By John Edwards

Looking at Machine Vision

Before we can enter a world in which cars and trucks drive themselves, autonomous aircraft dot the skies, and robots pitch in to perform a virtually endless array of tasks, these systems will need to have a way of reliably and safely interacting with the surrounding world. Machine vision is the technology that will give future autonomous systems the ability to detect and react to various types of objects, terrains, and situations.

Signal processing lies at the heart of machine vision, opening ways of acquiring, processing, analyzing, and understanding images and other high-dimensional data from the real world. In multiple research areas, today's machine vision developers are pioneering systems that in years ahead promise to make life more faster, safer, healthier, and more convenient in an almost endless number of areas.

CUTTING THROUGH THE CLUTTER

Object recognition is one of the most pressing challenges facing computer vision researchers, since a robot or other type of machine manipulating something in the real world needs to do more than simply recognize an item—it also must be able to perceive the object's precise orientation.

To enhance the ability of robots to determine the orientation of specific objects, researcher Jared Glover (Figure 1) turned to a lesser-known and semineglected statistical construct known as the *Bingham distribution*. While a graduate student in the Massachusetts Institute of Technology's (MIT's) Department of Electrical Engineering and Computer Science, Glover and coresearcher Sanja Popovic developed a new robot vision algorithm, based on the Bingham distribution, that he says turned out to be 15% more accurate at identifying familiar objects in cluttered scenes than the best previous models. (Glover graduated MIT in May 2014. Popovic, also an MIT graduate, currently works at Google.)

Glover focused his research on a single basic question: How can a robot detect objects within a cluttered environment? "I started working on specific object detection, meaning my system was looking for objects that the robot already has a model of in its database," Glover says. "The robot knows the 3-D (three-dimensional) shape of the object it's looking for, it's just trying to find that shape in the clutter."

[FIG1] **As an MIT graduate student, Jared Glover developed a new robot vision algorithm based on the Bingham distribution. (Photo courtesy of Jared Glover.)**

SIGNAL PROCESSING LIES AT THE HEART OF MACHINE VISION, OPENING WAYS OF ACQUIRING, PROCESSING, ANALYZING, AND UNDERSTANDING IMAGES AND OTHER HIGH-DIMENSIONAL DATA FROM THE REAL WORLD.

In noisy and jumbled landscapes, accurate orientation detection hinges on precise alignments using multiple cues, such as 3-D point positions, surface normals, curvature directions, edges, and image features. Glover observed that other than brute force optimization, no existing alignment method existed that could merge all of this information together in a meaningful way.

The researcher identified the Bingham distribution as a useful tool because it enables an algorithm to squeeze more information out of each ambiguous, local feature. By connecting the Bingham distribution to the classical least-squares alignment problem, the researchers were easily able to fuse information from both position and orientation information into a principled, Bayesian alignment system that they called the *Bingham procrustean alignment*.

In his research, Glover used a Microsoft Kinect camera to identify locations in an image where color or depth values change abruptly—likely object edge locations. The work was then narrowed down to taking two sets of points—the model and the object—and determining whether one could be superimposed on the other.

Most algorithms, including Glover's, will make an initial immediate attempt at aligning the points. If both sets of points really do describe the same object, they can be quickly aligned by rotating one of them around the right axis. For any given pair of points—from the model and the

Digital Object Identifier 10.1109/MSP.2014.2343983
Date of publication: 15 October 2014

object—one can effectively determine the probability that rotating one point by a particular angle around a particular axis will align it with the other. The challenge is that the same rotation might also move other pairs of points farther away from each other. Glover, in his research, showed that the rotation probabilities for any given pair of points can be described as a Bingham distribution, which can then be merged into a single, comprehensive Bingham distribution.

Getting noise under control proved to be one of the researcher's major challenges "If you have noise, say, noisy estimates on the object's depth, and, if that noise is different from the first time you saw it to the second time you saw the object—because you see it from a different view or under different lighting—then the system might not have an accurate model for the noise, and so it will get confused," he says.

Nonetheless, in experiments using visual data about particularly cluttered scenes, the algorithm identified 73% of the objects in a given scene, compared to 64% from the best existing algorithm. With further research and sources of information, Glover believes the algorithm's performance can be improved even more.

"Besides increasing accuracy and robustness, the biggest challenge is relationship understanding," he explains. "If a robot can understand, for example, that the bowls are on top of each other, or things are touching each other in a certain way, or wrapped around each other, that's the kind of information that is going to be necessary for it to manipulate objects in the real world."

HAUPTMANN'S RESEARCH TEAM DEVELOPED MATHEMATICAL MODELS THAT LET THEM COMBINE CRITICAL INFORMATION, SUCH AS APPEARANCE, FACIAL RECOGNITION, AND MOTION TRAJECTORIES.

