle trajectories.

Important to note is that errors and noise in optical flow are averaged out to some extent as a result of time integration performed to determine particle trajectories. Thus, the particular method used to produce optical flow is not crucial for the three problems we consider and has been verified experimentally. Temporal scale for analysis is determined by the integration time  $t$ . In practice,  $t$  should depend on the rate of change of the flow field, with a higher rate of change of flow field resulting in smaller time scales and vice versa. In our experiments, we fixed  $t =$  two seconds or 60 frames for all scenarios.

Particle advection produces a flow map, a function  $\phi_{t_0}^t(\mathbf{x}_0) = \mathbf{x}(t; t_0; \mathbf{x}_0)$  relating the position  $\mathbf{x}$  of a particle at time  $t_0 + t$  to its original position  $\mathbf{x}_0$  at the initial time  $t_0$ . That is, the flow map fully describes the trajectory of each particle, which does not necessarily correspond to a person in a crowd but to a small region in the scene exhibiting a collective pattern of motion. In sections with coherent motion, the flow maps show qualitatively similar behavior, but trajectories experiencing different behavior are from sections with different coherent motion. These qualitative differences define flow segments. Our primary mathematical tool for finding these qualitative differences is called a “finite time Lyapunov exponent,” or FTLE, we use to define Lagrangian coherent structures.<sup>21</sup> The FTLE is essen-

tially a number that reflects how two neighboring particles separate from one another over time and is computed using the maximum eigenvalue  $\lambda_{max}$  of the Cauchy-Green deformation tensor  $\Delta$ , obtained from the Jacobian matrix for the flow map,  $D\phi_t^t(\mathbf{x})$ . More precisely, the largest FTLE with integration time  $t$  is

$$\sigma = \frac{1}{t} \ln \sqrt{\lambda_{max}(\Delta)}$$

where

$$\Delta = D\phi_{t_0}^{t_0+t}(\mathbf{x})^T D\phi_{t_0}^{t_0+t}(\mathbf{x})$$

Computing the FTLE at every point produces the FTLE field, a scalar field that immediately exposes any regions in the scene with differing flow by finding particle trajectories that start close together but end far apart. In practice, the particle advection approach allows implementation of the algorithm in both forward and backward time, meaning the flow segments are the same regardless of which direction the flow is moving.

Combining the FTLE fields for both forward and backward motion yields vivid results (see Figure 3). We use a watershed algorithm to segment the FTLE field, making it find the exact number of flow segments. This process is repeated by moving the sliding temporal window to obtain segments for subsequent time steps.

The end result is a net segmentation showing each region exhibiting a single clearly defined characteristic flow pattern. Such a result is not possible through segmentation based solely on optical flow, because optical flow captures only motion between two frames. On the other hand, particle advection motion in several frames is integrated over time and nicely captured by the scalar FTLE field. Figure 4 includes several results in which the motion in crowded pedestrian and traffic scenes is properly segmented through our method; each row shows a frame from a different video sequence, along with subsequent segmentation. Regions of different colors signify qualitative changes in the flow,

and dark blue represents areas with no coherent flow. A clear example is the traffic scene in the last row, with dark blue representing regions outside the lanes, and red, green and light blue representing movement in each direction.

*Mesoscopic scale (behavior detection).* Beyond a global understanding of pedestrian flow in crowds, detection of abnormal events or behavior is important, generally for the sake of public safety. We use the local interactions of multiple people to identify regular patterns of motion, in addition to any anomalies.<sup>17</sup> A fundamental component of our approach (see Figure 5) in this setting is a social force (fluids-based mathematical) model for describing pedestrian movement, as pioneered by Helbing and Molnar<sup>10</sup> almost 20 years ago.

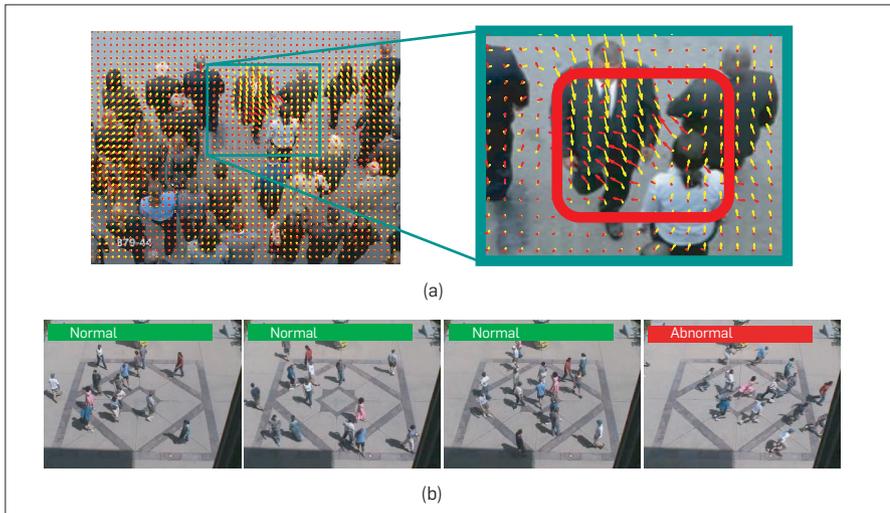
The central idea hinges on Newton’s second law of motion—force equals mass times acceleration, or  $F = ma$ . In it, each individual in the scene reacts to forces that produce motion. These forces can be deconstructed into two parts: the personal-desire force (individuals striving to get to their desired destinations) and the interaction force (exerted on individuals by other individuals or things in the scene). Thus, pedestrian  $i$  changes velocity according to

