Surveillance video analytics for large camera networks

Peter Tu, Fred Wheeler, Nils Krahnstoever, Thomas Sebastian, Jens Rittscher, Xiaoming Liu, Amitha Perera, and Gianfranco Doretto

Intelligent video research focuses on automatic camera calibration and person detection, tracking, and reacquisition.

Extensive networks of surveillance cameras are increasingly deployed in public and private facilities, with tremendous potential value for safety and security. However, because manual monitoring of large numbers of video sources is not feasible, surveillance imagery is often simply directed to mass-storage devices, to be used only forensically and for data mining. On-line computer-based video analysis represents an alternative to this traditional approach. This new paradigm continually apprises security personnel of who is on site, where they are, and what they are doing. It offers the prospect of increased productivity and highly advanced site-wide security.

Our approach to video analytics for security starts by establishing camera geometry. We determine the relative location of each camera with respect to a world coordinate system, and so can readily coordinate observations from different cameras. The next step is to detect people under a variety of imaging conditions, including both crowds and situations with dynamic backgrounds. Reliable detection of individuals on a frame-by-frame basis enables tracking over time. However, gaps in camera coverage require reacquisition of individuals to be tracked as they move between discontinuous regions of coverage.

In our work with automatic camera calibration, we assume a dominant ground plane and, further, that people walking on it are of some nominal height. We have shown that—even in the presence of significant noise—camera position, orientation, and focal length can be estimated by simple observation of individuals walking on site.[1] If the background is relatively static, moving foreground objects can be detected and classified as either human or non-human.[2] Where such foreground-background segmentation is not possible, methods can be employed to continually scan an image for persons.[3] We have found
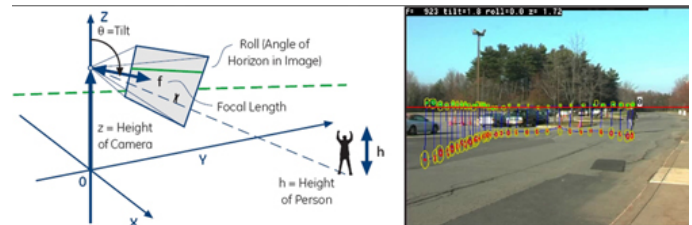
Figure 1. Observations of individuals enable automatic estimates of focal length, camera height, pan and tilt angles.


Figure 2. A crowd can be automatically segmented into individuals using on-line global optimization.

that, in crowded conditions, a global approach to person detection is warranted.[4] For person reacquisition, we use both face recognition and matching based on general appearance.[5]

With regards camera calibration, the relationship between the image locations of feet and head are governed by a homology that can be represented by a 3×3 matrix. Multiple head and foot observations and careful analysis of associated errors enable estimation of the foot-to-head homology. Assuming the camera has
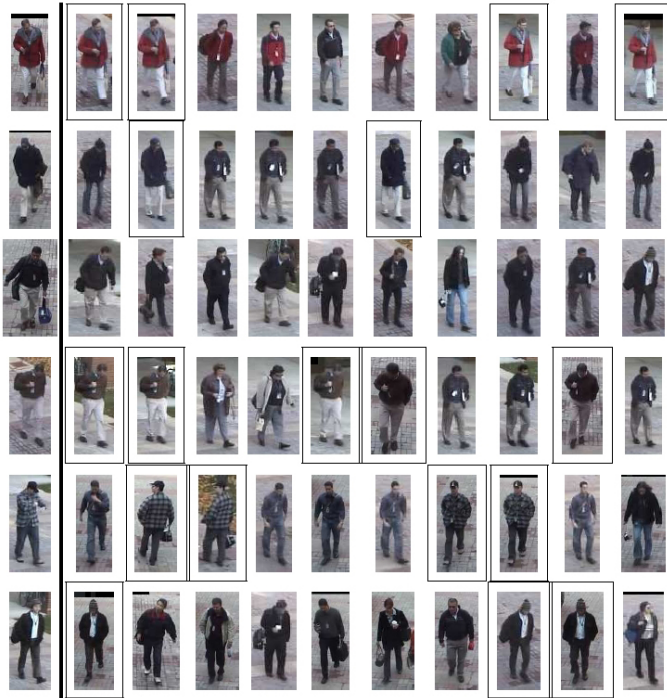
Figure 3. For person reidentification, probe images in the first column (far left) are succeeded by a series of likely matches based on appearance. Boxed images indicate correct responses.