IDENTIFYING FACES IN THE CROWD

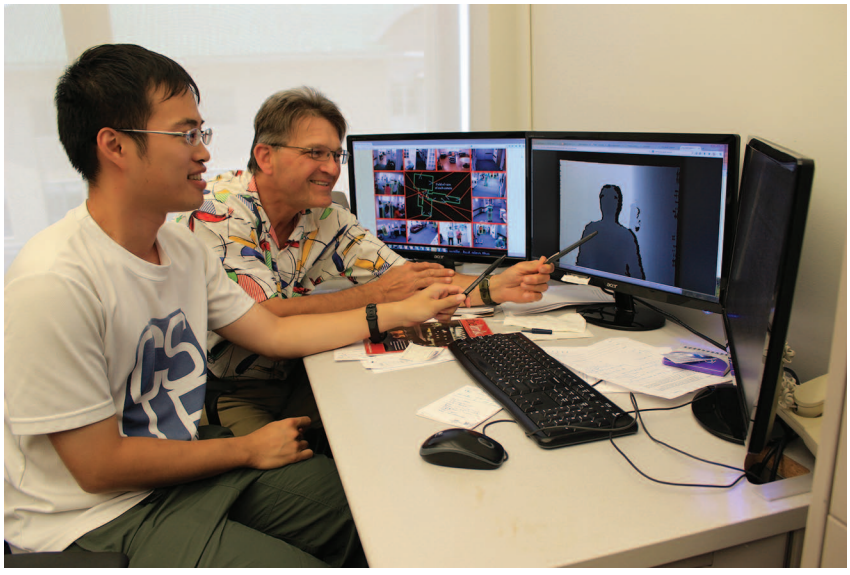
Multicamera, multiobject tracking has been an area of intense research for over a decade. Yet few automated techniques have been tested on objects located outside of well-controlled lab environments. To make tracking technology more useful in a potentially wide range of commercial and civic applications, researchers at Carnegie Mellon University recently developed an algorithm designed to track the locations of multiple individuals in complex, indoor settings via a network of video cameras.

Alexander Hauptmann (Figure 2), principal systems scientist in the Carnegie Mellon Computer Science Department, notes that developing an effective motion tracking system required overcoming a number of challenges. Something as apparently simple as tracking a person based on the color of the clothing worn proved to be frustratingly difficult because the apparel color can appear different to cameras in assorted locations due to lighting variations. Likewise, a camera's view of an individual can be blocked by people passing in hallways, by furniture or other stationary objects, or when someone enters a room or other area not covered by cameras. All of these situations, and others, make it necessary for individuals to be regularly reidentified by the system.

Hauptmann's research team developed mathematical models that let them combine critical information, such as appearance, facial recognition, and motion trajectories. Using all of this information is key to successful tracking, Hauptmann says, but facial recognition provided the greatest help. "The core tracking was a particle filter tracker, based on appearance," Hauptman remarks. When the researchers removed facial recognition data from the tracking, accuracy collapsed from 88% to 58%, not significantly better than existing tracking algorithms.

"The idea of particle filter tracking is that you don't commit to any one thing, so what you're tracking could be anywhere in the space, yet it's more likely to be here and less likely to be there," Hauptmann says. "So you always have these distributions of possible places for each particle that you're tracking and in the end you find that, overall, this is the most likely place for a particular person," he adds.

The algorithm's input consists of a set of person detection results at each time instant. "The person detection results from different camera views mapped to a common 3-D coordinate system using camera calibration and ground plane parameters provided," Hauptman says. Each person detection result is described by a color histogram. "Our algorithm's main task is to predict a label for each result," Hauptmann



[FIG2] **Alexander Hauptmann, principal systems scientist in the Carnegie Mellon Computer Science Department (right), and a student view motion-tracking system images. (Photo courtesy of Carnegie Mellon University.)**

explains. "To perform the prediction task, our algorithm incorporates two main innovative components, which are manifold learning in appearance space, with spatiotemporal constraints, and trajectory inference by nonnegative discretization.

The algorithm was able to automatically follow 13 individuals within a nursing home, with the residents' consent, despite the fact that people occasionally moved out of camera range. The researchers used 6 min of footage recorded by 15 cameras in a nursing home in 2005 to develop the algorithms and test the system. The team took advantage of multiple cues within the video, including trajectory, clothing color, person detection, and, most critically, facial recognition. "We thought it would be easy," Hauptmann said of multicamera tracking, "but it turned out to be incredibly challenging."

