Chapter N

FROM BACKGROUND SUBTRACTION TO THREAT DETECTION IN AUTOMATED VIDEO SURVEILLANCE

Joshua Eckroth
The Ohio State University, Columbus, OH

Dikpal Reddy
University of Maryland, UMIACS, College Park, MD

John R. Josephson
The Ohio State University, Columbus, OH

Rama Chellappa
University of Maryland, UMIACS, College Park, MD

Timothy N. Miller
The Ohio State University, Columbus, OH

INTRODUCTION: PERSISTENT VIDEO SURVEILLANCE

As sensors, such as video cameras, become cheaper and easier to deploy and network, the opportunity increases for using networks of such sensors to provide useful information for military operations, such as video surveillance for facilities protection, and "persistent surveillance" ("persistent ISR", "persistent stare") to maintain sensory contact with targets of interest (Pendal, 2005). However, without assistance from automation, humans will be overloaded by information, and unable to use it effectively.

"What information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention, and a need to allocate that attention efficiently among the

overabundance of information sources that might consume it." (Simon, 1971)

It is difficult for humans to vigilantly monitor a large number of video feeds for extended periods without fatigue, or complacency, especially if they are tasked with recognizing the significance of rare events in a complex stream of events (Warm, *et. al.*, 1996; Grier, et. al. 2003; Pattyn, *et. al.*, 2008). Moreover, sometimes events of interest cannot be recognized simply from the video, without knowing where a specific camera is pointing, at a restricted area, for example. Keeping track of the location and significance of a camera's field of view imposes an additional cognitive burden. The burden is even greater if recognizing events of interest requires mentally noting entities as they become visible in different cameras, identifying some entities seen in separate views as the same entity, identifying some entities that newly appear in some views as the same entities previously seen in other views (mentally tracking entities in "object space"), and accessing mental maps to understand the significance of motion paths.

Automation can help. It can potentially provide users with alerts based on recognizing indicators of threatening behavior, including behavior that cannot be detected with a single camera. For example, the movement of an entity from place to place in relation to the map might show a pattern indicative of scouting the perimeter of a facility, where no single camera view shows anything suspicious. Many unsolved technical problems remain, however, including problems about how to extract needed information from video imagery, and problems of how to process acquired information to climb the levels of abstraction from individual camera-centered frameworks to a world-centered framework, and from motion description ("Object O_{5604} moved along path P_{52} from location L_{12} at time T_{55} to L_{13} at T_{66}.") to behavior description ("O_{5604} proceeded slowly north on road R_3 to the intersection with R_7, turned East on R_7….") to recognition of indicators of threat ("O_{5604} slowly circled the facility.").

This paper describes some recent progress in technology for video surveillance that specifically addresses methods for background subtraction to detect changes from frame to frame in video, tracking of viewed objects in camera-centered "image space," tracking of objects in world oriented "object space" using information from multiple cameras, and methods for "climbing" levels of abstraction in descriptions of behavior.

BACKGROUND SUBTRACTION AND TRACKING USING COMPRESSIVE CAMERAS

Introduction

Most cameras installed for surveillance applications are used for simple vision tasks such as detection and tracking, pose estimation, and 3D reconstruction from silhouettes, and very few for higher level tasks such as activity recognition. For example, much of simple tracking can be done on silhouette blobs. In many of these tasks, the amount of data collected is huge for the purpose of the application. For example, a typical background subtraction algorithm uses fully sampled images for background and foreground, whereas the silhouette image obtained from it only occupies a small region of the image. Similarly, when tracking, we are interested in only a small region of the image that is moving and do not care about the other parts of the image. Nevertheless the entire image is commonly sensed and transmitted to a central location for processing. A typical camera network scenario involves the network observing a scene and then relaying it to a central unit for processing. This means that a huge amount of data, which is ultimately useless, is first collected and then transmitted over a channel, thus wasting sensing and bandwidth resources. In this paper, we present a solution to alleviate this problem. We show our approach on two vision applications: background subtraction and tracking. In both of these applications, we use a "compressive camera" to observe the images and then process is done on these measurements.

