Research Statement

Andrew Gallagher (andrew.c.gallagher@gmail.com)

Abstract

My interests are primarily in the field of computer vision, defined broadly as "knowing what is where by looking." This field has close associations with image science, image processing, and machine learning. Image science is the study of image formation, including elements of optics and representations of color and light. Image processing deals with the representation and sampling of images and the tools for summarizing image information. Machine learning addresses issues related to representing probability distributions, typically over a large number of dimensions. Each of these areas are important components of computer vision systems that can, for example, recognize the face of a person.

I especially enjoy working on problems that cross the boundaries of multiple fields of research. Recently, my work is in the arena of using context (information outside of merely the pixel values corresponding to the object of interest) for recognizing people and their demographics. Graphical models are used to consider both the appearance and the contextual features. Not only has the research led to improved algorithms for identifying people (e.g. labeling gender and age), but interesting sociological observations also result. For example, from [5], we see that the average distance between a face and the nearest person in a group shot is 306 mm, a fact that is interesting to both vision researchers and anthropologists who study issues of personal space.

My ongoing research continues to focus on image understanding by using machine learning methods to train algorithms. I continue to be especially interested in interpreting images (both single and multiview) of people in social situations, and am currently working to identify social relationships, pose and actions. I feel that tremendous progress will be made in the realm of computer vision in the next decade, and I intend to be heavily involved in these achievements.

I am experienced in managing project teams and mentoring students, though generally on a short-term basis. I am interested in the more long-term mentoring opportunities and greater impact that a professor can have. I thoroughly enjoyed the teaching opportunities that I had at Carnegie Mellon, and feel teaching should be a component of the next period of my career. In addition, I am interested in the research freedom that is afforded to a professor. I believe that the goal of a professor should be to do excellent research and to educate the public, and I embrace this mission.

My strengths include working well with people, communicating well, creatively self-directing my project work and initiating new projects. In my time at Kodak, I have achieved the highest wage grade possible for a non-executive scientist (Senior Principal Research Scientist). I enjoy working on challenging problems. I have a solid background in intellectual property and patenting (over 100 patent applications filed and over 80 U.S. issued patents so far).

Background and Kodak Research Contributions.

I joined Kodak in 1996 with the Image Science Career Development Program. This two-year program included 16 weekly hours of training in image science (both chemical and digital) in addition to two yearly assignments in different areas of the Kodak Research Labs. I almost immediately became involved in Kodak's digital photofinishing efforts. My mentors were excellent researchers like Hsien Lee, Bob Goodwin, Ted Gindele, Majid Rabbani, and Bob Gray. In my first years at Kodak, my work focused on image enhancement (i.e. improving either the quality or accuracy of the image) and then my work transitioned into computer vision and related applications applications.

In my first few years at Kodak, most of my efforts were directed toward algorithm development for improving the visual quality of images. Along with my associates, I developed proprietary algorithms for:

- adaptive sharpening (adaptive to scene content and noise)
- dynamic range compression (for seeing detail in highlight and shadow)
- falloff compensation (for correction of flash and lens falloff)

(a) Original Image (b) Pedestal (c) Texture (d) Enhanced Image

Figure 1: An original image with high dynamic range has poor balance for both the background and the subject. Nonlinear filtering is applied to extract a pedestal signal and a texture signal. The pedestal signal approximates the scene illumination and is manipulated to adjust the dynamic range, while the texture signal is preserved to maintain image details. Finally, the signals are recombined to produce an enhanced image with a pleasing balance for both the scenery and the subject.

- perspective modification (based on vanishing point analysis)
- flesh-sensitive scene balance
- red-eye reduction

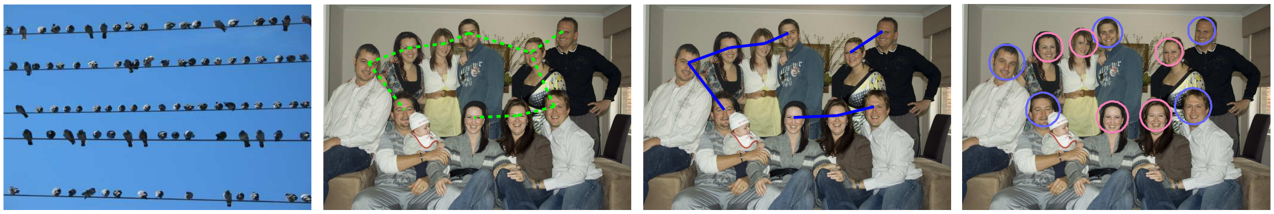
Describing one example, dynamic range compression is an approach for handling the problem that results from a scene having a dynamic range (i.e. ratio of the light portions to the dark portions of a scene) that is greater than can be properly displayed on output media. For example, the input image in Figure 1 shows an image where the shadowed subject is too dark and the sunlit background is too bright for both to be properly displayed. I developed proprietary non-linear filtering [7, 8] to separate the image into a "pedestal" signal containing mostly broad smooth areas approximately corresponding to the lighting in the image, and the "texture" signal, containing the detail in the scene. The lighting of the scene is manipulated to compress the lighting ratio, and the enhanced image is produced by recombining the modified pedestal and texture signals. A well-known academic publication with a similar intent to this work is by Durand and Dorsey [1]. In addition to designing the algorithms, I also quantified the robustness through blind experiments.