$$a = \frac{dv_i}{dt} = F_p + F_{int}$$

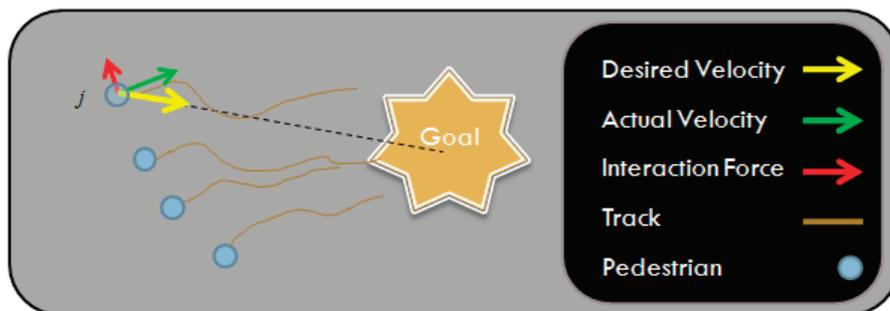
where  $F_p$  and  $F_{int}$  refer to personal and interaction forces, respectively. In a given scene, since individuals are all relatively the same size, the masses are assumed to be one. Quantifying these forces (see Figure 6a for an example) allows our method to establish the ongoing behavior in the crowd, enabling detection of any behavior out of the ordinary (Figure 6b).

Note that in very dense crowds, pedestrians follow group velocity and goals,<sup>12</sup> but as density decreases, personal interest plays a greater role in pedestrian motion. Hence, at the mesoscopic scale, our algorithm may use scenes with mid-to-high crowd density, provided the interaction force is not negligible, meaning behavior is still fluid-like.

The algorithm itself starts with particle advection, followed by computation of the forces. Each person in



**Figure 6. (a) Optical flow (yellow) and computed interaction vectors (red) for pedestrians with opposing directions; (b) frames of a sequence where the observed behavior suddenly becomes abnormal (people running in panic) in the last frame.**



**Figure 7. Scheme for computing interaction force.**

the crowd has a desired direction and speed, but individual direction and speed is limited by the surrounding pedestrians. The actual velocity  $v_i$  of a particle in the  $(x_i, y_i)$  coordinate is obtained from the spatial-temporal average of optical flow. On the other hand, the desired velocity  $v_i^p$  is given by the optical flow for that particle. Hence, the personal-desire force is