Figure 4. Image from a short video sequence (left) and enhanced image (right) constructed using super-resolution methods over multiple images

no skew and its principle ray is at the center of the image, the height, focal length, and tilt can be computed based on an eigen-decomposition of the homology. When cameras with overlapping fields of view have joint observations, the relative positions of the camera centers can be determined. This calibration information allows for rejection of many false person detections and facilitates camera-to-camera handoff (see Figure 1).

Our crowd segmentation method is based on the ability to develop hypothetical individuals from multiple part-based feature detection.[4] In place of a local-greedy-detection approach that often lacks sufficient contextual information, we have developed an on-line global approach with an expectation-maximization algorithm. We use camera calibration information that constrains the hypothetical space but does not require prior estimates of the number of people in the crowd, and remains robust in the face of partial occlusions (Figure 2).

For person reacquisition, we rely on general appearance matching and, when facial images can be captured, face recognition. Our appearance-matching method uses an articulated model-fitting approach to directly compare arms, torsos, and legs.[5] Such comparisons are based on signatures that capture illumination invariant color descriptions and structural descriptors in spite of transient phenomena such as folds and wrinkles (Figure 3). Face recognition efforts focus on acquiring and reconstructing the best possible representations from surveillance video. We have shown that active shape and appearance models, coupled with super-resolution, can be used to synthesize an enhanced facial image (Figure 4).[6] These can then be identified using standard commercial face-recognition engines.

The next step, beyond robust person detection, tracking, and identification, will be recognition of behaviors and events. Automatic activity analysis may be achievable, based on observed articulated motion. Analyzing interactions between individuals and objects may make possible detection of violent activity, medical emergencies, vandalism, and other significant events. Looking still further to the future, capture of physiological measurements, such as pulse and changes in facial temperature, may enable inferences regarding intention and aggression.

Author Information

Peter Tu, Fred Wheeler, Nils Krahnstoever, Thomas Sebastian, Jens Rittscher, Xiaoming Liu, Amitha Perera, and Gianfranco Doretto
GE Research
Niskayuna, NY

References
1. N. Krahnstoever and P. R. S. Mendonca, *Bayesian autocalibration for surveillance*, **Int'l Conf. on Computer Vision (ICCV05) II**, pp. 1858–1865, 2005.
2. N. Krahnstoever, P. Tu, T. Sebastian, A. Perera, and R. Collins, *Multi-View Detection and Tracking of Travelers and Luggage in Mass Transit Environments*, **IEEE Workshop on Performance Evaluation for Tracking and Surveillance (PETS)**, 2006.
3. N. Dalal and B. Triggs, *Histograms of Oriented Gradients for Human Detection*, **IEEE Conf. on Computer Vision and Pattern Recognition II**, pp. 886–893, 2005.

4. J. Rittscher, P. H. Tu, and N. Krahnstoever, *Simultaneous estimation of segmentation and shape*, **IEEE Conf. on Computer Vision and Pattern Recognition (CVPR 05) 2**, June 2005.

5. N. Gheissari, T. Sebastian, P. Tu, J. Rittscher, and R. Hartley, *Person Reidentification Using Spatiotemporal Appearance*, **IEEE Conf. on Computer Vision and Pattern Recognition (CVPR 06)**, 2006.

6. X. Liu, P. Tu, and F. Wheeler, *Face Model Fitting on Low Resolution Images*, **British Machine Vision Conf. (BMVC 06)**, September 4-7 2006.