A compressive camera is a device that has been built on the principles of compressed sensing (Candes, 2006). Such a camera measures, not the image pixels like a conventional camera, but a small number of random linear projections of the image. This inherently reduces the sensing capacity, since the same images can now be sensed in a compressed form. We present an innovative approach that utilizes the redundancy present in vision applications to further reduce the number of sensed measurements and hence the bandwidth and storage requirement. Vision tasks are performed, not on the full images, but on the compressed measurements directly, without the need to reconstruct the full images. By working over compressed measurements, and utilizing the redundancy of the problem, a significant reduction is achieved in the amount of data that needs to be sensed. In the next section, we provide a brief introduction to compressed sensing theory and compressive cameras. We then present our approach to background subtraction and tracking on compressed measurements.

Compressive cameras and computer vision applications

Compressive Sensing. Suppose we have an image x of size $N \times 1$ (i.e., vectorized), then we can represent the image in some basis Ψ as

$$x = \Psi\theta \tag{1}$$

where θ is a sparse coefficient vector (K -sparse). vector θ has very few large non-zero components, indicating that the image can be compressed. Wavelets are an example of such a basis.

In the CS framework (Candes, 2006) we do not measure the K largest elements of θ but we instead measure $M < N$ linear projections of the image x onto another basis Φ.

$$y = \Phi x = \Phi\Psi\theta \tag{2}$$

where y are the compressed measurements. Since $M < N$ the system of equations are underdetermined, but utilizing the sparsity of θ we can recover the signal by solving the following l_1-optimization problem called Basis Pursuit.

$$\hat{\theta} = \arg\min \|\theta\|_1 \; s.t. \; y = \Phi\Psi\theta \tag{3}$$

Compressive cameras. Based on compressive sensing (CS) theory, a prototype single pixel camera (SPC) was proposed in (Wakin, 2006). The SPC hardware is specifically designed to exploit the CS theory, and differs from a conventional camera by using only a single optical photodiode (infrared, hyperspectral, or visual) along with a digital micro-mirror device (DMD). It also combines the sampling and compression process, unlike a conventional camera. The DMD is used to generate the random linear projections described in the theory. A compressive camera, such as SPC, provides, not the pixels, but the compressed measurements. This is especially useful in bandwidth constrained networks, where the compressed measurements can be transmitted to a central location for processing, and where the image can be recovered.

Background subtraction. Background subtraction is fundamental in detecting and tracking objects, and has applications in surveillance, 3D modeling, etc. In previous background subtraction algorithms, the background and foreground are fully sampled images from conventional cameras. After the difference operation, the background images are discarded or used in future background models. This approach of discarding fully sampled images can be very expensive in normal imaging, and is particularly so in hyperspectral imaging. Using a compressive camera, such as SPC, can help with the problem. But sensing the

foreground and background images in a compressed form, and then reconstructing the images to perform background subtraction, can be computationally expensive. Instead, in (Cevher, 2008) we showed that background subtracted silhouettes can be obtained directly from the compressed measurements. While background subtraction on compressed images is not new (Aggarwal, 2006), unlike previous approaches, we sense the images directly in a compressed form using the SPC architecture. We also show that, since the difference images which we desire are sparse in the spatial domain (Fig.1. in Cevher, 2008), they should be sparser than complete images in the appropriate basis. This permits sensing the images with even fewer measurements than those that compressive cameras already permit.

The results of this background subtraction approach can be seen in Fig. 4.1. The background image (top row) and the test image (second row) are reconstructed from the compressed measurements. The difference of these images is shown in third row. The fourth row shows the difference image reconstructed directly from the compressed measurements. The columns correspond to measurement rates M/N of 50%, 5%, 2%, 1% and 0.5% respectively.

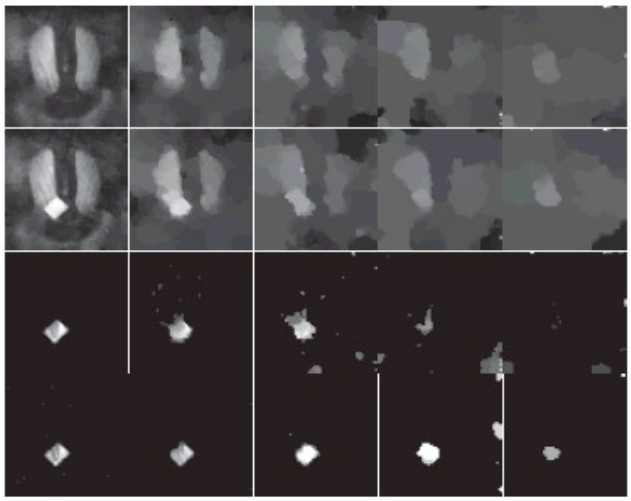

Figure 4.1.
Background subtraction experimental results using a SPC

We adapt to the changes in background by updating the background model, not by reconstructing the background images, but by using the compressed measurements themselves (Fig. 2. of Cevher, 2008). This

allows us to adapt to various changing backgrounds without the need to reconstruct the full images from compressive measurements. We applied our algorithm on data collected under changing illumination. We showed that our adaptive background subtraction algorithm, over compressive measurements, under changing illumination, works as well as working with normal full images (Fig. 6. of Cevher, 2008).

