Internet Vision

By SHAI AVIDAN
Guest Editor

SIMON BAKER
Guest Editor

YING SHAN
Guest Editor

The Internet has increasingly become a multimedia phenomenon. Huge volumes of images and video are available through photo and video sharing sites, news and sports sites, geo-mapping sites, informational sites, etc. For the field of computer vision (the automated analysis of images and video), the Internet offers both: 1) new application areas and 2) potential solutions to existing hard problems such as object recognition and scene understanding.

The goal of this special issue is to provide the reader with a general sense of the research being conducted in "Internet vision," the field at the intersection of computer vision and the Internet. While Internet vision is a relatively new field,[1] a number of significant advances have already been made.

The goal of this Special Issue is to provide the reader with a general sense of the research being conducted in Internet Vision, defined as the intersection of Computer Vision and the Internet and this issue includes coverage of a number of significant advances in this field.

The first four papers of the issue describe new applications that are motivated by the rise of the Internet. The main focus of these applications is to extend computer vision algorithms to Internet scale or to adapt them to the unique settings of the Internet.

The first two papers are:

- "Scene reconstruction and visualization from community photo collections," by Noah Snavely, Ian Simon, Michael Goesele, Richard Szeliski, and Steven M. Seitz;
- "Infinite images: Creating and exploring a large photorealistic virtual space," by Biliana Kaneva, Josef Sivic, Antonio Torralba, Shai Avidan, and William T. Freeman.

Both of these papers present methods of visualizing and navigating large collections of photographs. The first paper considers photo collections of a single location, such as Trafalgar Square in London, U.K. It first determines the relative 3-D positions and orientations of the cameras and computes a sparse 3-D model of the scene. This geometric information then allows users to transition intuitively and smoothly between photos in the collection. The problem of determining the relative positions of images is a long-standing problem in computer vision, but scaling it to Internet scale and dealing with the huge variety in the quality of the photos found on the Internet is what makes this system unique.

The second paper presents a system for exploring large collections of photos in a virtual 3-D space, where the system does not assume the photographs are of a single real 3-D location, or that they were taken at the same time. Instead, photos are organized in themes, such as city streets or skylines, and users are free to navigate within each theme using intuitive 3-D controls that include move left/right, zoom, and rotate.

The third paper is:

- "Toward large-scale face recognition using social network

[1]The guest editors organized the First Workshop on Internet Vision in June 2008 (http://www.Internetvisioner.org/).

Digital Object Identifier: 10.1109/JPROC.2010.2050376

context," by Zak Stone, Todd Zickler, and Trevor Darrell.

This paper considers the problem of face recognition on social networks and photo sharing sites, most notably Facebook. Users of such sites may wish to search for photos of themselves or their friends. They may also wish to filter incoming feeds in the same manner. The size of social networks presents a significant challenge to face recognition. On the other hand, the structure of the social network also presents contextual information that can help perform recognition more accurately.

The fourth paper is:

- "Contextual internet multimedia advertising," by Tao Mei and Xian-Sheng Hua.

Advertising is the main source of revenue for many Internet companies. Targeting advertisements to the demographics of the viewers is naturally as important on the Internet as for more traditional media. Automated analysis of the imagery being displayed on webpages is an important new application area for computer vision.

The next six papers describe how the Internet helps advance research in the field of computer vision. All these papers treat the Internet as a giant data source that can be harvested to help solve hard computer vision problems that require large training sets, most notably object recognition and scene understanding. Besides its pure size, the Internet has two key properties that make it a particularly good data source: 1) it is relatively representative of the kind of imagery that people care about, and 2) there are a variety of ways to extract labels or annotations for the data, as described in more details in the papers themselves.

The fifth paper is:

- "It's all about the data," by Tamara L. Berg, Alexander Sorokin, Gang Wang, David Alexander Forsyth, Derek Hoiem, Ian Endres, and Ali Farhadi.

This paper explains how labeled training data are vitally important for many computer vision algorithms and presents three case studies of how the Internet can be used to obtain high-quality training data.

The sixth paper is:

- "Learning object categories from Internet image searches," by Rob Fergus, Li Fei-Fei, Pietro Perona, and Andrew Zisserman.

This paper shows how the results returned by an image search engine can be used to learn models for object recognition.

The seventh and eights papers are:

- "LabelMe: Online image annotation and applications," by Antonio Torralba, Bryan C. Russell, and Jenny Yuen;
- "I2T: Image parsing to text description," by Benjamin Yao, Xiong Yang, Liang Lin, Mun Wai Lee and Song-Chun Zhu.

Both papers argue that obtaining large, annotated photo collections is a key step in developing better computer vision applications but take different approaches towards reaching this goal. The seventh paper shows how every person with an Internet connection can help advance computer vision research. The key idea is to develop an online tool that allows users to annotate images quickly and easily. The tool is publicly available and, as a result, the authors were able to obtain large amounts of detailed image annotations that were then used in a variety of applications. The eighth paper describes the development of an image parsing system, a system that attempts to convert an image into a textual description. A key component to this system is a database of carefully annotated Internet data, in this case generated under the control of an independent organization.

