Introduction to the Special Section on Real-World Face Recognition

Gang Hua, *Senior Member, IEEE*, Ming-Hsuan Yang, *Senior Member, IEEE*, Erik Learned-Miller, *Member, IEEE*, Yi Ma, *Senior Member, IEEE*, Matthew Turk, *Senior Member, IEEE*, David J. Kriegman, *Senior Member, IEEE*, and Thomas S. Huang, *Fellow, IEEE*

1 Introduction

In recent years, there has been an emerging demand for robust face recognition algorithms that are able to deal with real-world face images. This is largely due to two factors. First, consumers have shown an increasing desire to annotate or tag their digital photos to facilitate organization, access, and online sharing of their personal albums. Industrial responses to this consumer desire can be exemplifed by successful commercial face recognition systems included in applications and web sites such as Google Picasa, Windows Live Photo Gallery, Apple iPhoto, face.com, PolarRose, etc. Second, the growing applications in public security also call for robust face recognition technologies that can identify individuals from surveillance cameras in uncontrolled situations.

In consumer digital imaging, face recognition must contend with uncontrolled lighting, large pose variations, a range of facial expressions, make-up, changes in facial hair, eyewear, weight gain, aging, and partial occlusions. Similarly, in scenarios such as visual surveillance, videos are often acquired in uncontrolled situations or from moving cameras. These factors have been the focus of face recognition research for decades, but they are still not well resolved.

In addition to these market forces, face recognition also represents an ongoing set of key scientific challenges. How can we match or exceed the performance of humans on face on many real-world face recognition tasks (e.g., recognition of people known to the human viewer)? How can we learn a good model of a face from a small number of examples? How can we achieve the level of robustness exhibited by human face recognition? These questions and new applications for the technology promise to keep this area active for the forseeable future.

Standard face recognition systems often start with a set of labeled *gallery faces*. When a new *probe image* is provided, it is matched against the gallery faces to be recognized as a known face or rejected. In a well-controlled setting, face images can be carefully captured for both gallery and probe faces. In a moderately controlled setting, we may have quality control over either the gallery faces or the probe faces, but not both. In an uncontrolled setting, we lose control of both.

Face recognition in well-controlled settings has been extensively studied and is relatively mature. Earlier face recognition methods often directly appled pattern recognition and machine learning techniques on informative face features and are only effective when the probe and gallery images are frontal. More recently, to achieve higher recognition performance, many works have started to consider more precise geometric, shape, lighting, and reflectance models of faces. Notwithstanding the successes of these techniques, there remains much room for improvement of face recognition in real-world scenarios since, overall, much less attention has been paid to these less controlled settings.

Two general misconceptions are that face recognition is a solved problem and, on the other hand, that the uncontrolled scenarios are too difficult to address in practice. Neither is true in our opinion. A primary purpose of this special section is to deliver the following message to the community: Although significant progress has been made in the last few decades, there remain plenty of challenges and opportunities ahead.

In the consumer digital imaging domain, practical face annotation systems are emerging based on existing face recognition technologies. This brings human factors in the assistanted annotation system since good user interface (UI) and user experience (UX) design are essential in order to compensate for possible failures of the face recognition algorithm. Last but not least, in many of these applications

- *G. Hua is with the IBM T.J. Watson Research Center, Hawthorne, NY 10532. E-mail: ganghua@gmail.com.*
- *M.-H. Yang is with the Department of Electrical Engineering and Computer Science, University of California, Merced, CA 95344. E-mail: mhyang@ucmerced.edu.*
- *E. Learned-Miller is with the Computer Science Department, University of Massachusetts, Amherst, MA 01003. E-mail: elm@cs.umass.edu.*
- *Y. Ma is with Microsoft Research Asia, Beijing, China, and is on leave from the Electrical and Computer Engineering Department, University of Illinois at Urbana-Champaign, Urbana, IL 61801. E-mail: mayi@microsoft.com.*
- *M. Turk is with the Computer Science Department, University of California, Santa Barbara, CA 93106. E-mail: mturk@cs.ucsb.edu.*
- *D.J. Kriegman is with the Computer Science and Engineering Department, University of California, San Diego, La Jolla, CA 93093. E-mail: kriegman@cs.ucsd.edu.*
- *T.S. Huang is with the Electrical and Computer Engineering Department, University of Illinois at Urbana-Champaign, Urbana, IL 61801. E-mail: huang@ifp.uiuc.edu.*

For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org.