It should be noted that our algorithm can also recover the appearance of the objects by reconstructing a single auxiliary image. Hence, depending on the application, we can either recover silhouettes only, or with minimal computation, recover the appearance of the object. We believe that this is a significant improvement over previous sensing and background-subtraction methods.

In summary, our background-subtraction algorithm, not only works on compressive measurements, but also takes advantage of the sparsity of the problem to decrease the amount of sensing required, and hence the requirements for storage and bandwidth. Further, the approach can adapt to changes in background.

Tracking. Many tracking algorithms use background subtracted silhouettes to track objects. The silhouette image corresponds to some object which is in motion and which we would like to track. In many scenarios the background subtracted silhouette is sparse in the image domain. This sparsity translates directly into sparsity of the object parameter (such as location, volume, etc.) in space. Since we would like to estimate a sparse parameter which corresponds to a sparse region in an image, it is waste of sensing resources to observe complete images using conventional cameras, perform background subtraction to throw away most of the image, and then track. In (Reddy, 2008) we showed tracking as a sparse approximation problem, and we argued that to do so we need to measure, not the entire image, but a few random projections of the image vectors appropriately picked.

To illustrate the idea, we show our results on the simpler case of background subtracted silhouette images. We relate the object parameter (such as location) and the corresponding image pixel intensity using Eq. 7. and Eq. 8. of (Reddy, 2008), respectively. This relation can be visualized in Figure 4.2.

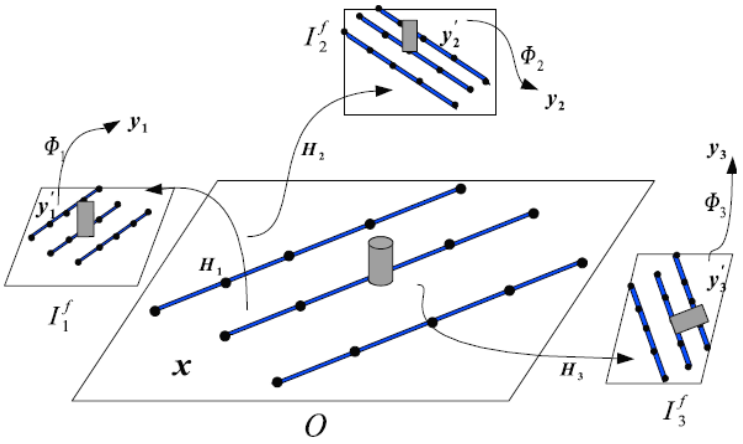

Figure 4.2.
Ground plane tracking scenario

We tested the algorithm on an outdoor scene with walking people. A few images of the test scene are shown in Fig. 4.3.

Figure 4.3.
Outdoor scene where moving people are tracked using background subtracted silhouette images

We resolve the position of occluding objects by fusing the silhouette images from multiple views in a manner similar to (Khan, 2006).

One of the principal advantages of this approach is that it easily scales to large number of cameras since the complexity of the algorithm depends on the grid size that is used for localization. The speedup of this approach is shown in Fig. 4.4.

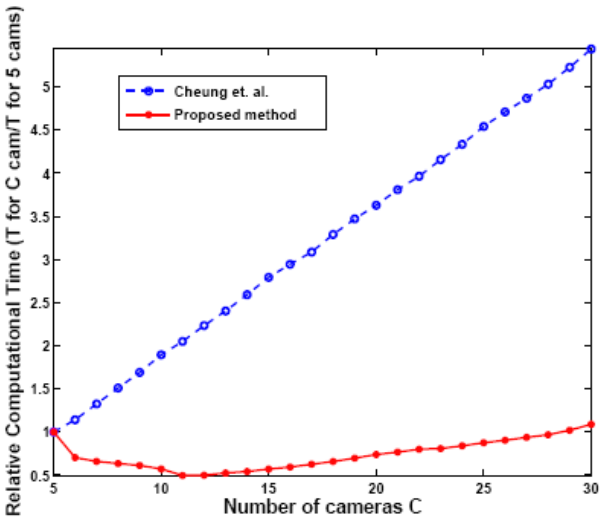

Figure 4.4.
Speedup achieved as the number of cameras increases

The results of tracking people in an outdoor scene are shown Fig. 4.5.

