Defense-Related Research Contributions and Interests

Richard J. Radke, Rensselaer Polytechnic Institute

My main scientific interests are in the field of computer vision, the study of how to design algorithms that can process and interpret images in the same ways that people do. In the worlds of government and industry, this research often is categorized as "intelligent video" or "video analytics". I am motivated by societally-important applications from both the surveillance/remote sensing and biomedical imaging communities. In this document, I summarize my research contributions that have direct relevance to DoD interests. PDF versions of accepted/published papers are available online at http://www.ecse.rpi.edu/~rjradke/pubs.htm. Hyperlinks to selected papers are also provided in red in the text below. This document is best viewed in color.

In 2007, I was chosen as one of the 12 members of the FY07 DARPA Computer Science Study Panel (CSSP). The CSSP program's goal is to familiarize the participants with DoD practices, challenges and risks, encouraging the development of cutting-edge computer science technologies of interest to the government. My involvement with the CSSP provided a unique insight into national security related computer science research in the DoD, and included briefings at a SECRET classified level at the National Military Command Center, the Naval Network Warfare Command, seven of the nine US combatant commands, NSA, CIA, DIA, NGA, and DHS, among many others. I also received close-range tours of current and future military combat systems at defense contractors and Army, Navy, and Air Force bases around the country. The CSSP is accompanied and mentored by a group of distinguished retired military and intelligence officers who put the briefings into context and introduce the participants to appropriate DoD contacts based on common research interests. Through the CSSP, I hold a SECRET security clearance.

Integrating EO and LiDAR Images. Three-dimensional range scanning technologies such as LiDAR (Light Detection and Ranging) will play a decisive role in addressing irregular strategic challenges such as urban and guerilla warfare. Integrating these new technologies into warfighting systems and evaluating their effectiveness will be critical for maintaining total battlespace awareness in complex environments. For example, early LiDAR systems have already proven useful for detecting improvised explosive devices in Iraq and imaging beneath dense forest canopies in South America. LiDAR technologies promise dramatic advancements in wide-area persistent surveillance, locating/tagging/tracking problems, and the generation of MASINT (Measurement and Signal Intelligence).

However, due to the high cost of the most sophisticated range sensors, little academic research exists on transformative high-end technologies such as portable flash LiDAR systems. While exciting new devices are being developed, substantial research challenges remain in determining how information from different types of range sensors can be fused, as well as integrated with ground-level digital images of the same environment. Our goal is to study and solve fundamental problems of multimodality **calibration**, **registration**, **data integration**, **model construction** and **information exploitation** that will allow these devices to achieve their full potential in C4ISR applications.

Our research group has several years' experience designing state-of-the-art computer vision algorithms for representing and aligning scanning LiDAR data, funded by the Army Intelligence and Security Command and DARPA. For example, our current algorithms can accurately register multiple high-resolution scans of the same scene from different perspectives into a

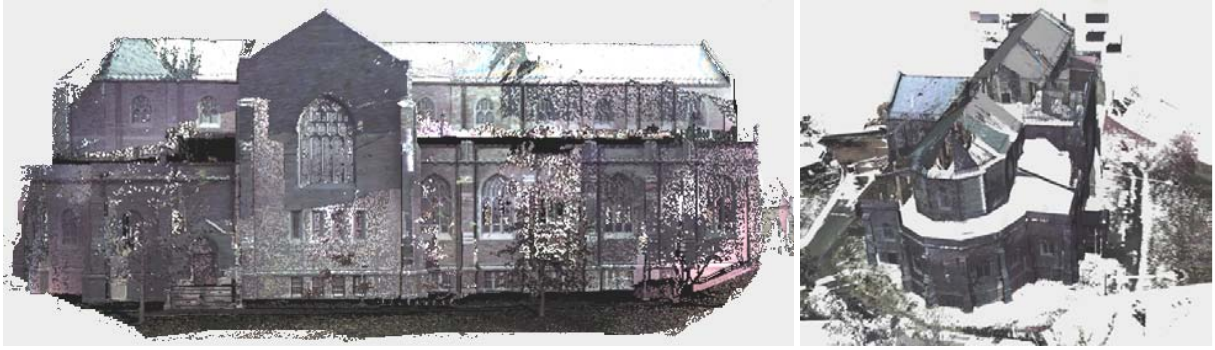
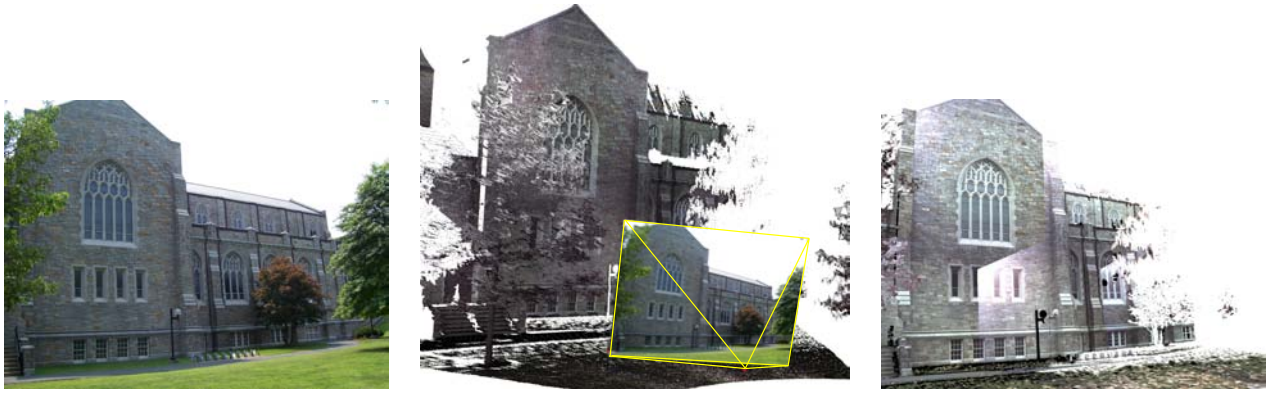


