Research Statement

Hector J. Santos-Villalobos *

My research interests span the areas of signal and image processing, and computer engineering. More precisely, I desire to investigate within the topics of digital color imaging, pattern recognition (e.g. biometrics, biomedical imaging), storage and retrieval of high-dimensional data (i.e. shape recognition), usability engineering, semantic web, and autonomous systems. In addition, I long to be part of outreach programs, and mentor students from underrepresented populations.

During my first three years at Purdue University, I worked on two projects sponsored by a leading printer manufacturer: the Print Quality Troubleshooting (PQTS) tool, and the Color Printing Workflow Troubleshooting (CPWT) tool. The PQTS is a web-based application developed to enable customers to self-solve many of their print quality issues. This project in particular allowed me to experience teamwork and interdisciplinary collaboration. In addition, it was gratifying and stimulating, because the tool currently serves thousands of customers. For the second project, I had to identify and propose solutions to critical flaws in the web-based services offered by the sponsor. Based on my assessment, we recommended the development of the CPWT. Such tool is analogous to the PQTS but aimed at solving more complex issues. Both tools, the PQTS and the CPWT, established groundbreaking design guidelines for web-based troubleshooting applications. At present, I am working on a project sponsored, in part, by a National Science Foundation (NSF) grant (CCF-0728929). The motivation was to develop a comparison method for 2-D objects, which shapes can be modeled as Gaussian mixture (GM) densities. More precisely, my contribution to this project was a novel method to estimate the likelihood that two GMs belong to the same distribution. The originality of the method is that instead of comparing the mixtures directly, we compare distance samples drawn from the mixtures (i.e. a distance sample is the pairwise distance between two point samples drawn from a GM). As of right now, we are using this method for developing tools for fingerprint and halftone pattern matching, but implications for further applications are virtually endless.

My future research plans aim to: 1) Develop of new recognition paradigms that utilize my GM comparison method for fingerprint matching, printing and image forensics, image-context matching, and so on. 2) Further study efficient methods for dimensionality reduction and projections. 3) Research the use of intelligent systems for an efficient representation, accessibility, and association of scientific knowledge (e.g. evidence-based treatment tools for medical care of humans and animals). 4) Develop learning paradigms for knowledge-based systems that mimic human learning. Such research has the potential to get funded by private companies and public institutions like the NSF, and the National Security Agency (NSA).

Current and recent research

My research experience began in 2004 as a master's student at the University of Puerto Rico. During this time, Purdue University and University of Puerto Rico had joined efforts to design and develop a fully automated digital printing workflow. My piece of the puzzle was to develop a knowledge-based system that identifies style-dependent inconsistencies generated by variable content in digital documents. In 2005, my digital printing and knowledge-based systems expertise led me to become part of the Electronic Imaging Systems Laboratory (EISL) at Purdue University as doctoral student. As part of an interdisciplinary group, I contributed to the PQTS with ideas to improve the usability of the tool, and coded realistic simulations of

*Contact information: Tel. (787) 466-4460, E-mail: hector.j.santos@gmail.com

print quality issues. In 2007 I began working with the CPWT which aimed to help customers self-solve color issues. In contrast to print quality issues, which are generated at the printer color issues arise from misuse or malfunction of any of the components in the printing workflow (i.e. operating system, imaging application, printer driver, printer) or simple limitations of the printer. I was responsible for conceiving a solution, which included an assessment of the digital printing workflow in personal computers, apply knowledge elicitation methods to capture issue resolution strategies from a color specialist, and the design and development of the actual web application.

Currently, I am working on a theory-centered project, motivated by the need to develop an efficient comparison method for 2-D shapes. More precisely, we wanted to compare shapes represented with planar GM densities specifically asking the question *"Can we estimate the similarity between the shapes of two planar GMs efficiently?"* Previous research proved that the distribution of distances is a lossless representation of generic GMs. This noteworthy finding in conjunction with the fact that distances are invariant to rotations and translations, allowed us to utilize distance distributions to avoid the typical problem of having to align 2-D shapes before comparing them. Then, we used a goodness-of-fit test to estimate the likelihood that the distribution of distances of two GMs was indeed the same distribution. After an assessment of several goodness-of-fit tests, we selected the Kolmogorov-Smirnov test, which evaluates the distance samples directly and performs with a linear time complexity. We found that, in some instances, the set of distance samples did not appropriately represent the distance distribution of a GM. To mitigate this, we generated several distance-sets from both shapes and measure their similarity. Then, the median of all measurements is the estimate for shape similarity. My assessment of the method showed a robust recognition performance. This is an efficient shape comparison method that can be utilized by low computational power devices (e.g. smartphones) or for a quick search of large shape databases (i.e. fingerprint matching for a national security system).

Research plans

The ingredients of my research philosophy are hard work, dedication, adaptability, and collaboration. My success in research is directly related to 1) the effort allocated to understanding the problem and defining solutions; 2) the dedication to comprehending and embracing the needs of the sponsors and the vision of the leadership; 3) the adaptability to changes in research priorities and unexpected setbacks; and 4) the interaction with other experts to complement and enhance current ideas and approaches.

My preferred modus operandi is to break up a research project into a few short-term projects with clear objectives and deliverables. Also, I plan research around the intersection of problems and techniques. I believe that if we focus our research plans solely on a technique, we endanger our ability to do research beyond the development of that particular technique. Similarly, if we only focus our research around a well-defined problem, it becomes difficult to contribute as more and more researchers investigate the problem. My research methodology for shape matching is an example of the combined problem-technique approach in that it can be extended to several issues like fingerprint matching, printer and scanner forensics, image context recognition, and so on.

My approach will be to search for new areas of research where I can transfer previous knowledge and expertise along with the further development of new applications that employ my current classification

technique. For example, I would like to investigate new computationally efficient methods for dimensionality reduction and projections. Commercial electronic devices with low computational power crave this type of technology. Also, applications with a colossal demand for data processing (e.g. a nationwide face identification system) require efficient algorithms to reduce the load of their computations. In addition, a topic of interest to the NSA is the development of a web-based test bed for shape representation and classification methodologies. Such framework will facilitate the integration and benchmarking of new recognition technologies. I would also contribute with knowledge-based systems that simplify the sharing of knowledge between the scientific community. For example, health professionals urge for tools that accurately retrieve the latest research findings about a drug for evidence-based treatment. The Cyber-Enabled Discovery and Innovation program from the NSF is a potential sponsor. Another important piece of my research plan is the development of autonomous intelligent systems. I am fascinated by the fact that some basic human behaviors take years to develop, yet we often want or expect artificial intelligence programs to perform at the level of a human in minutes or days. One of my long-term research goals is to develop an intelligent system that mimics the slow learning process of the human brain. Such project could be funded by the NSF Robust Intelligence Program.

Final thoughts

Purdue University has given me the opportunity to expand my research experience by being part of multiple interdisciplinary research projects sponsored by both private and public institutions. As a result, not only have I learned to design well-defined research projects that exhibit granting or contract potential, but also have developed relationships that could result in potential research collaborations with both industry and academia. I believe the benefits of working along side with partners from both the academic and industry realms will expand beyond the research community and will positively impact students, the university, and the general community. In addition, I feel committed in engaging in research activities that promote the development of students and communities from underrepresented populations. Beyond any doubt, my passion and vocation is to help develop well prepared students and to contribute to science with creative and valuable ideas.