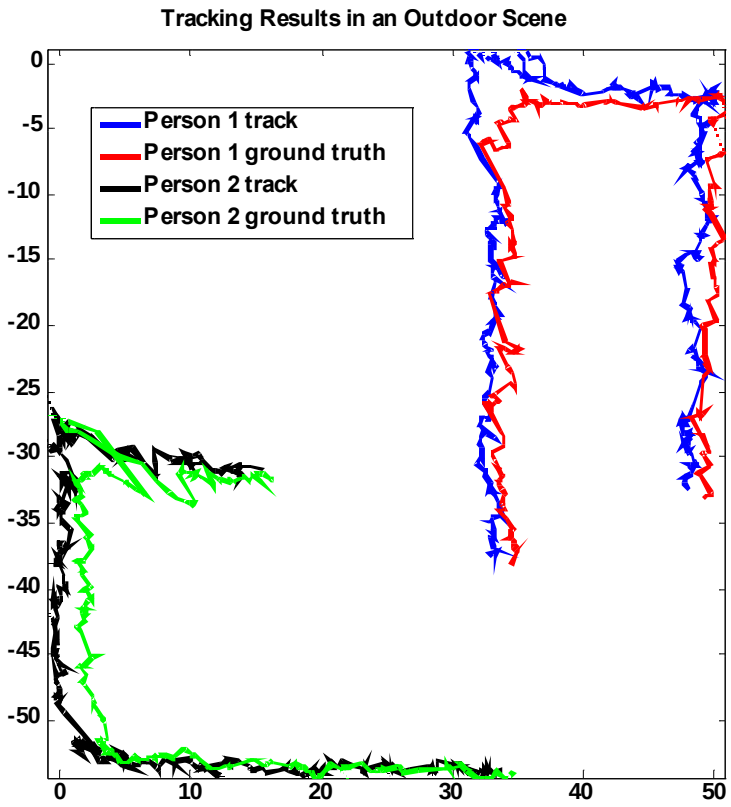

Figure 4.5.
Tracking results in an outdoor scene over 500 frames using difference of compressed measurements of foreground and background image

Summary: We have described methods for background subtraction and tracking using the compressed measurements of a compressive camera.

USE OF ABDUCTIVE INFERENCING FOR TRACKING IN OBJECT SPACE AND THREAT DETECTION

Introduction

As technology advances, and video cameras become cheaper and easier to deploy, video surveillance will use increasing numbers of cameras, often with overlapping views, while gaps in coverage will become smaller. In principle, smaller gaps will make it harder for objects of interest to escape detection, and enable objects to be tracked for longer periods of time. Longer tracks enable more complex forms of behavior to be recognized,

and increase the likelihood that the location of an object of interest will be known when the time comes to act. In principle, overlapping views enable 3D from stereo, reduced positional uncertainty, and improved recognition of object characteristics and behavior. However, for these potentials to be realized, a number of technical problems must be solved.

In this section, we describe methods for tracking in world-oriented coordinates (object space) using information from multiple cameras, and methods for ascending levels of abstraction in descriptions of behavior. The inputs are assumed to be the outputs of single-camera background-subtraction and tracking algorithms in the form of reported "detections" (detected movements) and track updates.

Tracking in object space

When tracking is done in "image space," using single-camera data, the problem remains of determining which of these tracks represent the same external objects. Actually, this problem is not much different from that of determining which single-camera detections represent the same external entities. In both cases, the information for making the determinations consists of information about the fields of view of the cameras in world-oriented coordinates, the locations of the detections or track in the images, the visible characteristics of entities detected or tracked, and knowledge of entities types and terrain conditions that constrain movements, and thereby sometimes enable reidentification of entities after they pass through gaps in sensor coverage.

Since track-update reports are just entity-detection reports with the additional information associating a new detection with a sequence of earlier ones, the possibility arises of doing the tracking in object space, especially if detections can be mapped to world-oriented locations, for example when entities can be assumed to be on the ground plane. There are several potential benefits of tracking in object space, rather than relying only on single sensors for tracking. These include:

- benefits from using world units (e.g., meters) when measuring track velocities. This enables kinematic constraints from physics and object types to be expressed and to be used to constrain expected and possible movements.
- Processing in object space can determine which cameras should see certain entities, and then those cameras (or the data from those cameras) can be queried to determine if an entity was seen as expected. If

it was not, then a putative entity may be noise, or there might be occlusion; error in estimated sensor location, alignment or sensitivity; or sensor failure. Depending on the operational demands of the application, these possibilities can be evaluated explicitly, and systems adapted accordingly.