These algorithms became the core of the Kodak Perfect Touch processing, which resulted in a huge increase (millions of dollars of increased revenue) in Premium orders, even at the time when film sales were beginning their steep decline. Commercialization of these algorithms included algorithm development, software development, and design of psychovisual experiments (to ensure customer acceptance). These algorithms were adapted to operate on multiple color spaces (both digital camera images and scans of photographic film). This work was commercially successful, and although external publication was not encouraged, the novelty of this work was such that it resulted in over 50 U.S. Patents. Ted Gindele and I won the Kodak's Eastman Innovation Award in 2005. This award was roughly equivalent to "Kodak Inventor(s) of the Year" and included a $50,000 cash award to split between the honorees. Today, the ANSEL image enhancement algorithms in Kodak Perfect Touch are embedded into nearly all of Kodak's consumer products, including cameras, printers, frames, and on-line services.

While attending Rochester Institute of Technology from 1998-2000 for a Master's degree, I began to expand the scope of my interests from image enhancement to include the field of image understanding, with work on vanishing point detection, redeye correction and flesh detection. In 2005 I build an early version of the system for recognizing people in consumer image collections. Within a few months I had an impressive demo that was shown to Walter Mossberg of the Wall Street Journal, Martha Stewart, and Kodak CEO Antonio Perez (among others) and received positive external press.

Thesis Research.

In early 2006, my career took an exciting turn as I began studies towards a Ph.D. degree at Carnegie Mellon University in Pittsburgh under the direction of Prof. Tsuhan Chen. My course work focused on modern computer vision methods, and



(a) birds on a wire (b) nearest neighbors (c) minimum spanning tree (d) gender classification

Figure 2: Just as birds naturally space themselves on a wire (a), people position themselves in a group image. By extracting contextual features that capture the structure of the group of people, such as the nearest face (b) and minimum spanning tree (c), inferences can be made about the role of each person in the group. Among several applications, this context is used to determine the gender of the persons in the image (d).

machine learning with a heavy dose of probability. I particularly enjoyed the presentation of the material in a way that filled in many of the 'holes' in my background knowledge that occurred by randomly learning skills and ideas required for succeeding in research at Kodak.

My university research focused on using context (non-pixel information) for understanding images. Traditionally, in the field of face processing, algorithms had existed that made decisions about a face based on the appearance of that face alone (i.e. its pixel values). My thesis research is fundamentally different in that I introduced the idea of "social context" to better understand images of people. Social context is a context that describes people, their culture, and the social aspects of their interactions. By incorporating social context into graphical models for computer vision analysis, computers can bring the common sense and intuition as humans do when interpreting a scene. Further, I have demonstrated in a number of situations, analysis of social context allows one to use computer vision as a tool to learn more about people. I successfully defended my thesis "A Framework for Using Context to Understand Images of People" in May 2009. This work resulted in numerous publications in top-tier venues, and a best paper award at a conference workshop.

Clothing Cosegmentation: When looking for a particular person in a set of personal photographs, face recognition is a valuable tool. However, other contextual features such as clothing can provide identity-specific information as well. In [3], I show that considering segmented clothing as context can improve recognition accuracy. Further, recognition of people improves clothing segmentation.

First Names Carry Social Information: How can we find an image of a person named Mildred when we have no examples at all? It sounds like an impossible problem, until we recognize that first names carry a great deal of information about demographics. Specifically, Mildred is likely an older woman because the name was most popular as a name for baby girls in the early nineteenth century. In [4], we study the relationship between first names, age and gender, and show how to incorporate social statistics (in this case, the first dame data from the Social Security Administration's baby name database) into a real vision problem. Not only can we assign names to faces (e.g. Lisa versus Mildred), but knowledge of a person's first name improves on our appearance based age estimate. This paper has been required reading in vision courses at Brown University (CSCI 2951) and SUNY Stony Brook (CSE591).