0162-8828/11/$26.00 © 2011 IEEE Published by the IEEE Computer Society

there may be rich additional contextual information and meta-data that one can leverage to improve face recognition. A multidisciplinary exploration may be required to deliver real working systems.

2 Review Process

The motivations for organizing this special section were to better address the challenges of face recognition in real-world scenarios, to promote systematic research and evaluation of promising methods and systems, to provide a snapshot of where we are in this domain, and to stimulate discussion about future directions. We solicited original contributions of research on all aspects of real-world face recognition, including:

- the design of robust face similarity features and metrics,
- robust face clustering and sorting algorithms,
- novel user interaction models and face recognition algorithms for face tagging,
- novel applications of web face recognition,
- novel computational paradigms for face recognition,
- challenges in large scale face recognition tasks, e.g., on the Internet,
- face recognition with contextual information,
- face recognition benchmarks and evaluation methodology for moderately controlled or uncontrolled envi-ronments, and
- video face recognition.

We received 42 original submissions, four of which were rejected without review; the other 38 papers entered the normal review process. Each paper was reviewed by three reviewers who are experts in their respective topics. More than 100 expert reviewers have been involved in the review process.

The papers were equally distributed among the guest editors. A final decision for each paper was made by at least two guest editors assigned to it. To avoid conflict of interest, no guest editor submitted any papers to this special section.

3 The Papers

Six papers were accepted through the rigorous review process, for an overall acceptance rate of 14.3 percent. The accepted papers can be placed into three categories: face recognition in real-world watch-list visual surveillance systems, 3D modeling for pose variant face recognition, and design of robust face similarity features and metrics for face recognition in consumer photos. In the following, we briefly summarize the papers in each category.

In "Toward Development of a Face Recognition System for Watch-List Surveillance," Kamgar-Parsi et al. examine the problem of designing a face recognition system for a watch-list visual surveillance system where a small set of people needs to be identified from a large number of people passing through surveillance cameras. Their approach is to use view morphing to automatically generate borderline faces to define the face space of a person, and then to train classifiers based on the borderline faces from each person in the watch-list. The method attacks a real-world face recognition problem with an interesting and solid approach.

3D face recognition has been regarded as a natural solution to pose variation. In "Using Facial Symmetry to Handle Pose Variations in Real-World 3D Face Recognition," Passalis et al. propose using facial symmetry to handle pose variation in 3D face recognition, while in "Unconstrained Pose Invariant Face Recognition Using 3D Generic Elastic Models," Prabhu et al. propose a generic 3D elastic model for pose invariant face recognition. Both are plausible approaches for using 3D information to assist in face recognition under large pose variations. While historically 3D face recognition has been criticized for lack of real-world 3D sensory cameras, this issue may be resolved in the future with inexpensive 3D sensors as evidenced by the PrimeSense sensor used in the Xbox Kinect from Microsoft.

The other three accepted papers all deal with face recognition in photos and images in the wild. In "Describable Visual Attributes for Face Verification and Image Search," Kumar et al. present a face verification algorithm based on a representation that uses a set of describable attributes. Their classifier achieves very good results on two publicly available benchmarks, namely Labeled Faces in the Wild (LFW) and the Public Figures (PubFig) data set. Since their method was first published at ICCV '09, there has been a lot of work using attribute-based representations for various visual recognition problems beyond face recognition.

In "Effective Unconstrained Face Recognition by Combining Multiple Descriptors and Learned Background Statistics," Wolf et al. propose an approach for face verification in the wild that combines multiple descriptors with learned statistics from background context. Its face verification accuracy ranked first on the LFW benchmark. In "Scalable Face Image Retrieval with Identity-Based Quantization and Multireference Reranking," Wu et al. present a scalable face image retrieval system using identity-based quantization to build a visual representation and multiple references for reranking. It builds a good foundation to tackle the problem of searching for face images over the internet image corpus.

The face recognition research community has built a variety of solid benchmarks to evaluate different algorithms. It is vital for researchers to leverage these databases to conduct solid and convincing experimental validation and compare with the state-of-the-art. Yet there also comes a time when performance on a benchmark reaches ceiling performance or methods become overengineered for nuances of a data set, and modest performance gains may be indicative of overfitting.