- Focus of attention can be directed to locations of interest, expressed in world coordinates, so that cameras can be directed to pan/zoom/focus to corroborate detections, anticipate the arrival of entities in the field of view, refine location estimates, or pick up additional details.

However, one potential drawback of mapping detections to object space before tracking is that image details will need to be transmitted to the processing location, or left out of consideration by the tracking algorithm. 2D shapes and color histograms are examples of image details that may be useful for tracking.

Tracking using abductive inferencing

We investigated methods for tracking that use abductive inferencing to generate and evaluate hypotheses for explaining detection reports.

Abduction. Abductive inferencing (inference to the best explanation) is a distinctive pattern of reasoning. It is ubiquitous in ordinary life, and in the trained reasoning of intelligence analysts, diagnosticians, accident investigators, and scientists (Josephson & Josephson, 1996). It can be considered a part of commonsense logic. Researchers have developed various kinds of computational models of abductive inferencing, although they have not always referred to it as abductive inferencing. Our work treats abduction as a kind of knowledge-based problem solving, where the reasoning process is analyzed into tasks, methods, and subtasks to achieve problem-solving goals, and where the choice of a method (and knowledge representation) for a task depends on such considerations as the forms in which knowledge is available, and the operational demands on quality and timeliness of solutions (Chandrasekaran & Johnson, 1993).

In using abductive inferencing for tracking from detection reports, the reports are the inputs that need to be explained. Hypotheses available for explaining detections are of three types: noise (e.g., spurious detections, light reflections, camera jitters), movements of known entities, or first detections of an entity.

Noise. Several different types of "noise" occur, and must be either filtered out in image processing, or accounted for. Some types of noise are highly transient. These include glints of solar reflection and pixel-level noise from physical sensors. Another type of noise consists of real objects that are fixed in overall position, but move in response to wind (e.g., vegetation, flags). Another type is a result of shadows of objects visible in the scene (and moving with them) or shadows of objects that are out of view, such as aircraft and clouds. Another type consists of real object, but of types that are not of interest for the application, such as birds, small animals, and tumbleweed.

Small and transient differences between video frames can be filtered out as probable noise, and not reported as detections. Thresholds can be set empirically to minimize both false positives and false negatives, taking into account estimated costs for either type of mistake. With feedback, thresholds might be set adaptively based on successes in interpreting detections into tracks, and failures to detect entities that are otherwise indicated to have been in the field of view, either by other sensors, or by filling in the gaps in a dashed line.

Noise consisting of the effects of wind on fixed objects will probably best be treated as explicit hypotheses about vegetation, flags, and such, or at least as belonging to the class of fixed objects that stay in place, but occasionally wiggle. An advantage of doing so is that hypotheses about such objects can be maintained in the world estimate, so subsequent detections with similar characteristics in the same location will not draw additional computational or attentional resources, and the likelihood can be reduced of confounding such objects with nearby moving objects of interest.

Shadows of visible objects deserve special treatment. Shadows tend to look very much like the background, but dimmer. This similarity can be exploited to help filter them out, treating them as part of the background (Sexton and Zhang, 1993). However, sometimes shadows are too dark for the background to show through, and other methods are needed. One can imagine an abductive approach that treats shadows as parts of a complete 3D scene reconstruction that includes an illumination model. Shadows are then predicted by the illumination, together with the opacity of the shadow-causing objects, and explain why certain segments of surfaces are darker, shaped as they are, and move with the objects; these properties of surface segments, in turn, explain properties of 2D image segments. However, to our knowledge, a treatment of shadows of this sort has not been achieved technically.

Similar to the effects of wind on fixed objects, small objects that are not of interest for the application will probably best be treated as explicit hypotheses about birds, small animals, wind-blown objects, and so on. Explicit recognition of such phenomena can spare computational and attentional resources, and enable better discrimination of objects that really are of interest.

None of these methods for handling noise have been implemented yet in our experiments, except for the filtering out of small and transient differences between video frames, so they do not cause detections, and the filtering out of detections that do not find subsequent interpretation as belonging to tracks that move significantly from their initial positions.