$$F_p = \frac{1}{\tau} (v_i^p - v_i)$$

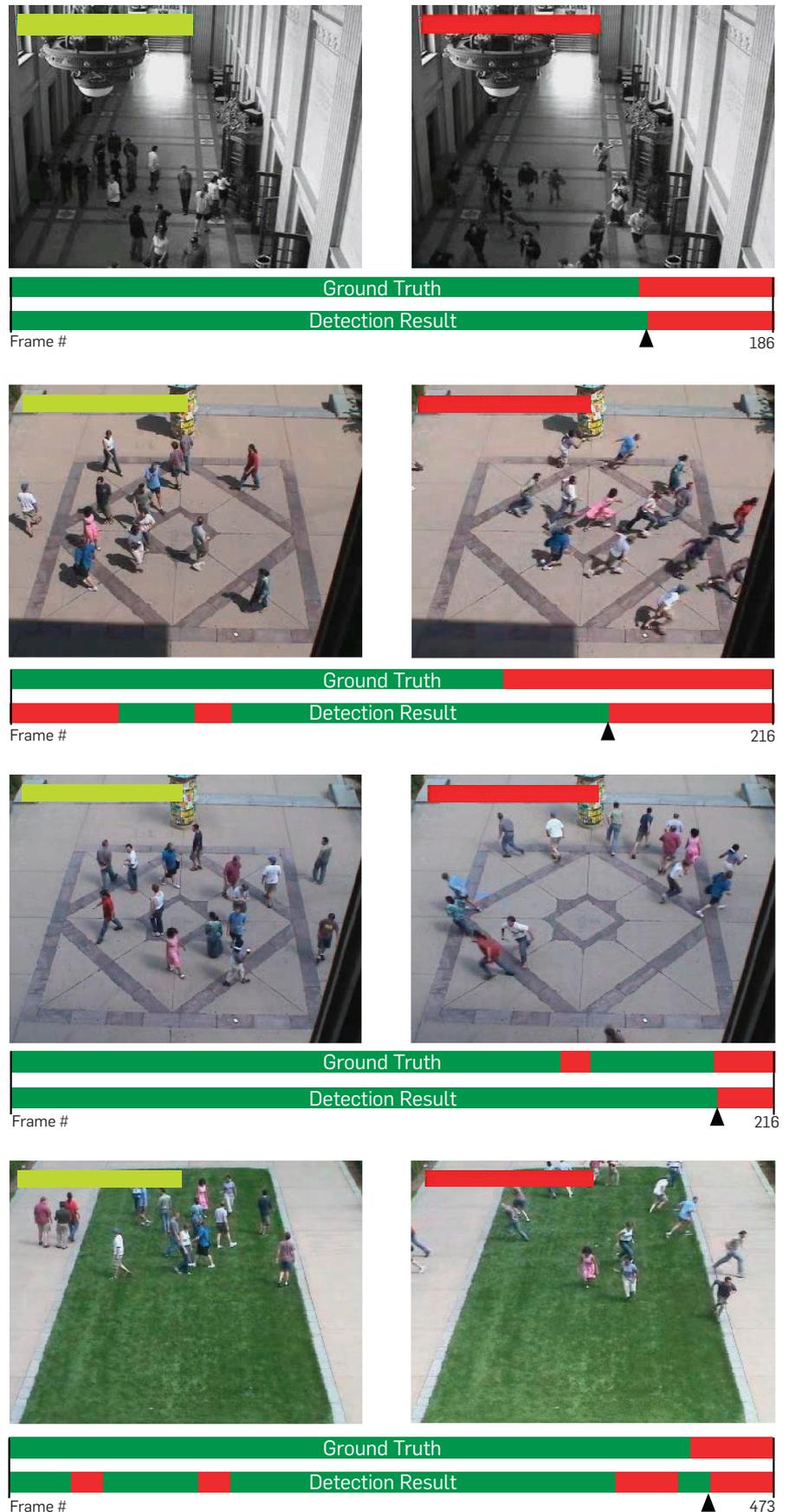
where  $\tau$  is a relaxation parameter. Thus, the interaction force (see Figure 7) is given as

$$F_{int} = \frac{dv_i}{dt} - F_p$$

These forces together yield a sufficient description of the motion in the scene based on the acting forces.

Specification of the forces determining the motion in the scene provides understanding of synergy between interacting particles but does not, by itself, secure evidence of changes in behavior; for example, normal interaction forces on a stock-market trading floor may differ drastically from those of pedestrians on the street. Using this technique to detect and localize any changes in behavior, the computer must first learn the “normal” behavior for the scene, for which our algorithm takes a bag-of-words approach. (In the same way a document can be considered a bag of words, a video can be considered a bag, or collection, of spatial-temporal cuboids, for which the interaction force is computed.) The idea is for the algorithm to use a training set of videos and match the interaction forces with given dynamics. A video in question can then be compared with those from the training set, and changes from the regular behavior in the scene are easily identified by the computer.

To improve the fidelity of the results, optical flow is smoothed by a Gaussian filter, where the standard deviation of the Gaussian distribution is empirically set to half the width of the typical person in the crowd. This smoothing compensates for the inaccuracies of optical flow in textureless regions. Moreover, using a bag of video words for several frames could also reduce the effects of inaccurate instantaneous optical flow.



**Figure 8. Frames from different sequences, showing (left) normal behavior (green) and (right) abnormal escape panic (red), comparing ground truth to abnormal behavior detection.**

Sample results of the algorithm are in Figure 8; the videos for these experiments are from the University of Minnesota and show walking pedestrians as the normal behavior. At the end of each video the pedestrians suddenly run in all directions to escape the scene. The figure shows detection of abnormal behavior by our method (indicated as black triangles) compared with the ground truth. In most cases, panic detection occurs immediately following the change in behavior. The receiver operating characteristic curves in Figure 9 show a clear advantage of our method over simply using the optical flow to detect abnormal behaviors.

*Microscopic scale (individual track-*

*ing).* At the “atomic” level, a surveillance analyst is interested in automatically following a person in a high-density crowd, a very challenging problem, as the object our algorithm is tracking is subject to occlusion, and other nearby objects may lead the tracker away from the original object. Figure 10 shows tracking results using our method in which individuals are correctly tracked in four video sequences involving hundreds of people; each image shows the tracks overlaying a single frame of the video.