After working on the project for nearly a decade, Hauptmann notes that a series of relatively small technology and technique advancements can have as big an impact in an area like object tracking as a major breakthrough. "There's a big disconnect in computer vision between things that are published and things that work," he says. "What tends to get published are really novel ways of thinking about it—novel theories and novel algorithms."

But real life isn't quite that simple. Camera angle, for instance, can make a big difference in results. "Most research is done in a lab with a good camera position, so if a person eating is directly facing the camera, you've got high enough resolution to see their mouth moving, track points around the corners of their mouth and so on," Hauptmann says. "This approach is really impressive in that sort of laboratory situation, but when you take it into the real world it's a different story, and that's why this project took us so long."

After years of hard work, Hauptmann regards the project as a success. "Our algorithm exhibited the robust localization and tracking of persons-of-interest not only in outdoor scenes, but also in a complex indoor real-world nursing home environment," he says.

Hauptmann acknowledges, however, that the algorithm still has some limitations. "Our objective function does not have a spatial locality constraint on a trajectory," he says. "Therefore, our algorithm is not effective in very crowded sequences where each person wears the same color clothes." Another challenge is that optimization converging to a severe local optima, making initialization crucial. "Bad initialization may cause the performance to degrade," Hauptmann says.

While real-world deployment of the technology is still years away, Hauptman sees identification applications in venues beyond nursing homes, ranging from casinos to prisons. "We're still improving the accuracy," he says. "We're trying to get it so that we can easily apply it to other places."

SPEEDY CELL SORTING

Machine vision can also be used to recognize and differentiate objects as small as a biological cell. Researchers at the University of California, San Diego (UCSD), say that with the assistance of computer vision and hardware optimization they are now able to analyze and sort cells up to 38 times faster than with previous methods.

The approach, based on research originated at the University of California, Los Angeles (UCLA), improves imaging flow cytometry, a technique that uses a microscope-mounted camera to capture the morphological features of up to thousands of cells per second. The technology sorts cells into different categories, such as benign or malignant cells, based on their shape and structure. "The idea is, can we, at 50,000 frames per second, accurately identify each cell?" says Ryan Kastner, a UCSD professor of computer science (Figure 3).

Algorithms currently used take anywhere from 10 s to 0.4 s to analyze a single frame, making imaging flow cytometry far too slow for routine clinical use. The researchers' new approach promises to speed processing rates up to between 11.94 ms and 151.7 ms, depending on the hardware used. For enhanced performance, the team created a custom field-gate programmable array (FPGA). Low-range performance results, still significantly faster than currently achievable rates, were obtained by using an off-the-shelf graphics processing unit (GPU).

[FIG3] Ryan Kastner, a UCSD professor of computer science, leads a project aimed at speeding cell analysis and sorting. (Photo courtesy of UCSD.)

Four stages are necessary to perform the morphological analysis necessary for high-speed cell sorting: Blob Search, Image Interpolation and Adjustment, Find Center, and Coordinate Conversion/Radius Extraction. "Each module had to be carefully designed to achieve our performance targets," Kastner says. Yet reaching for maximum speed also required making some tradeoffs. "For example, at the end of the process, histogram equalization works better than image adjustment for contrast enhancement," Kastner explains. "Histogram equalization requires more complex processing leading to a lower throughput. Therefore, we sacrificed quality for performance."

The Blob Search module analyzes the images to detect the cell's area. The module then transforms the monochrome cell image into a binary digital image (only the pixels representing the cell are highlighted). The module then creates a histogram and crops a 20×20 pixel image around the cell.

To improve the fidelity of the analysis, the selected cell area from the Blob Search module is resized by a factor of ten. The Interpolation step also generates a higher contrast image by linearly adjusting the brightness level. The resized 200×200 image is input to the Find

(continued on page 117)

Sherman Karp (shermankarp@msn.com) is currently a private consultant. He was the principal scientist of the Defense Advanced Research Agency in the early 1980s. He also coauthored *Fundamentals of Electro-Optic Systems Design: Communications, Lidar, and Imaging*, (Cambridge Press, New York). He is a Life Fellow of the IEEE.

Joseph M. Aein (joe.aein0097@verizon.net) is now retired. He was employed at the Institute for Defense Analyses, Arlington, Virginia, followed by the RAND Corp., Santa Monica, California, where he participated in technology and systems evaluations in support of the U.S. Department of Defense. Among these were his participation in the Defense Advanced Research Agency Miniature GPS Receiver and GPS Guidance Package efforts. He is a Life Fellow of the IEEE.

REFERENCES

[1] B. Parkinson and S. T. Powers, "The origins of GPS: Part 1" *GPS World*, vol. 21, no. 5, pp. 30–41, May 2010.