Alternatively, some new works or operational scenarios may push the envelope in directions that are not well represented with existing benchmarks; in such cases, authors may need to develop alternative benchmarks and justify this need in subsequent publications. Interestingly, real-world face recognition methods that achieve state-of-the-art performance on data sets like LFW may actually perform worse on constrained, frontal data sets like FERET. We should not be surprised by this, and we should embrace methods for where they are effective.

4 CONCLUSION

Through the editorial process of this special section, it has been our observation that the joint efforts of the whole face recognition research community have made many applications of real-world face recognition achievable, but there are still many challenges to address and opportunities to explore.

Claims that face recognition is a solved problem are overly bold and optimistic. On the contrary, claims that face recognition in real-world scenarios is next to impossible are simply too pessimistic, given the success of the aforementioned commercial face recognition systems. We hope this special section on Real-World Face Recognition will serve as a reference point toward an objective evaluation of the community's progress on face recognition research.

ACKNOWLEDGMENTS

We thank all authors for their enthusiastic contributions to this special section. We also thank the 100+ reviewers for their thoughtful reviews of the submissions, which were vital to ensure the quality of this special section. We greatly appreciate Editor-in-Chief Professor Ramin Zabih's timely help in resolving many issues that occurred during the review process. Last but not least, we thank Andy Morton for his help in all aspect of the editorial administration.

Gang Hua
Ming-Hsuan Yang
Erik Learned-Miller
Yi Ma
Matthew Turk
David J. Kriegman
Thomas S. Huang
Guest Editors

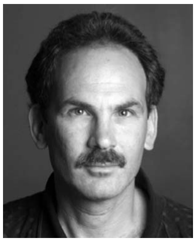
Gang Hua was enrolled in the Special Class for the Gifted Young of Xian Jiaotong University (XJTU) in 1994 and received the BS degree in automatic control engineering from XJTU in 1999. He received the MS degree in control science and engineering in 2002 from XJTU, and the PhD degree from the Department of Electrical and Computer Engineering at Northwestern University in 2006. He is currently a research staff member at the IBM Research T.J. Watson Center. Before that, he was a senior researcher at Nokia Research Center, Hollywood, from 2009 to 2010, and a scientist at Microsoft Live Labs Research from 2006 to 2009. He is an associate editor of the *IEEE Transactions on Image Processing* and *IAPR Journal of Machine Vision and Applications* and a guest editor of the *IEEE Transactions on Pattern Analysis and Machine Intelligence* and the *International Journal on Computer Vision*. He is an area chair of the IEEE International Conference on Computer Vision, 2011, an area chair of ACM Multimedia 2011, and a Workshops and Proceedings Chair of the IEEE Conference on Face and Gesture Recognition 2011. He is the author of more than 50 peer reviewed publications in prestigious international journals and conferences. As of August 2011, he holds three US patents and has 17 more patents pending. He is a senior member of the IEEE and a member of the ACM.

Ming-Hsuan Yang received the PhD degree in computer science from the University of Illinois at Urbana-Champaign in 2000. He studied at the National Tsing-Hua University, Taiwan, the University of Southern California, and the University of Texas at Austin. He is an assistant professor in electrical engineering and computer science at the University of California, Merced. He was a senior research scientist at the Honda Research Institute working on vision problems related to humanoid robots. He received the Ray Ozzie fellowship in 1999 and the Google faculty award in 2009. He coauthored the book *Face Detection and Gesture Recognition for Human-Computer Interaction* (Kluwer Academic, 2001) and edited a special issue on face recognition for *Computer Vision and Image Understanding* in 2003. He served as an area chair for the IEEE Conference on Computer Vision and Pattern Recognition, Asian Conference for the IEEE International Conference on Computer Vision, and the AAAI conference on Artificial Intelligence in 2011. He is an associate editor for the *IEEE Transactions on Pattern Analysis and Machine Intelligence* and *Image and Vision Computing*. He is a senior member of the IEEE and the ACM.

Erik Learned-Miller (previously Erik G. Miller) is an associate professor of computer science at the University of Massachusetts, Amherst, where he joined the faculty in 2004. He spent two years as a postdoctoral researcher at the University of California, Berkeley, in the Computer Science Division. Learned-Miller received the BA degree in psychology from Yale University in 1988. In 1989, he cofounded CORITechs, Inc., where he and cofounder Rob Riker developed the second FDA cleared system for image-guided neurosurgery. He worked for Nomos Corpo-ration, Pittsburgh, Pennsylvania, for two years as the manager of neurosurgical product engineering. He received the Master of Science (1997) and PhD (2002) degrees from the Massachusetts Institute of Technology, both in electrical engineering and computer science. In 2006, he received a US National Science Foundation CAREER award for his work in computer vision and machine learning. He is a member of the IEEE.