Inspired by research on evacuation dynamics,<sup>6,13</sup> our method uses a scene-structure-based force model that likens pedestrians to particles, such that the forces acting on them determine

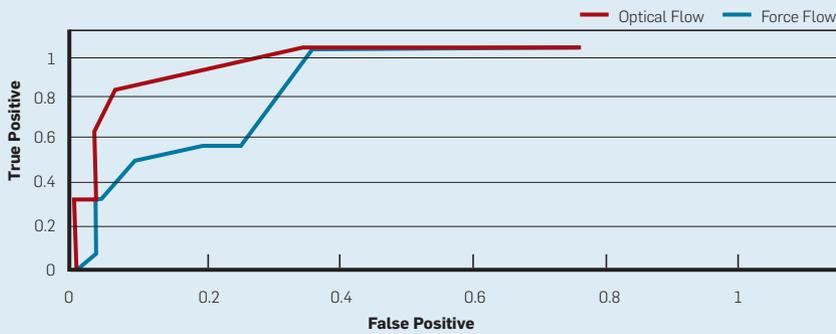
their direction and velocity. The algorithm computes the probability that a particular particle will move from one position to another, building on floor fields that provide information about the scene.<sup>2</sup> To make this clear, we make three assumptions about the flow influencing the individual’s behavior: First, the person has a goal (place to get to and clear direction how to get there) and, in the absence of obstacles, will go there directly; this is the effect of what is called the “static floor field.” Second, the person avoids permanent fixtures (such as trash cans and walls) and virtual barriers (such as opposing crowd flow) as a consequence of what is called the “boundary floor field.” And third, the person can move toward the goal only as the flow of the crowd allows; this motion and direction is the influence from the dynamic floor field.

A basic assumption on the static floor field, based on the observation that directions of motion in high-density crowds have dominant trends, is that crowd behavior remains constant during tracking. However, the static floor field can be updated periodically to respond to changes in the dominant trends. To respond to any instantaneous change in crowd flow, the model uses the dynamic floor, which is representative of instantaneous crowd behavior in the vicinity of the target. The main limitation of the floor-field tracking model is the inability to handle locations with no dominant trend (such as a crowded museum) and locations with more than one dominant trend (such as pedestrian crossings).

We begin our description of the method with the inference that people in crowds are constantly avoiding collisions. Hence, the boundary floor field is repulsive and computed easily through particle advection and the FTLE field, as described earlier in terms of segmentation of crowd flow. The edges of the computed segments give the boundaries of the flow, leading to the resulting boundary floor field (see Figure 11).

Computation of the static floor field (Figure 11d) is performed only once for a given video using a small subset of all video frames. The first step provides a representation of the instantaneous changes in motion, or “point flow,” achieved by calculating

**Figure 9. ROC curves for detection of abnormal behaviors in the University of Minnesota data set; the area under the social force curve (red) is 0.96, and the area under the optical flow curve (blue) is 0.84.**



**Figure 10. Tracking individuals using our method in (a–c) marathon scenes and (d) a crowded train station.**

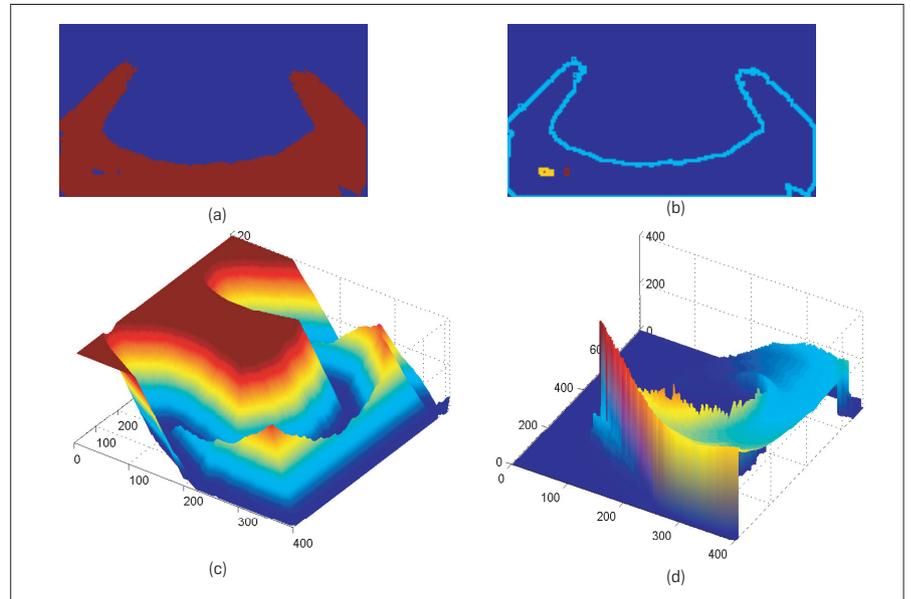
the average optical flow for each location over the entire subset of frames. Our algorithm can then place a grid of particles over the scene and determine the preferred direction of each particle based on the motion of neighboring point-flow vectors. If the influence is great enough to move the particle to the next cell, then the algorithm continues the process until the velocities are not significant enough to move it to the next position (see Figure 12). This process is used by the algorithm to find the sinks, defined computationally as the points where particle motion ceases to exist. In terms of crowd behavior, sinks are the desired goals or locations of the individuals in the crowd (such as preferred exits and frequently visited areas dominated by the flow of the crowd). The sinks, as well as the shortest distance needed to reach them, produce the static floor field.