People in Groups: In many social settings, images of groups of people such as shown in Figure 2 are captured. In [5], I explain how the structure of this group provides meaningful context for reasoning about individuals in the group, and about the structure of the scene as a whole. For example, men are more likely to stand on the edge of an image than women. Instead of treating each face independently from all others, contextual features are exploited that encapsulate the group structure locally (for each person in the group) and globally (the overall structure of the group). This "social context" allows one to accomplish a variety of tasks, such as such as demographic recognition, calculating scene and camera parameters, and even event recognition. This paper has been required reading in vision courses at Brown University (CSCI 2951) and University of Texas at Austin (CS395T).

Other Research Projects.

I have enjoyed working with a 20 co-authors on a range of topics related to image science, image processing, and image understanding, including:

Image Forensics: I developed an effective forensic tool for determining if an image has been interpolated [6]. With further development, the same algorithm can distinguish between digital camera captured images (which contain traces of interpolation from the color filter array on the image sensor) and computer rendered images (which generally do not). Still further, by looking for local evidence of resampling, forged regions within an image can effectively be recovered.

Sensor Design: In my early work on dynamic range compression, the images originated by scanning film negatives, which have a large dynamic range. With this in mind, I co-invented (with Kodak researchers Ted Gindele and Dave Nichols) a series of color filter array patterns for digital cameras that have clear photosites or photosites with staggered responses, instead of the typical red, green and blue photosites of the Bayer filter array pattern [10, 9]. As a result, the dynamic range of the sensor is improved without sacrificing resolution.

Markov Random Field Inference: More recently, several colleagues (Dhruv Batra, Devi Parikh and Tsuhan Chen) and I had an interesting theoretical result. We developed an improved approximation method for performing inference in undirected graphical models based on decomposing the underlying graph into a set of spanning outerplanar subgraphs. By solving this family of component problems and merging the solutions, we achieve better solutions than the previously best known algorithms.

Research Interests.

To date, my research has focused on image processing and computer vision, and the closely related areas of machine learning and graphical models. The mathematics and science involved are great in scope, and there are many fertile areas for research. I would like to further explore several directions, such as:

People in Images: I continue to be especially interested in interpreting images (both single and multiview) of people in social situations, and am currently working to identify social relationships, pose and actions. In particular, I am interested in extending my earlier work with pose models such as [2], and use this pose to better understand the actions and interactions in the image.

Planar Graph and Delaunay Image Representations: Many computer science algorithms begin by forming a graph over nodes that represent variables in a problem. In images, graphs can be formed by performing Delaunay triangulation over the interest points found in the image. Do these triangulations inherently provide a good representation of the image? Essentially, this research would be directed at ways of representing the spatial relationships between objects in scenes (e.g. cows stand on grass, not on golf courses).

Scenes over Time: I am interested in how scenes change over time. By looking at millions of images on the Internet, we can gain an understanding of the static and movable portions of the scene. In turn, this information can be used in editing algorithms (e.g. to remove the tourists from the landmark).

Decision Analysis Research: I am also interested in research involving decision-making and probability that is not directly involved with images or vision. For example, I have begun analyzing a dataset of football plays from the National Football League in an effort to determine the optimal decisions (punt, field goal attempt, or attempt first down) given the field position. Not surprisingly, many of the tools (probability models, filtering, graph theory) that are useful in computer vision and image processing are useful in these other areas as well.

Funding and Budgets.

In my role Kodak I also oversaw several research grants to Cornell University, working with Professors Tsuhan Chen and Ashutosh Saxena. With this experience, combined with my 14 years in industry, I feel I have a good understanding of what research is of interest to industry.

References

[1] F. Durand and J. Dorsey. Fast bilateral filtering for the display of high-dynamic-range images. In *SIGGRAPH*, 2002.

[2] V. Ferrari, M. Marin-Jimenez, and A. Zisserman. Pose search: Retrieving people using their pose. *Proc. CVPR*, 2009.

[3] A. Gallagher and T. Chen. Clothing cosegmentation for recognizing people. In *Proc. CVPR*, 2008.

[4] A. Gallagher and T. Chen. Estimating age, gender, and identity using first name priors. In *Proc. CVPR*, 2008.

[5] A. Gallagher and T. Chen. Understanding images of groups of people. In *Proc. CVPR*, 2009.

[6] A. C. Gallagher. Detection of linear and cubic interpolation in JPEG compressed images. In *CRV*, 2005.

[7] A. C. Gallagher and E. B. Gindele. Method for preserving image detail when adjusting the contrast of a digital image, U.S. Patent No. 6,317,521, 1998.

[8] A. C. Gallagher and E. B. Gindele. Method of spatially filtering a digital image using chrominance information, U.S. Patent No. 7,280,703, 2002.

[9] A. C. Gallagher and D. Nichols. Method and apparauts to extend the effective dynamic range of an image sensor device, U.S. Patent No. 6,909,461, 2002.

[10] E. B. Gindele and A. C. Gallagher. Sparsely sampled image sensing device with color and luminance photosites, U.S. Patent No. 6,476,865, 2001.