Yi Ma received the bachelor's degree in automation and applied mathematics from Tsinghua University, Beijing, China, in 1995. He received a master's degree in electrical engineering and computer sciences (EECS) in 1997, a second master's degree in mathematics in 2000, and the PhD degree in EECS in 2000, all from the University of California, Berkeley. He is currently a professor in the Department of Electrical and Computer Engineering at the University of Illinois at Urbana-Champaign and, since January 2009, has also served as research manager for the Visual Computing Group at Microsoft Research Asia, Beijing, China. He received the David Marr Prize in 1999, a US National Science Foundation CAREER award in 2004, and a US Office of Naval Research Young Investigator award in 2005. He is an associate editor of the I*EEE Transactions on Pattern Analysis and Machine Intelligence* and the *International Journal of Computer Vision*. He is a senior member of the IEEE.

Matthew Turk received the BS degree from Virginia Tech, the MS degree from Carnegie Mellon University, and the PhD degree from the Massachusetts Institute of Technology. He worked for Martin Marietta Denver Aerospace from 1984 to 1987 on vision for autonomous robot navigation. In 1992, he moved to Grenoble, France, as a visiting researcher at LIFIA/ENSIMAG, then took a position at Teleos Research in 1993. In 1994, he joined Microsoft Research as a founding member of the Vision Technology Group. In 2000, he joined the faculty of the University of California, Santa Barbara, where he is now a full professor in the Computer Science Department and former chair of the Media Arts and Technology Graduate Program. He co-directs the UCSB Four Eyes Lab, where the research focus is on the "four I's" of Imaging, Interaction, and Innovative Interfaces. He is a founding member and former chair of the advisory board for the International Conference on Multimodal Interfaces and on the editorial board of the *Journal of Image and Vision Computing* and the *ACM Transactions on Intelligent Interactive Systems*. He was a general chair of the 2011 IEEE Conference on Automatic Face and Gesture Recognition. In 2011, he received the Fulbright-Nokia Distinguished Chair in Information and Communications Technologies. He is a senior member of the IEEE.



David J. Kriegman received the BSE degree in electrical engineering and computer science from Princeton University in 1983. He received the MS degree in 1984 and the PhD degree in 1989 in electrical engineering from Stanford University. Since 2002, he has been a professor of computer science and engineering in the Jacobs School of Engineering at the University of California, San Diego (UCSD). Prior to joining UCSD, he was an assistant and associate professor of electrical engineering and computer science at Yale University (1990-1998) and an associate professor with the Computer Science Department and Beckman Institute at the University of Illinois at Urbana-Champaign (1998-2002). He was founding CEO and presently serves as chief scientist of Taaz, Inc. He was chosen for a US National Science Foundation Young Investigator Award, and has received best paper awards at the 1996 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), the 1998 European Conference on Computer Vision, and the 2007 International Conference on Computer Vision (Marr Prize, runner-up), as well as the 2003 Paper of the Year Award from the *Journal of Structural Biology*. He served as program cochair of CVPR 2000 and general cochair of CVPR 2005. He was the editor-in-chief of the *IEEE Transactions on Pattern Analysis and Machine Intelligence* from 2005-2008. He is a senior member of the IEEE.

Thomas S. Huang received the ScD degree from the Massachusetts Institute of Technology (MIT) in electrical engineering, and was on the faculty of MIT and Purdue University. He joined the University of Illinois at Urbana-Champaign in 1980 and is currently the William L. Everitt Distinguished Professor of Electrical and Computer Engineering, Research Professor of Coordinated Science Laboratory, Professor of the Center for Advanced Study, and cochair of the Human Computer Intelligent Interaction major research theme of the Beckman Institute for Advanced Science and Technology. He is a member of the National Academy of Engineering and has received numerous honors and awards, including the IEEE Jack S. Kilby Signal Processing Medal (with A. Netravali) and the King-Sun Fu Prize of the International Association of Pattern Recognition. He has published 21 books and more than 600 technical papers in network theory, digital holograpy, image and video compression, multimodal human computer interfaces, and multimedia databases. He is a fellow of the IEEE.