Computing the dynamic floor field means discovering the behavior of the crowd around the individual. To do this, the algorithm uses the optical flow for a subset of video frames and performs particle advection. If a particle changes its position between frames, then the value of interaction between those frames is increased by one, and zero interaction is assumed at the first frame in this sequence. The individual's interactions in a local neighborhood are thus captured for that interval of time (see Figure 13).

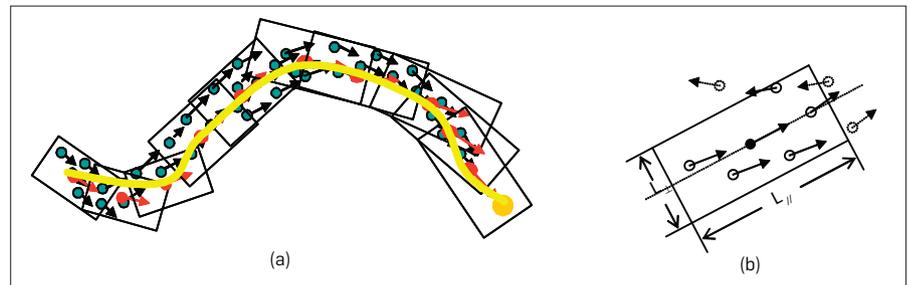
To bring the three floor fields together for the purpose of tracking, the algorithm divides image space into cells, so each cell is occupied by one particle. The probability that a particle at cell  $i$  will move to neighboring cell  $j$  is then computed and combined with appearance information to complete the tracking. This method depends on computation of the influences from the static, dynamic, and boundary floor fields, denoted  $S_{ij}$ ,  $B_{ij}$ , and  $D_{ij}$ , respectively, with each needed for accurately modeling the interaction of individuals and their preferred direction of motion. Described precisely, the probability that a particle will move from  $i$  to  $j$  is

$$p_{ij} = C e^k D_{ij} e^{k_s S_{ij}} e^{k_b B_{ij}} R_{ij}$$

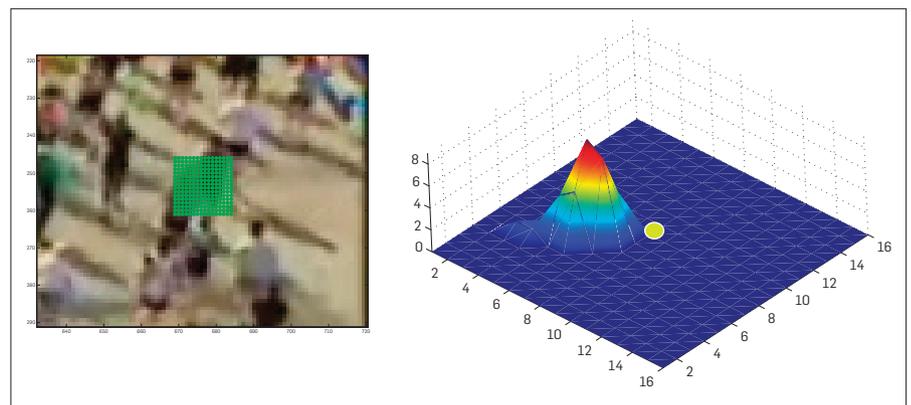
where  $k_d$ ,  $k_s$ , and  $k_b$  are the coupling



**Figure 11.** For the marathon sequence in Figure 10c: (a) crowd-flow segmentation obtained through particle advection; (b) the corresponding edge map; (c) the boundary floor field; and (d) the static floor field.



**Figure 12.** (a) The sink-seeking process. Red arrows signify the point flow influenced by neighboring points; the yellow curve is the sink path. (b) The sliding window used to find sinks; the solid circle is the point under consideration; hollow circles inside the box are neighboring points and outside the box non-neighboring points.



**Figure 13.** (left) Region for computing dynamic floor field, where green chip is a target individual; (right) dynamic floor field reflects strong relationship between the yellow cell and neighboring cells at the peak.

strength of the object to the respective field,  $C$  is a normalization constant, and  $R_{ij}$  is a similarity measure for the initial and updated appearance templates.

Experimentally, a target individual is selected by a surveillance analyst

by computing the gray-scale appearance template for a rectangular region, called a chip, surrounding the individual, with average chip size  $14 \times 22$  pixels. The algorithm computes the position of the target at the next time instant according to the probable loca-

tion, as determined by the computed floor fields; the appearance similarity is then computed by normalized cross correlation, and the appearance template is automatically updated. Figure 14 charts results for 50 tracks in a video of a marathon, showing objects are tracked correctly in most cases.

