Research Statement

Matthew N. Dailey

February 11, 2004

Perspective

Building intelligent machines has been my lifelong ambition. Toward this end, my research draws on machine learning and our knowledge of the human visual system to build better computer vision systems. My ultimate goal is to answer two superficially unrelated questions:

- How do humans make sense of their visual sensations? and
- How do we build intelligent machines that improve peoples' lives?

These questions are actually complementary, because progress on one furthers the other. Visual perception is one of the most amazing human abilities and one of the most challenging near-term problems facing artificial intelligence. Clearly, not all intelligent systems require vision. But just as clearly, availability of robust, general machine vision would enable a host of applications that increase our safety, productivity, and leisure time.

Until very recently, progress in machine vision research was crippled by a severe shortage of compute power. On the 1 MIPS machines commonly available in the 1970s, simply *copying* the pixels in a 640×480 image took nearly a second. Only in recent years, with the advent of cheap 1000 MIPS machines, have some of the simplest computer vision algorithms become practical for real-time applications in uncontrolled environments.

But the times are changing at a dramatic pace. Due to exponential yearly growth in computing power, we are at the cusp of a revolution in many fields of computer science, especially computer vision and, more generally, machine intelligence. Technology pundits like Hans Moravec predict that machines will exceed human brain power in the 21st century.

Without new algorithms to harness this growth, however, the coming increases in raw compute power will be worthless. As a trivial example, we have no algorithm today that, given a set of images containing "chairs," is capable of identifying the chairs in a second set of novel images. Moreover, it is difficult to imagine how we would even begin to tackle the problem, even with unlimited compute power at our disposal. Despite this, any 3-year-old child can perform the same task effortlessly!

So although we are faced with an extremely difficult ill-posed computational problem, we have an example of a robust, fully-implemented, computational solution to that problem: the human visual system. For this reason, I advocate a multidisciplinary approach to vision. I believe that without an adequate understanding of the human visual system, we will never produce machines that see as well as we do. While technologists focusing on the development of practical systems need not be constrained to mimic the brain, it behooves them to pay close attention to emerging developments in neuroscience and psychology. My work demonstrates that there need not be a great divide between scientists inventing practical technology and scientists expanding our understanding of the brain.

My approach is straightforward. Regardless of the application, all of my work contains a common thread: I develop statistical models appropriate for the problem at hand, estimate the

parameters of said models from training data, and build decision-making systems informed by probabilistic inference under those models.

In the remaining sections of this statement, I first provide an overview of my prior research aimed at modeling certain aspects of the human visual system. I then describe my prior industrial research aimed at producing commercially successful intelligent systems. In the final section, I explain how I plan to build on all of this work to make further progress in computer vision, machine intelligence, and robotics.

If we are successful in this endeavor, society will accrue many benefits. In the near term, machine vision will enable robots to take care of our personal chores, enhance the abilities of industrial robots, and improve our security. In the long run, the possibilities are limitless.

Academic Research

My first major research project at UCSD, with my adviser, Garrison Cottrell, was motivated by the evidence from psychology and neuroscience that the human brain contains a small region apparently dedicated to face recognition. We wondered how the computational requirements of face recognition might prompt a developing brain to allocate specialized resources to face recognition without any innate bias to do so. We therefore explored the relationship between face recognition and non-face object recognition in neural networks trained on various categorization tasks. I built a machine vision system loosely modeled on what is currently known of the human visual system. My model consisted of a Gabor filter transform followed by dimensionality reduction and a "mixture of experts" neural network architecture, whose task was to classify images of peoples' faces and simple objects like coffee cups and books. I found that my system tended to develop a specialized face recognition "module" only when it was asked to *identify* individual faces (as opposed to classifying them as face or not-a-face) but lump objects into basic categories like "cup" (as opposed to identifying individual cups). Furthermore, my face recognition modules had a bias toward using the lower spatial frequency components of the input. We used these results to theorize that an interaction between task requirements (face identification) and the type of input best suited to the task (low spatial frequency input) is responsible for the development of a specialized face processing area in the brain. This work is documented in detail in Dailey and Cottrell (1999), and has led to a fruitful long-term collaboration with psychologists and neuroscientists in the U.S. and Canada on the computational basis of visual expertise.

More recent research with Cottrell tackled a different problem. I built on the work of Curtis Padgett to produce one of the first machine vision systems capable of accurately reading the emotional expressions on peoples' faces. The Ekman and Friesen "universal" facial expressions of emotion are specific patterns of muscle movements for happiness, sadness, fear, anger, surprise, and disgust. They have been tested for validity in dozens of cultures around the world. Our system, EMPATH, was the first to classify the Ekman and Friesen expressions of emotion as accurately as people do.

One of the main goals of EMPATH was to explain a mysterious phenomenon observed by psychologists investigating human visual perception: categorical perception of facial expressions. These researchers had found that humans apparently do not see facial expressions as graded mixtures of emotions, but as discrete categories. For example, when shown a face containing a mixture of surprise and happiness expressions, humans tend to see either happiness or surprise but not both. This observation supports the theory that the human visual system has evolved specifically to recognize human faces and very specific patterns of facial muscle movements. Suffice it to say that that is a troubling prospect for many theories in psychology and cognitive science. We therefore performed a series of experiments on EMPATH to determine how it would behave if it were a human subject in the categorical perception experiments. We found, surprisingly, that EMPATH exhibited the same behaviors as humans do. This was true even though the machine was simply trained to do a pattern recognition task by experience, without access to thousands of years of evolution. We were

thus able to show that "mysterious" behavior observed by psychologists is actually an emergent phenomenon in a system that learns by experience. Our early work on this project is documented in Cottrell, Dailey, Padgett, and Adolphs (2000), and a greatly improved effort is described in detail in Dailey, Cottrell, Padgett, and Adolphs (2002). More recent work using EMPATH to explain differences between Japanese and American subjects in interpreting facial expressions appears in my dissertation and will be published soon.