Figure 1: Our current research enables the accurate fusion of many scanning LiDAR images into a common coordinate system in a matter of minutes to automatically construct large-scale building models.



(a) (b) (c)

Figure 2: RPI algorithms for EO-LiDAR registration. (a) Digital (EO) image from an uncalibrated camera. (b) EO image automatically registered into coordinate system of LiDAR scan from different viewpoint. (c) LiDAR image synthetically re-rendered from estimated viewpoint of EO camera, showing correct localization.

common frame of reference in a matter of minutes to construct large-scale building models, as illustrated in Figure 1 [Smith et al. 2008]. We are also able to automatically determine the location and orientation with respect to a 3D LiDAR model from which a digital image was acquired, as illustrated in Figure 2- a kind of "visual GPS" system [Yang et al. 2007].

Under a DARPA-funded project, we are currently designing advanced algorithms for integrating EO data (e.g., visible imagery acquired from digital cameras), high-resolution scanning LiDAR data (e.g., acquired from ground-level surveys or possibly "air-breathing" aerial platforms), and flash LiDAR data (e.g., ground-level scans acquired by portable warfighter- or operative-carried sensors). A major focus is on algorithms for statistically well-justified hypothesis testing, enabling queries about the presence or absence of general classes/shapes/aspect ratios of objects or specific objects (e.g., CAD-type models). In all cases, the query is against a probabilistic data model that preserves and properly combines the uncertainty from all of the original component scans. This uncertainty is critical for making correct and principled decisions about scenes containing clutter, complete or partial occlusions (e.g., foliage or camouflage netting), or undersampled regions (e.g. from LiDAR rays nearly parallel to object surfaces), as well as providing analysts with reliable confidence-associated answers instead of yes/no results. One of our early approaches was presented in [Yapo et al. 2008].

Our most recent result is a procedure for verifying 3D object detections in LiDAR data using two complementary metrics [Doria et al. 2009, in review]. We take, as input, a triangulated mesh representation of a 3D model, one or more LiDAR scans of a scene, and an estimated transformation of the model into the scene. We wish to evaluate the hypothesis that the object is present at the given position. The verification procedure is independent of the method that produced the location hypothesis, so it is objective and unbiased in deciding if the position is indeed reasonable and correct. The advantage of the dual metric is that we can answer two questions simultaneously. We use a measure of *consistency* to determine if the object is in a position that makes sense physically. We use a measure of *confidence* to determine, if indeed the object is at a reasonable position, how much of it we have observed. The values produced by our consistency and confidence measures are both between 0 and 1, so they are easy to interpret for any data set. In contrast, the cost function value resulting from a

(a) (b)

Position	ICP Mean Distance	Consistency	Confidence
Correct	0.057	0.589	0.579
Incorrect	0.094	0.077	0.252

(c)

Figure 3: (a) Car model registered to correct position in scene. (b) Car model registered to incorrect position in scene. (c) The ICP mean distances are similar and difficult to interpret. Our new consistency and confidence measures are always in the range [0,1] and clearly disambiguate the correct and incorrect positions, even in the presence of substantial missing data.