Some tracking methods (though not ours) depend mainly on appearance information, but in crowded scenes appearance is not enough, as neighboring objects may have similar appearance. Figure 15a shows the appearance similarity surface for a marathon scene; the surface is relatively flat, so which runner is being tracked from

frame to frame is uncertain. However, by combining the dynamic and static floor fields (Figures 15b and c) with the appearance surface our method obtains a surface (Figure 15d) providing the best match for the tracked individual. Figure 16 also shows that, when using all three floor fields, the tracking error is consistently low, but if using only one floor field, the error increases, often significantly.

**Conclusion**

We have devised methods for segmenting motion, detecting abnormal behavior, and tracking individuals in video scenes of high-density crowds.

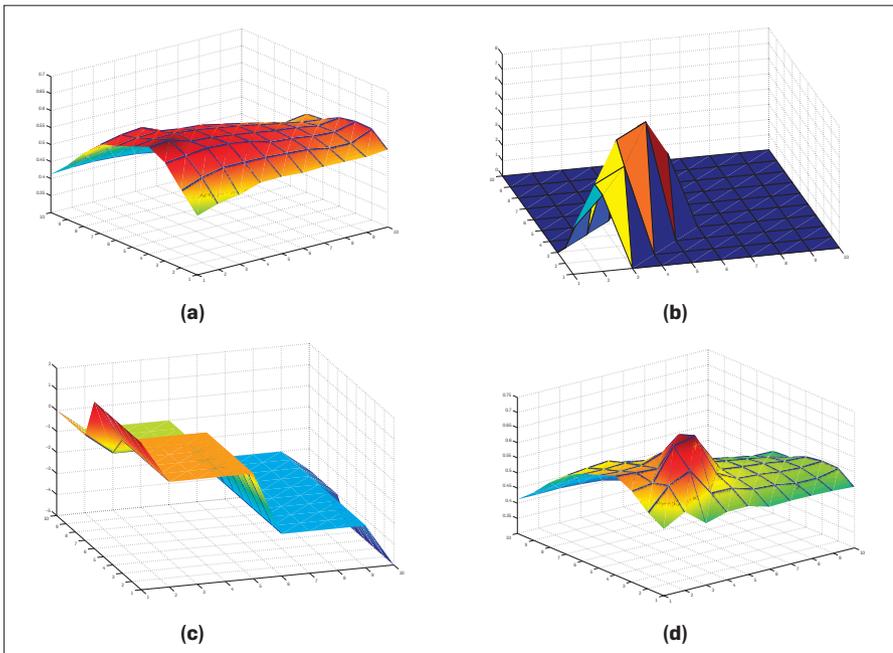
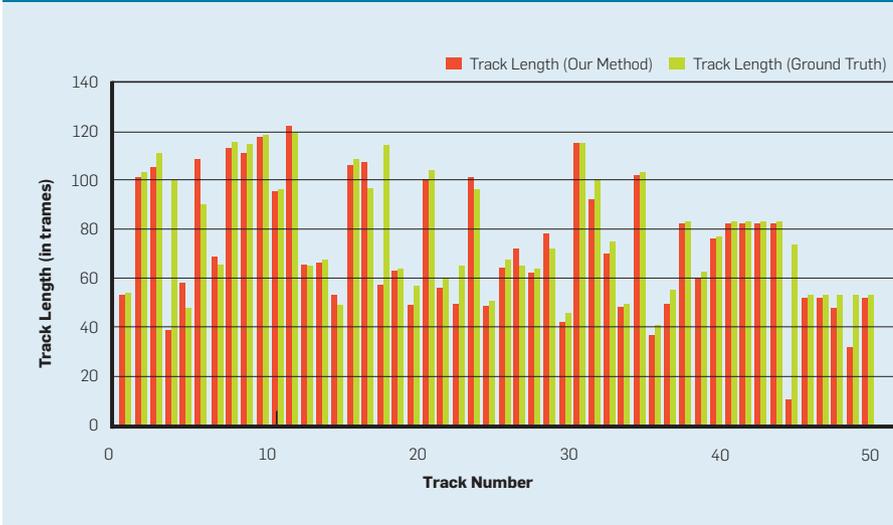
Our underlying supposition is that people in crowds appear to move according to the flow, like particles in a liquid. Hence, we gaze through a hydrodynamics lens to analyze video scenes in various scenarios on three different length scales. Each of our methods relies to some extent on the optical flow and associated particle advection adapted from the Lagrangian approach to fluid dynamics.

Our experimental results have been excellent, and we expect the underlying hydrodynamics theme can be taken further to solve other problems in visual surveillance of high-density crowds. Ultimately, we envision the ability to predict potentially hazardous situations in crowded scenes, though it is work for the future. Training a computer to decipher and understand crowd behavior from a video sequence is extremely challenging; aside from having to sort through a plethora of digital information, there are also questions specific to each of the three problems—segmenting motion, detecting abnormal behavior, tracking individuals—discussed here.