[2] B. Parkinson and S. T. Powers, "The origins of GPS: Part 2" *GPS World*, vol. 21, no. 6, pp. 8–18, June 2010.

[3] S. Pace, G. P. Frost, I. Lachow, D. R. Frelinger, D. Fossum, D. Wassem, and M. M. Pinto. (1995). The global positioning system: Assessing national policies. Santa Monica, CA: RAND Corporation. [Online]. Available: http://www.rand.org/pubs/monograph_reports/MR614

[4] R. B. Langley, "The evolution of the GPS receiver," *GPS World*, vol. 11, no. 4, pp. 54–58, Apr. 2000.

[5] L. B. Stotts and J. Aein, "Status of DARPA guidance and control program," invited paper in *Proc. Cruise Missile Association Annu. Meeting and Symp.*, Apr. 1989, Washington, DC, pp. 205–228.

[6] L. B. Stotts, J. Aein, and N. Doherty, "Miniature GPS-based guidance technology," in *Proc. Guidance and Control Panel 48th Symp. Advances in Techniques and Technologies for Air Vehicle Navigation and Guidance*, May 1989, Lisbon, Portugal.

[7] L. B. Stotts and J. Aein, "Guidance technology useful for military communications," presented at Military Communications Conf. (MILCOM 89), Classified Session, 15–18 Oct. 1989.

[8] N. Dahlen, T. Caylor, and E. Goldner, "High performance GGP for multiple dual-use applications," in *Proc. 1996 National Technical Meeting of the Institute of Navigation*, Santa Monica, CA, Jan. 1996, pp. 63–74.

[9] N. J. Dahlen, T. L. Caylor, and E. L. Goldner, "Tightly coupled IFOG-based GPS guidance package," *J. Inst. Navigation*, vol. 43, no. 3, Fall 1996, pp. 257–272.

[10] C. Volk, J. Lincoln, and D. Tazartes, "Northrop Grumman's family of fiber-optic based inertial navigation systems," in *Proc. IEEE/ION PLANS 2006*, Apr. 25–27, 2006, San Diego, CA, pp. 382–389.

[11] G. Pavlath, "Fiber optic gyros past, present, and future," in *Proc. 22nd Int. Conf. Optical Fiber Sensors (OFS)*, Y. Liao, H. Ho, W. Jin, D. D. Sampson, R. Yamauchi, Y. Chung, K. Nakamura, and Y. Rao, Eds., *Proc. SPIE*, vol. 8421. Bellingham, WA: SPIE, 2012, p. 842102.

[SP]

[special REPORTS] (continued from page 11)

Center module. The outputs of this module are two images, the initial image interpolated and the linearly adjusted image after interpolation.

The Find Center module attempts to more accurately locate the cell's center. It finds the center of the cell by converting input images into binary image and counting the number of nonzero pixels in each row and column. The module processes the two output images from the Interpolation module and averages both to identify the center point. This is done to improve accuracy as specular noise can affect the results of either in-put. The Find Center module transforms these images into binary images by adaptively thresholding at different intensity values to separate the inner cell area and cell wall.

At the last stage, the system determines morphological properties of the cell using the interpolated image and its corresponding center point. It converts the resized image from Cartesian coordinates into polar coordinates. The darkest pixels found on a line from the cell center at each angle are considered the cell wall.

The researchers found that they obtained significantly faster performance with the FPGA than with GPU. The result didn't come as a total surprise, since FPGAs, unlike GPUs, can be custom-tailored to match the algorithm.

While designing the FPGA the researchers carefully studied each step and made changes designed to enhance efficiency and performance. Kastner notes that when mapping to custom hardware, it's important to balance algorithm complexity against result accuracy. "Algorithms incorporating a large number of decisions points, or that have to make multiple passes over the data, can lead to a slow and inefficient FPGA," he says.

Still, when correctly implemented, an FPGA can be used to perform operations at stunning speeds (the UCSD algorithm needs fewer than 500 µs to detect a cell and calculate its radius).

The researchers' ultimate goal is to analyze cell properties in real time and then use the information to accurately sort the cells. To achieve this capability, the sorting decision must be made in fewer than 10 ms. With the new approach promising sorting rates as low as 11.94 ms that target is now tantalizingly close.

Kastner is optimistic that the new technology will eventually be used in wide range of clinical applications. "This has to potential to lead to numerous breakthroughs," he says. "We are collaborating with UCLA and their industrial partners to commercialize the technology."

AUTHOR

John Edwards (jedwards@johnedwardsmedia.com) is a technology writer based in the Phoenix, Arizona, area.

[SP]