Figure 4: (a) A LiDAR scan of a building, containing many "shadows" created by occlusion. (b) A close-up of a "shadow" in the LiDAR scan. (c) An image of the same environment from a different perspective. (d) Using the visual evidence from the image, new, highly-detailed 3D structure is generated.

registration algorithm such as Iterative Closest Points generally depends on the scale, sampling density, and parameterization of the problem, and is impossible to interpret as an absolute measure of match quality. Figure 3 shows that our metrics are much more discerning of the actual quality of a match, enabling a user to make a well-informed decision about the likelihood of an object's presence or the need for more scans of the scene to answer the question more conclusively.

LiDAR data can be difficult to automatically interpret due to the common artifact of missing data in the form of holes or "shadows" produced by occluding foreground surfaces during the scanning process. Our group has recently designed a LiDAR "inpainting" algorithm that uses a single additional image of the scene from a different perspective to automatically fill in convincing high-detail structure in these shadow regions [Becker et al. 2009, in review]. We first create an example database of image patch/3D geometry pairs from the non-occluded parts of the LiDAR scan, describing each uniform-scale region in 3D with a rotationally invariant image descriptor. We then iteratively select the best location on the current shadow boundary based on the amount of known supporting geometry, filling in blocks of 3D geometry using the best match from the example database and a local 3D registration. Figure 4 illustrates an example of our LiDAR inpainting algorithm, showing the generation of realistic, high-detail 3D geometry in a heavily occluded region. We believe the technique improves the utility of LiDAR scans for planning and situational awareness.

Our group owns a high-end scanning LiDAR system, which we use to acquire building models and test data. We hope to extend our current expertise to the investigation of portable flash LiDAR sensors for constructing and planning the acquisition of scene models from ground-level perspectives- a problem that to the best of our knowledge has not yet been studied by the research community. The critical challenges are the accurate alignment of a huge number of low-resolution flash LiDAR frames into a detailed three-dimensional model, and the integration of such data with EO images or scanning LiDAR models of the same scene. We are extremely interested in joint research partnerships with government/military agencies involving flash LIDAR data.

Distributed Computer Vision in Camera Networks. I am particularly interested in computer vision problems that occur in networks of a large number (tens to hundreds) of cameras dispersed throughout an environment. Such camera networks are common today in surveillance and security applications (e.g., subways and airports). Typically, the cameras are fixed in place, and the images from them are fed to a central location for processing (e.g., a powerful computer, or a cluster of them). In contrast, my research is motivated by computer vision problems that occur in wireless sensor networks. Such networks will be essential for 21st century military, environmental, and homeland security applications. For example, camera nodes may be dropped out of helicopters onto a battlefield, quickly placed by soldiers as they advance through terrain, or distributed throughout a hazardous environment by mobile search-and-rescue robots. However, such scenarios pose many challenges to traditional computer vision.

One novel aspect of my group's work is that we approach computer vision problems in a *distributed* manner. Each camera only communicates with its neighbors that see part of the same scene, and no camera has full information about the entire network. By contrast, much of the existing academic work on systems with many cameras assumes that they are all connected to a single computer. This is not practical or advantageous in the wireless sensor networking scenario. Our goal is to develop distributed algorithms that approach the performance of the best centralized algorithms, without having a single master processor. We also explicitly take into account that the camera nodes may be part of a power-constrained wireless sensor network, which means that the messages that the cameras send must be as efficient as possible, and only strictly necessary messages should be sent. These types of communication constraints are not usually a consideration in centralized computer

vision algorithms. Hence, my work lies at the intersection of conventional computer vision and sensor networking.

My students and I have investigated several problems in camera networks. First, how can a new camera entering the network identify the other cameras with which it shares visual overlap (i.e., views of the same scene from different perspectives)? We have designed a protocol in which each camera independently composes a short "digest" of distinctive features in its image and sends this around the network to see if any cameras see some of the same features [Cheng et al. 2007]. Our algorithm finds a set of good features and a way to describe them concisely, given a fixed length that each camera is allowed for its message. Second, how can the cameras that share visual overlap determine their position and orientation in 3D space, relative to each other and their environment? We showed how the camera network can be accurately calibrated using a method where each camera only communicates with its neighbors with which it has visual overlap, using a fast initial estimate [Devarajan et al. 2006] followed by a belief-propagation algorithm [Devarajan and Radke 2007] that ensures the distributed estimate is consistent (see Figure 5). A general overview of my group's research in this area is presented in [Devarajan et al. 2008]. This project was supported by an NSF Career award.