Mapping detections from image space to object space. We assume that camera locations, alignments, angular fields of view, and image distortion characteristics, have been predetermined accurately enough for practical purposes. This information about the cameras is then available for use in mapping detections to object space. Where the fields of view of cameras overlap, 3D surfaces can be reconstructed from stereo processing, and these 3D surfaces can be transformed to surfaces in object space using the camera information. However, sometimes there will be no overlap, and 3D surface reconstruction is computationally expensive. Moreover, sometimes we are exclusively interested in tracking objects moving on the terrain surface, when viewed from relatively high up. We have been experimenting with data sets of this sort. Under these circumstances, the locations of detections can be mapped from image space to object space by applying a warp, that is, a perspective transformation, that maps image locations to locations in the ground plane. We have implemented this method for our experiments. Tracking is then done in the object space (in the ground plane). Where fields of view overlap, detections are intermingled in the ground plain, and tracking uses detections whose sources are different cameras. Where fields of view do not overlap, tracking is also done in the object space, but using detections from single sources. A warp is invertible, so tracks in object space can be mapped back to image space for display.

Tracking experiments

We treated detections, with locations mapped into object space, as the items that need to be explained by abductive processing. As we described, hypotheses available for explaining these detections are of three types: noise, movements of known entities (a known track continues to there), or

first detections of an entity. A first detection will typically correspond to an object entering the field of view, but it might be a previously unmoving object that begins moving (e.g., an auto in the parking lot).

Tracks may split or join. Such phenomena actually occur in the data used in our experiments. People walk in a group, which is tracked as a single object until they go their separate ways. People walking separately converge and walk together. People get out of autos, or enter them. Explicitly noting splitting and joining events is important for classifying such events explicitly, as multiple people exiting a vehicle for example, both because classifying such events may be useful for recognizing significant behavior, and because it helps with decomposing tracks into segments over which the tracked object has a consistent type (e.g., person, vehicle, group of people). Thus, the desired output at the first stage of abductive processing consists of "track segments," where a track segment goes from a track origination point, or from a splitting or joining event, to another splitting or joining event, or to a track ending point.

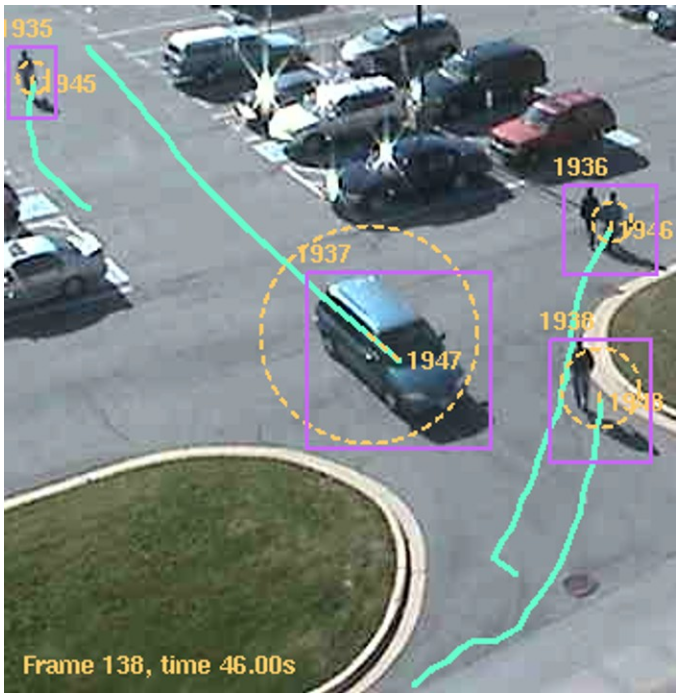

Figure 4.6
Tracking multiple entities

Our experiments used video data from two fixed cameras mounted on the building at the ARL facility in Adelphi Maryland, and looking downward at a parking lot. The fields of view partially overlap. Individual persons, groups of people, and vehicles are seen moving about (see Figure 4.6). These entities may initially appear by entering the field of view, or from within the field of view, such as when a vehicle begins moving after remaining parked for some time. (Detections come from a background-subtraction process, so unmoving objects are not detected.) We see splits and joins when groups of people disband or form, and when entities cross paths or obscure one another.

Figure 4.7 shows the results of tracking a vehicle based on detections from two different cameras. The detections were mapped to object space, as described previously, by warping the images to the ground plane. The tracking was done in object space, with the results mapped back into the camera image spaces using the inverse warps, for display. The results were also mapped to the image space of an overhead photo, which is shown in the figure. The straight lines in the overhead image correspond to the edges of the views from the two cameras.

