2013
Volume 9, Number 1s

ACM Transactions on

Multimedia Computing, Communications and Applications

Association for Computing Machinery

Advancing Computing as a Science & Profession

ACM Transactions on Multimedia Computing, Communications and Applications

2013 • Vol. 9 • No. 1s

Articles 31-53

Navigating the Worldwide Community of Photos

RICHARD SZELISKI, Microsoft Research
NOAH SNAVELY, Cornell University
STEVEN M. SEITZ, University of Washington and Google Inc.

The last decade has seen an explosion in the number of photographs available on the Internet. The sheer volume of interesting photos makes it a challenge to explore this space. Various Web and social media sites, along with search and indexing techniques, have been developed in response. One natural way to navigate these images in a 3D geo-located context. In this article, we reflect on our work in this area, with a focus on techniques that build partial 3D scene models to help find and navigate interesting photographs in an interactive, immersive 3D setting. We also discuss how finding such relationships among photographs opens up exciting new possibilities for multimedia authoring, visualization, and editing.

Categories and Subject Descriptors: I.4.8 [**Image Processing and Computer Vision**]: Scene Analysis; I.3.3 [**Computer Graphics**]: Picture/Image Generation; H.5.2 [**Information Interfaces and Presentation**]: User Interfaces

General Terms: Algorithms

Additional Key Words and Phrases: Image-based rendering, image-based modeling, visualization

ACM Reference Format:

Szeliski, R., Snavely, N., and Seitz, S. M. 2013. Navigating the worldwide community of photos. ACM Trans. Multimedia Comput. Commun. Appl. 9, 1s, Article 47 (October 2013), 4 pages.
DOI: http://dx.doi.org/10.1145/2492208

1. INTRODUCTION

While the last decade has seen an incredible explosion of multimedia content on the Web, our ability to index, search, and navigate these huge collections has also dramatically increased. Most multimedia search and indexing schemes rely on metadata, such as tags and anchor text, as seen on websites such as Flickr and YouTube. Another popular way to index photographs is by geo-location, since users can browse related photos and find visual information of interest. However, simply locating photos on a map with push-pins does not provide a particularly intuitive way to navigate related photos. A third class of techniques relies on visual similarity, which, while useful, is still a very challenging problem to solve.

A different line of multimedia content development has been the proliferation of photographic VR (virtual reality) sites, where large immersive panoramas can be navigated interactively. While

This work is supported by several agencies; see Agarwal et al. [2011] for the complete list.
Authors' addresses: R. Szeliski, Microsoft Research, One Microsoft Way, Redmond, WA 98052; email: szeliski@microsoft.com; N. Snavely, Department of Computer Science, Cornell University, Ithaca, NY 14853-7501; email: snavely@cs.cornell.edu; S. M. Seitz, Department of Computer Science and Engineering, University of Washington, Seattle, Washington 98195-2350; email: seitz@cs.washington.edu.

DOI: http://dx.doi.org/10.1145/2492208

Fig. 1. The Photo Tourism system takes collections of Internet images (originally found using keyword search) and reconstructs a sparse 3D point model of the scene along with the camera locations associated with each photograph. This information can then be used to interactively browse the collection of photos.

individual high-resolution panoramas can provide a great amount of visual detain and richness, an even more immersive experience can be created from navigable panoramic videos [Uyttendaele et al. 2004], as evidenced by the popularity of services such as Google Street View and Microsoft StreetSide.

How can we best combine the sheer volume and richness of user-generated photographs with the immersive navigation capabilities of VR photography? Our answer was the creation of the Photo Tourism system [Snavely et al. 2006], which takes roughly geo-located photographs from either individual users or the Web and creates a navigable 3D model in which these photographs can be explored.

In this brief overview, we summarize the components and attributes of the Photo Tourism system (Figure 1) and then describe a variety of additional applications and scenarios that it enables. For a longer survey article on these topics (and more detailed references), please see Snavely et al. [2010].

2. PHOTO TOURISM

The genesis of Photo Tourism lay both in previous work we had done in panoramic photography [Szeliski and Shum 1997] and view morphing [Seitz and Dyer 1996], as well as our own personal interests in travel and photography. The question we set out to answer was how can we create a rich immersive and navigable experience from varied collections of photographs taken at locations such as a tourist sites? Could we create an experience that wasn't a full 3D model and also had some of the photographic qualities seen in traditional slide shows, for example, pan-and-zoom "Ken Burns" effects? Some of our early experiments used photographs we had personally taken while "strafing" a façade so we could stabilize the background while fading amongst the various foregrounds (café tables, moving people). We also experimented with various levels of 3D models (proxies), since we knew that these were important for high quality image-based rendering.