A future challenge is to take the algorithms we have developed and analyzed as computer simulations and make them practical for actual deployments of real networks of cameras, which will require teaming with software engineers, networking researchers, and sensor designers to build relatively cheap wireless cameras. A second longer-term goal is to develop efficient algorithms for higher-level applications on camera networks. Once the cameras all know where they and their neighbors are, we can start to investigate collaborative tasks such as tracking multiple objects as they move through an environment, detecting changes in the world, creating maps of terrain, or automatically detecting and relaying information about important events.

Video Analytics in Crowded Scenes. The automated analysis of images and video of crowded scenes is becoming increasingly important for understanding patterns of activity in urban areas. Crowds can arise in benign situations such as busy streets, sporting events, public celebrations, transit hubs, and retail environments, as well as in more dangerous or confrontational scenarios including political rallies, mobs, brawls, natural disasters, or mass evacuations. Our research in this area has two main thrusts: the automatic extraction of dominant patterns of motion in crowd video, and the detection and counting of multiple overlapping moving objects. In both cases, the raw input is simply a set of low-level feature trajectories automatically generated by tracking identifiable pixel patterns from frame to frame. The human visual system can easily infer small-scale and large-scale activity patterns from this basic information; we were inspired to craft algorithms that use the same input. A key advantage of our methods is that no strong models for the objects' shape, appearance, or motion are required, since such models would be very difficult to fit in highly crowded scenes.

Figure 5: (left) Several images of a scene captured by a simulated camera network. The goal of camera calibration is to determine the 3D positions and orientations of the cameras that took the images, using only the images themselves. (right) A subset of the distributed calibration result centered around a prominent church-like building in the scene. This overhead view shows the estimated camera positions in green and reconstructed points on the walls of the building in red.

(a) (b) (c)

Figure 6: In this scene, the dominant motion of the crowd is downwards, as correctly detected in (a). However, the algorithm also identified a substantial anomalous motion in (b)- a person who suddenly takes a U-turn to join a group walking downward. (c) Sample multi-object detection results from our coherent motion region algorithm. The red number indicates the aggregate count of independently moving objects (yellow boxes) detected in the video up to the given frame.

Our approach to detecting dominant motions [Cheriyadat and Radke 2008] begins by independently tracking low-level object features using an optical flow algorithm. Since our objective is to identify dominant, not individual, motions, we do not link, fix, or otherwise precondition these point tracks. Instead, we designed a clustering algorithm based on the similarity of point track segments measured using longest common subsequences. While many of the individual feature point tracks are unreliable, we show that we can automatically cluster them using an appropriate ordering and metric, and identify smooth dominant motions in a crowded scene by fitting polynomials to the cluster centers. Results on real video sequences demonstrate that the approach can successfully identify both dominant and anomalous motions in crowded scenes. These fully-automatic algorithms could be easily incorporated into distributed camera networks for autonomous scene analysis, such as detection of changes in the overall traffic pattern, or the anomalous motion of a single person in a crowd (Figure 6a,b.)

In related work, we proposed an object detection system that uses the locations of tracked low-level feature points as input, and produces a set of independent *coherent motion regions* as output [Cheriyadat et al. 2008]. We define a coherent motion region as a spatiotemporal subvolume fully containing a group of point tracks; ideally, a single moving object corresponds to a single coherent motion region. However, in the case of many similar moving objects that may overlap from the camera's perspective, many possible coherent motion regions exist. Therefore, we pose the multi-object detection problem as one of choosing a good set of disjoint coherent motion regions that represents the individual moving objects. This decision is based on a track similarity measure that evaluates the likelihood that all the trajectories within a coherent motion region arise from a single object, and is made using a greedy algorithm. Figure 6c shows sample video frames with our algorithm's results overlaid. This work is funded by the NSF and by the new DHS Center of Excellence for Awareness and Localization of Explosives-Related Threats (ALERT).

Change Detection and Understanding. Detecting regions of change in images of the same scene taken at different times is of widespread interest due to a large number of applications in diverse disciplines. Important applications of change detection include video surveillance, remote sensing, medical diagnosis and treatment, civil infrastructure, underwater sensing, and driver assistance systems. In 2005, I authored a systematic survey of image change detection algorithms, the first modern overview to cover both the surveillance and remote sensing literature [Radke et al. 2005]. This frequently-cited survey emphasizes the common processing steps and core algorithms used to attack the change detection problem, in an application-independent manner. I am particularly interested in extending image-based change detection algorithms to combined video and range imagery for defense applications.

Contact Information

Richard J. Radke
Associate Professor
Department of Electrical, Computer, and Systems Engineering
Rensselaer Polytechnic Institute
110 8th Street
Troy, NY 12180-3590

Phone: 518-276-6483
Fax: 518-276-8715
e-mail: rjradke@ecse.rpi.edu
http://www.ecse.rpi.edu/ rjradke