For crowd segmentation, our method makes use of flow maps corresponding to each particle, computing maximal Lyapunov exponents to reveal segments of coherent motion in the scene.<sup>1</sup> Our method performs well for steady flows, with no changes in geometry, but segmentation for unsteady flows is an open problem with several challenges. Coherent flow segments in crowds can change quickly, and to capture such changes, an algorithm must distinguish changes within segments from changes in segment boundaries. One location in a scene may also exhibit alternating collective patterns of motion, meaning several segmentations are needed to describe different modes within a single region. In addition, modeling abstract human behaviors that help define segments (such as courteous acts, social agreement, and individual intention) is difficult. Moreover, scenes can grow more complex, as moving/cluttered back/foregrounds are important in segmentations more discriminating than ours.

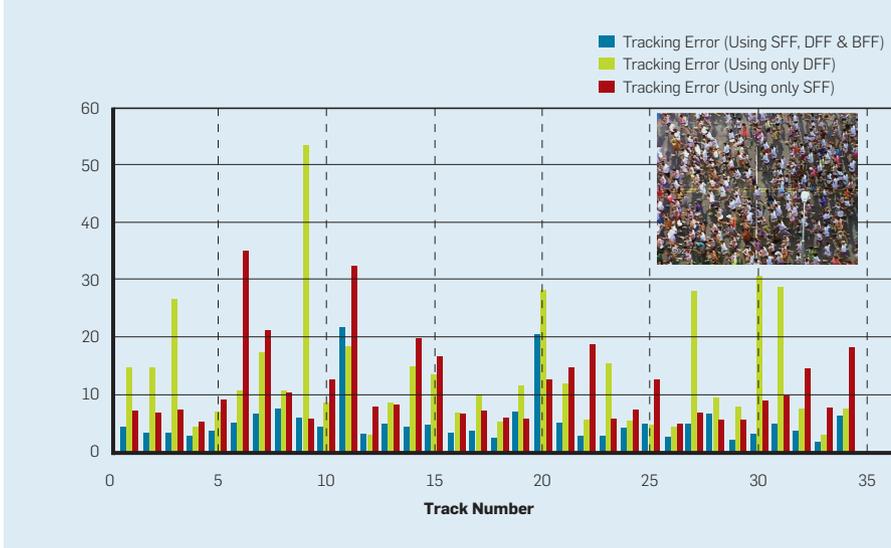
For detecting abnormal behavior, our method approximates the interaction forces in the crowd to build a model for the motion, detecting anomalies as deviations from the norm.<sup>17</sup>

**Figure 14. Computed track lengths vs. ground truth for a marathon sequence.**



**Figure 15. For tracking a runner in a marathon sequence: (a) appearance similarity surface; (b) dynamic floor field; (c) static floor field; and (d) final decision surface.**

**Figure 16. Average tracking error for each object in a marathon sequence using only dynamic (green), only static (maroon), and all three (blue) floor fields.**



This approach works well for detecting global changes in regular motion, but detecting smaller (more local) events is more difficult. Our method is also, by design, good at measuring the forces individuals exert on one another but is unable to recognize specific behaviors and distinguish the acceptable from the unacceptable. This limitation stems from an enormous variety in the behaviors observed in crowded scenes, along with the difficulty of distinguishing certain activities from other activities. Some behaviors are easily defined (such as bottlenecks or lanes), but formulating clearly defined behaviors for general crowd motion is difficult, as is categorizing unsteady flows, with the flow constantly changing.

For tracking in high-density crowds, our method exploits the influences of boundaries, neighboring pedestrians, and desired direction, along with appearance information, to identify the position of a target in subsequent frames.<sup>2</sup> Our algorithm produces excellent results for extremely crowded scenes, where the tracked individual is highly influenced by the flow of the crowd, but tracking in crowds that are less dense, allowing pedestrians to move against the flow, still involves many research problems; for example, crowd dynamics involve psychological aspects (such as preferences and habits) that influence individual behavior, thereby increasing scene complexity. Aside from the thoughts and intent of individuals, the constant interaction

among them makes it difficult to distinguish one from another. In addition, occlusions result in loss of observation of a target object, while the object's appearance (such as shape and color) varies, not only from one setting to the next, but also as a given setting evolves.

**Acknowledgment**

This article summarizes and incorporates three earlier publications: Ali et al.,<sup>1</sup> Ali and Shah,<sup>2</sup> and Mehran et al.<sup>17</sup> The research is partially supported by the U.S. Army Research Office, part of the U.S. Army Research Laboratory, under grant number W911NF-09-1-0255 and by the U.S. Department of Defense. ■