Figure 4.7
Vehicle track in three views

Limitations. The more general task of automated video surveillance involves many more elements than those investigated in our experiments to date. The number of video cameras employed may be many, and some or all of them may be mobile. Additionally, cameras may be actively controlled; they might be panned, zoomed, or focused to maximize information pickup from objects or regions of interest. To conserve power, a camera might be turned on only when needed, e.g., when an object is expected to come into the field of regard, or to corroborate uncertain information from another camera. The general problem also includes the need to ascending levels of abstraction in descriptions of behavior from track seg-

ments, through intermediate levels, to descriptions of threats and perhaps opportunities.

Climbing levels of abstraction in descriptions of behavior

Our general approach to processing at multiple levels of abstraction is based on viewing the problem as one of layered abduction, where the conclusions of one layer become data to be explained at the next higher layer (Josephson & Josephson, 1996, Ch. 10; Schenk, 1995; Banerjee, 2006). Our current design for automated video surveillance postulates the following layers, characterized by the types of hypotheses evaluated:

- Detected movements (JDL fusion Level-0).
- Track segments (JDL fusion Level-1),
- Classified track segments - tracked objects are classified (person, group of people, vehicle, etc.) and possibly identified individually (John Smith, UPS truck, vehicle with license plate WOOF). Splitting and joining events are classified by the types of their inputs and outputs (group of people splits into three separate people). (Also JDL fusion Level-1),
- Behavior - classified track segments are described in relation to terrain features and entity states (vehicle slowly exited the parking lot, pedestrian walked on the sidewalk), and split/join events are described in terms of entering and emerging from vehicles, and from groups of people. (Vehicle came to a stop in the parking lot, two people emerged and walked toward the building entrance.) (JDL fusion Level-2),
- Scripts – stereotypical patterns of activity (Schank & Abelson, 1977) - (person parking a vehicle in the visitor section of the parking lot and proceeding on foot to the building entrance; vehicle cruising the parking lot looking for an empty spot). The scripts are presumably prestored, and tagged with their threat significance, e.g., as benign or possibly threatening. Behavior that cannot be interpreted as part of a known script is considered to be anomalous, especially when a relatively complete library of scripts has been built up, and anomalous behavior triggers an alert. (JDL fusion Level-3).

A description of JDL fusion levels can be found in Steinberg, Bowman & White (1998). We have not yet implemented any of the layers above that of track segments, although we have previously experimented with layered abduction models for several other applications (*op. cit.*).

SUMMARY

We have described recent progress in technology for video surveillance that addresses: methods for background subtraction to detect changes from frame to frame in video, tracking of viewed objects in camera-centered "image space," tracking of objects in world oriented "object space", and methods for climbing levels of abstraction in descriptions of behavior.

COLLABORATIONS

The work on background subtraction and tracking using compressed images was done in close collaboration with Volkan Cevher, Marco Duarte and Prof. Richard Baraniuk of Rice University. The work on abductive tracking in object space benefited from collaboration with Sean McGinnis and John Hancock of ArtisTech, Inc. and Robert Winkler of ARL.

REFERENCES

Aggarwal, A., Biswas, S., Singh, S., Sural, S., Majumdar, A.K.: Object Tracking Using Background Subtraction and Motion Estimation in MPEG Videos. In: *ACCV*, Springer (2006) 121–130.

Banerjee, B. (2006). A Layered Abductive Inference Framework for Diagramming Group Motions, *Logic Journal of the IGPL* 14(2) 363-378, 2006.

Bulich, C., Klein, A., Watson, R., and Kitts, C. 2004. Characterization of Delay-Induced Piloting Instability for the Triton Undersea Robot. In the *Proceedings of the 2004 IEEE Aerospace Conference*, Big Sky MT.

Candes, E.: Compressive sampling. In: *Proceedings of the International Congress of Mathematicians*. (2006).

Casper, J. and Murphy, R. R. (2003). Human-robot interactions during the robot-assisted urban search and rescue response at the World Trade Center. *IEEE Transaction on Systems, Man and Cybernetics-Part B: Cybernetics*, Vol 33, No. 3, pp 367-385.

Cevher, V., Sankaranarayanan, A., Duarte, M.F., Reddy, D., Baraniuk, R.G., Chellappa, R.: Compressive Sensing for Background Subtraction. In: *ECCV*, Marseilles (Oct 2008).