Industrial Research

In addition to my academic research, I have worked on several industrial projects in machine learning, vision, and robotics. Some of this work is proprietary and cannot be published, but other work is in the patent process and will be published in academic conferences and journals as well. At HNC Software, I developed outlier detection techniques for flagging potential financial fraud when positive examples of fraudulent activity are very rare compared to the total number of examples. At Netrologic, Inc., I developed image processing and pattern recognition algorithms for images of leaves of flowering plants. At Burning Glass Technologies, I developed new techniques for extracting information from semi-structured documents using hidden Markov models (HMMs) and a novel statistical model for document similarity. I also helped develop a statistical model for estimating the supply of and demand for workers in various industries and job categories using a large database of resumes and job descriptions.

My most recent work has been on the design and implementation of a 3D visual perception system for mobile robots at Vision Robotics Corp. (VRC). The company's current project is a vision-guided vacuum-cleaning robot capable of autonomously mapping its owner's home, storing that map and planning an optimal cleaning strategy, then autonomously cleaning the home periodically. Thus far, we have produced a prototype and solved all of the major algorithmic problems. We are now working with two major appliance manufacturers to take the prototype to market. My main contribution has been to solve the "concurrent localization and mapping" problem for mobile robots. This is an active area of research in the academic mobile robotics community, wherein a robot is required to map an unknown environment by exploring it. Mapping is generally an easy problem if the position of the robot is known precisely at each point in time. Likewise, if a robot has a precise map of its environment, determining its location is a straightforward statistical estimation problem. But doing both simultaneously in an uncertain environment is a much more difficult challenge. With Pearse Ffrench at VRC, I extended a 2D laser range finder-based approach developed by Sebastian Thrun and colleagues at CMU, generalizing the algorithm to use Hans Moravec's 3D grids and a 6 degree-of-freedom robot motion model. In order to improve the system's localization accuracy to the point needed for precise navigation in tightly constrained spaces, e.g. between pieces of furniture, I additionally developed a novel algorithm for learning a database of visual "landmarks" and a statistical model for position estimation combining evidence from the 3D grid with evidence from the landmark database.

In addition to solving the concurrent mapping and localization problem, I have also developed a novel approach to the classical stereo edge matching problem. Stereo matching means finding corresponding features (especially edges or corners) in two or more images of the same scene. Finding these correspondences enables a robot to extract extract 3D information from two dimensional images. By recasting the problem of matching edge pixels in two images as probabilistic inference in a Bayesian network, my system is able to combine evidence from multiple independent pixel matching heuristics in a principled manner, demonstrably outperforming traditional rule-based techniques. PRI is currently in the process of patenting all of these innovations, and publication will follow shortly thereafter.

Future Research Plans

In the next phase of my career I intend to continue my work in vision, learning, and robotics. My work will draw inspiration from the biological and cognitive sciences where appropriate. Its primary goal will be to create innovative machine intelligence and computer vision technology having potential to improve peoples' lives.

My approach to research will be straightforward: find practical applications with obvious commercial, humanitarian, or educational value and a need for machine intelligence, then tackle the interesting problems underlying those applications. Though the specific applications I work on will of course be influenced by the strategic goals of my institution, here I highlight just a few potential applications areas for my research.

One possible application is an autonomous mine-clearing robot. This application has obvious humanitarian value. In some sense, the task is quite similar to home cleaning, in that an unfamiliar environment must be carefully mapped then covered in detail. But the move from mostly flat indoor environments, rich in easy-to-extract landmarks, to rugged outdoor terrain, devoid of straight lines or other distinctive shapes, increases the difficulty of the problem dramatically.

Another possibility is a wheelchair for the disabled capable of autonomous navigation and obstacle avoidance. A standard motorized wheelchair outfitted with a stereo pair of cameras would form an excellent platform for research on vision, reinforcement learning, reactive control, and planning.

An autonomous vehicle is another ripe application for vision research. Self-guided automobiles hold the promise of reducing traffic congestion and accidents in overcrowded cities, and vision is an obvious choice as one of a vehicle's sensory modalities. The problem is made easier by accurate localization via GPS, preexisting maps, and a well-structured environment with traffic signals, lane markings, and so on. But there are many obvious challenges as well, including rapid response to unexpected events and continual fail-safe operation.

Although the RoboCup soccer-playing robot competition is often thought of as an educational game for high school and undergraduate students, fielding a top-notch robotic soccer team requires innovative artificial intelligence software for multi-agent coordinated goal-seeking behavior. (One of the perennial worldwide top contenders in the mid-sized league is a very well-funded, serious scientific effort by Philips Electronics.) Soccer-playing robots would benefit from fast, robust, and accurate machine vision systems. This makes RoboCup yet another ideal platform for research on machine vision and intelligence.

Again, these applications are only meant to be examples. In my view, the particular applications are secondary to the technology itself. However, I am particularly interested in developing technology that addresses the unique problems faced by the people of southeast Asia, and will seek out opportunities for multidisciplinary collaboration on those problems.

To summarize, my future research will be based firmly on statistical modeling and estimation in machine vision and robotics. My work will draw inspiration from diverse fields including biology and cognitive science. It will also exploit my demonstrated ability to deliver on large software projects. The result will be practical technology with the ultimate potential to make our lives safer, more secure, and more rewarding.