The ninth paper is:

- "Image interpretation using large corpus: Wikipedia," by Mandar Rahurkar, Shen-Fu Tsai, Charlie Dagli, and Thomas S. Huang.

This paper shows how to use an online encyclopedia (Wikipedia) to provide an ontology with which to represent high-level world knowledge in images.

In all of the papers above, the Internet is being used as data source. Importantly, the raw imagery can be combined with human understanding of the world, whether embedded in captions, search engine indices, online encyclopedias, or in human brains (being accessed via labeling tools).

The tenth and final paper is:

- "Vision of a Visipedia," by Pietro Perona.

In this paper, the author presents a proposal for how the knowhow and hard work of computer vision researchers can be combined collaboratively in a manner similar to online encyclopedias such as Wikipedia to form a repository of human understanding of visual imagery.

Finally, the guest editors would like to thank all the reviewers and authors for their time and hard work, without which this special issue would not have been possible. The guest editors also would like to thank Tom Huang and Harry Shum for their continued support and encouragement during this endeavor. ■

ABOUT THE GUEST EDITORS

Shai Avidan received the Ph.D. degree from the School of Computer Science, Hebrew University, Jerusalem, Israel, in 1999.

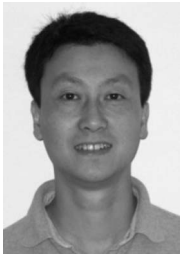
Currently, he is a Senior Research Scientist at the Creative Technologies Lab, Adobe Systems, Newton, MA, and, as of 2009, he is an Assistant Professor at the Faculty of Engineering, Tel-Aviv University, Tel-Aviv, Israel. He was a Postdoctoral Researcher at Microsoft Research, a Project Leader at MobilEye, a startup company developing camera based driver assisted systems, and a Research Scientist at Mitsubishi Electric Research Labs (MERL). He published extensively in the fields of object tracking in video sequences and 3-D object modeling from images. Recently, he has been working on Internet vision applications such as privacy preserving image analysis, distributed algorithms for image analysis, and media retargeting, the problem of properly fitting images and video to displays of various size.

Dr. Avidan was an Area Chair for the 2007 IEEE International Conference on Computer Vision, as well as the 2008 and 2009 IEEE Conference on Computer Vision and Pattern Recognition. In addition, he was a Program Chair for several workshops, including the Workshop on Privacy Research in Vision held in conjunction with the 2006 IEEE Conference on Computer Vision and Pattern Recognition, and the First Workshop on Internet Vision held in conjunction with the 2008 IEEE Conference on Computer Vision and Pattern Recognition.

Simon Baker received the B.A. degree in mathematics from Trinity College, Cambridge, U.K., in 1991, the M.Sc. degree in computer science from the University of Edinburgh, Edinburgh, U.K., in 1992, the M.A. degree in mathematics from Trinity College in 1995, and the Ph.D. degree from the Department of Computer Science, Columbia University, New York, NY, in 1998.

Currently, he is a Principal Researcher in the Interactive Visual Media Group, Microsoft Research, Redmond, WA. Before joining Microsoft in 2006, he was an Associate Research Professor at the Robotics Institute, Carnegie Mellon University, Pittsburgh, PA. His research interests include face recognition, modeling and tracking, human body modeling and tracking, super-resolution, 3-D reconstruction, vision for safe driving, projector-camera systems, and all aspects of video processing.

Dr. Baker was an Associate Editor for the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE from 2004 to 2008. He was Program Co-Chair for the IEEE Conference on Computer Vision and Pattern Recognition in 2007 (http://cvpr.cv.ri.cmu.edu/) and Program Co-Chair for the First and Second IEEE Workshops on Internet Vision in 2008 and 2009 (http://www.Internetvision.org/).



Ying Shan received the B.E. degree in chemical engineering, focusing on automatic process control, from Zhejiang University, Zhejiang, China, in 1990 and the M.S. and Ph.D. degrees in computer science from Shanghai Jiaotong University, Shanghai, China, in 1993 and 1997, respectively.

Currently, he is a Lead Applied Researcher at Microsoft AdLabs, Redmond, WA. He was a Postdoctoral Researcher at Microsoft Research from 1999 to 2001. Until 2006, he was a Senior Member of the Technical Staff in Sarnoff Corporation's Vision and Learning Laboratory, where he initiated, led, and contributed to a number of government and commercial projects. His research interests include computer vision, pattern recognition, machine learning, and computer graphics, with applications in online video ads, video surveillance, object/face detection, image registration, and 3-D reconstruction. He has published more than 25 peer-reviewed papers, holds 19 U.S. patents, and has ten others pending.

Dr. Shan is an active reviewer of top journals such as the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE and the *International Journal on Computer Vision*. He was on the program committee of major international conferences such as the European Conference on Computer Vision, Computer Vision and Pattern Recognition, and International Conference on Computer Vision. He was a Program Co-Chair for the First and Second IEEE Workshop on Internet Vision in 2008 and 2009. He is the recipient of Sarnoff's Recognition Award in 2003, Innovation Award from 2003 to 2005, and a two-time winner of Gold Star Award since he rejoined Microsoft in 2007.