**References**

1. Ali, S. and Shah, M. A Lagrangian particle dynamics approach for crowd flow segmentation and stability analysis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Minneapolis, June 18–23, 2007), 1–6.
2. Ali, S. and Shah, M. Floor fields for tracking in high-density crowd scenes. In *Proceedings of the 10th European Conference on Computer Vision* (Marseille, France, Oct. 12–18), Springer, 2008.
3. Andrade, E.L., Blunsden, S., and Fisher, R.B. Modeling crowd scenes for event detection. In *Proceedings of the 18th International Conference of Pattern Recognition* (Hong Kong, Aug. 20–24, 2006).
4. Bennett, A. *Lagrangian Fluid Dynamics*. Cambridge University Press, New York, 2006.
5. Brostow, G. and Cipolla, R. Unsupervised Bayesian detection of independent motion in crowds. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (New York, June 17–22, 2006).
6. Burstedde, C., Klauck, K., Schadschneider, A., and Zittartz, J. Simulation of pedestrian dynamics using a two-dimensional cellular automaton. *Physica A: Statistical Mechanics and its Applications* 295, 3–4 (June 2001), 507–525.
7. Chan, A.B. and Vasconcelos, N. Bayesian poisson regression for crowd counting. In *Proceedings of 12th IEEE International Conference on Computer Vision* (Kyoto, Sept. 27–Oct. 4, 2009), 545–551.

8. Chan, A.B. and Vasconcelos, N. Mixtures of dynamic textures. In *Proceedings of the 10th IEEE International Conference on Computer Vision* (Beijing, Oct. 17–20, 2005), 641–647.
9. Helbing, D. Traffic and related self-driven many-particle systems. *Review of Modern Physics* 73, 4 (Dec. 2001), 1067–1141.
10. Helbing, D. and Molnar, P. Social force model for pedestrian dynamics. *Physical Review E* 51, 5 (May 1995), 4282–4286.
11. Hughes, R.L. The flow of human crowds. *Annual Review of Fluid Mechanics* 3 (2003), 169–182.
12. Hughes, R.L. A continuum theory for the flow of pedestrians. *Transportation Research (Part B: Methodological)* 36, 6 (July 2002), 507–535.
13. Kirchner, A. and Schadschneider, A. Simulation of evacuation processes using a bionics-inspired cellular automaton model for pedestrian dynamics. *Physica A: Statistical Mechanics and its Applications* 312, 1–2 (Sept. 2002), 260–276.
14. Kotelenetz, P. *Stochastic Ordinary and Stochastic Differential Equations*. Springer, New York, 2008.
15. Kratz, L. and Nishino, K. Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Miami, June 20–26, 2009), 1446–1453.
16. Marques, J.S., Jorge, P.M., Abrantes, A.J., and Lemos, J.M. Tracking groups of pedestrians in video sequences. In *Proceedings of the IEEE Computer Vision and Pattern Recognition Workshop* (2003), 101.
17. Mehran, R., Oyama, A., and Shah, M. Abnormal behavior detection using social force model. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (Miami, June 20–26, 2009), 935–942.
18. Pellegrini, S., Ess, A., Schindler, K., and van Gool, L. You'll never walk alone: Modeling social behavior for multi-target tracking. In *Proceedings of the 12th IEEE International Conference on Computer Vision* (Kyoto, Sept. 27–Oct. 4, 2009).
19. Reisman, P., Mano, O., Avidan, S., and Shashua, A. Crowd detection in video sequences. In *Proceedings of the IEEE Intelligent Vehicles Symposium* (Parma, Italy, June 14–17, 2004), 66–71.
20. Sand, P. and Teller, S. Particle video: Long-range motion estimation using point trajectories. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (New York, June 17–22, 2006), 2195–2202.
21. Shadden, S.C., Lekien, F., and Marsden, J.E. Definition and properties of Lagrangian coherent structures from finite time Lyapunov exponents in two-dimensional aperiodic flows. *Physica D: Nonlinear Phenomena* 212, 3–4 (Dec. 2005), 271–304.
22. Tu, P., Sebastian, T., Doretto, G., Krahnstoeber, N., Rittscher, J., and Yu, T. Unified crowd segmentation. In *Proceedings of the 10th European Conference on Computer Vision* (Marseille, Oct. 12–18, 2008), 691–704.
23. Yilmaz, A., Javed, O., and Shah, M. Object tracking: A survey. *ACM Computing Surveys* 38, 4 (2006), 13.1–13.45.
24. Zhan, B., Monekosso, D.N., Remagnino, P., Velastin, S.A., and Xu, L. Crowd analysis: A survey. *Machine Vision and Applications* 19, 5–6 (2008), 345–357.
25. Zhao, T. and Nevatia, R. Tracking multiple humans in a crowded environment. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Washington, D.C., June 27–July 2, 2004), II-406–II-413.

**Brian E. Moore** (bmoore@math.ucf.edu) is an assistant professor of mathematics in the Department of Mathematics at the University of Central Florida, Orlando, FL.

**Saad Ali** (saad.ali@sri.com) is a computer scientist in the Vision Technologies department at SRI International Sarnoff, Princeton, NJ.

**Ramin Mehran** (ramin@cs.ucf.edu) is a Ph.D. student in the Computer Vision Lab at the University of Central Florida, Orlando, FL.

**Mubarak Shah** (shah@cs.ucf.edu) is Agere Chair Professor of Computer Science and founding director of the Computer Vision Lab at the University of Central Florida, Orlando, FL.