Chen, J, Y. C. and Thropp, J. E. 2007. Review of low frame rate effects on human performance. *IEEE Transaction on Systems, Man and Cybernetics* – Part A: Systems and Humans, Vol. 37, pgs 1063-1076.

Chandrasekaran, B. & Johnson, T. R. (1993). Generic Tasks and Task Structures: History, Critique and New Direction. In: David, J. M., Krivine, J. P. & Simmons, R., (Eds), *Second Generation Expert Systems*. New York: Springer-Verlag.

Endsley, Mica and Garland, Daniel 2000. *Situation Awareness Analysis and Measurement.* Mahwah, NJ: Lawerence Earlbaum Associates.

Fiorini, P. and Oboe, R. 1997. Internet-Based Telerobotics: Problems and Approaches. In the *Proceedings of the International Conference on Advanced Robotics*, Monterey.

Grier, R. A., Warm, J. S., Dember, W. N., Matthews, G., Galinsky, T. L., Szalma, J. L. & Parasuraman, R. (2003). The Vigilance Decrement Reflects Limitations in Effortful Attention, Not Mindlessness. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 45: 349-359

Hewish, M. 2000. Pilotless progress report – UAVs have made exceptional strides recently. *Jane's International Defense Review* – October 01, 2000. http://search.janes.com

Josephson, J. R. & Josephson, S. G., Eds. (1996). *Abductive Inference: Computation, Philosophy, Technology* New York: Cambridge University Press.

Khan, S. M., Shah M., 2006. A multi-view approach to tracking people in crowded scenes using a planar homography constraint. In: *ECCV*, 2006, vol. 4, 133–146.

Krotkov, E., Simmons, R., Cozman, F., and Koenig, S. 1996. Safeguarded Teleoperation for Lunar Rovers: From Human Factors to Field Trials. In *Proceedings of the IEEE Planetary Rover Technology and Systems Workshop*, Minn., MN.

Pattyn N, Neyt X, Henderickx D, Soetens E., 2008. Psychophysiological investigation of vigilance decrement: Boredom or cognitive fatigue? *Physiology & Behavior* [serial online]. January 28, 2008;93(1/2):369-378. Elsevier. Available from: Academic Search Complete, Ipswich, MA. Accessed April 15, 2009.

Pendall, Major David W., 2005. Persistent Surveillance and its Implications for the Common Operating Picture. *Military Review*: 41-50.

Reddy, D., Sankaranarayanan, A., Cevher, V., Chellappa, R.: Compressed sensing for multi-view tracking and 3-D voxel reconstruction In: *ICIP*, San Diego, CA (Oct. 2008).

Schank, R. & Abelson, R. (1977). *Scripts, Plans, Goals, and Understanding,* Hillsdale, NJ: Lawrence Erlbaum.

Schenk, T. (1995). A Layered Abduction Model of Building Recognition, in *Automatic Extraction of Man-Made Objects in Aerial and Space Images*, Gruen, A.; Kuebler, O.; Agouris, P. (Eds.), Birkhäuser, 1995.

Sexton, G.G.; Zhang, X., Suppression of shadows for improved object discrimination, *IEEE Colloquium on Image Processing for Transport Applications*, Volume , Issue , 9 Dec 1993, pp. 9/1 - 9/6.

Simon, Herbert, 1971. In Martin Greenberger (ed.) *Computers, Communications and the Public Interest,* pp. 40-41, Johns Hopkins Press.

Steinberg, A. N., Bowman, C. L. and White, Jr., F. E. (1998) "Revisions to the JDL Data Fusion Model," *Proc. 3rd NATO/IRIS Conf.*, Quebec City, Canada.

Voshell, M., Woods, D. D., and Phillips, F. 2005. Overcoming the keyhole in human-robot coordination: Simulation and Evaluation. *Proceedings of the Human Factors and Ergonomics Society* 49th Annual Meeting, 26 – 30 September, Orlando, FL.

Wakin, M.B., Laska, J.N., Duarte, M.F., Baron, D., Sarvotham, S., Takhar, D., Kelly, K.F., Baraniuk, R.G.: An architecture for compressive imaging. In: *ICIP*, Atlanta, GA (Oct. 2006) 1273–1276.

Warm, Joel S., Dember, William N., Hancock, Peter A. Vigilance and workload in automated systems. *Automation and human performance: Theory and applications* (1996).