After trying a variety of 3D models (3D meshes, piecewise-planar proxies for buildings and ground), we discovered that in most cases, a single planar proxy oriented to fit the majority of the 3D points provided a familiar transition (the "3D tile flip" seen in sports broadcasts) while maintaining good visual continuity. We also discovered that we could harvest images from photo sharing sites such as Flickr using location keywords and select from this plethora of images those that matched well based on feature points and 3D structure from motion. Along the way, we had to solve a number of challenging

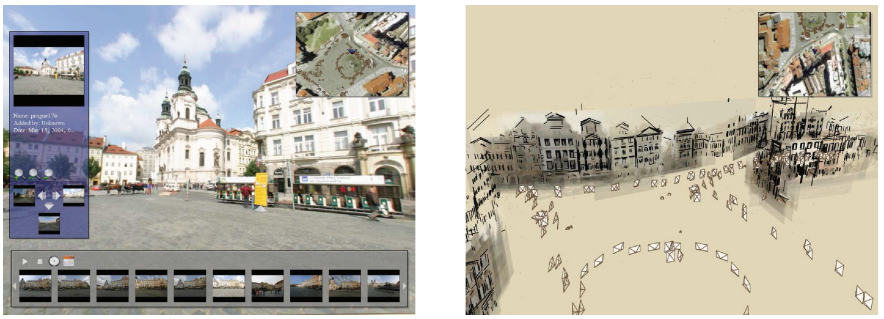

Fig. 2. The photo exploration interface (left) shows not only the current image but also a filmstrip of related images, a navigation pane that allows 2D movement, as well as an overhead map. The 3D overview (right) shows a sketchy rendering of the scene as well as the overhead map, which indicates both the current location and the locations of other cameras.

computer vision problems, such as reliably matching and reconstructing 3D scenes from unstructured data (using incremental seed-and-grow techniques) and robust plane fitting to the reconstructed 3D point clouds. We also developed new navigation techniques (both map-based and direct manipulation, such as "zoom here", "move sideways", etc.) as well as evocative renderings based on low-resolution washes and stylized 3D lines segments (see Figure 2).

In subsequent work [Snavely et al. 2008], we address a number of limitations in our original system. These included the confusing nature of the navigation based on single point-to-point transitions, as well as differences in lighting and exposure between overlapping images. We developed novel techniques to automatically create interesting paths through photo collections and also techniques to blend between exposure settings during transitions.

One of the interesting multimedia applications of our system was in the automatic region-based tagging of photographs and points of interest based solely on whole-image tags. Ian Simon (another graduate student in our group) and Steve Seitz developed a system that first clustered 3D points likely to correspond to particular tags and then turned these into bounding boxes that labelled objects of interest in Internet photographs (Figure 3) [Simon and Seitz 2008]. In related work, Ian, Noah, and Steve developed a system to select the most representative photographs to describe a particular landmark or point of interest [Simon et al. 2007]. The ability to construct large-scale 3D point clouds from casually acquired images, based on advances in computer vision techniques for feature matching and structure from motion, has re-energized the field of multiview stereo reconstruction. Our work, with an extended set of new collaborators, has shown that the quality of reconstructions based on collections of images could rival that obtained with commercial laser scanning equipment [Goesele et al. 2007]. It also inspired us toward the ambitious goal of reconstructing whole cities from community-based photographs. Our paper "Building Rome in a Day" [Agarwal et al. 2011] describes a distributed system for matching the millions of images required for this task and efficiently and stably performing the 3D reconstruction. In concurrent and ongoing work, we have also been improving the quality of the dense multiview stereo algorithms and specializing them to applications such as architectural modeling.

Photo Tourism has also inspired the creation of several commercial applications and large-scale photo sharing communities. The first of these is Photosynth (`http://photosynth.net/`), which was developed at Microsoft. This system allows users to automatically reconstruct partial 3D models from personal photographs and to upload the results to a Web site for sharing with a community of like-minded photographers.

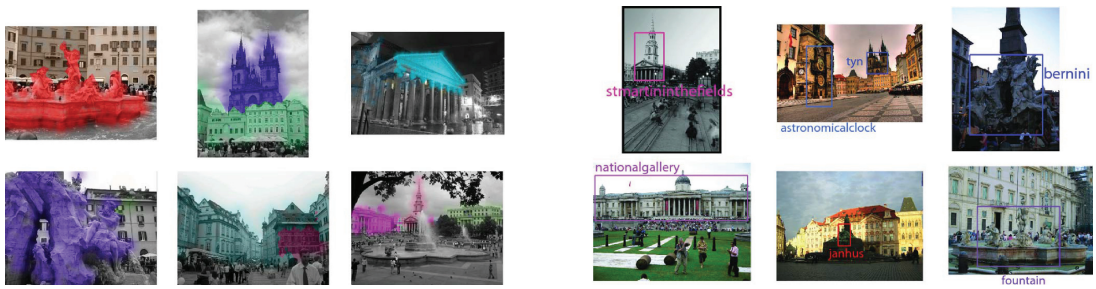

Fig. 3. Scene segmentation using the wisdom of crowds: the likelihood of a given image tag being associated with a particular 3D point, that is, the tag being present in an image where the point is matched, is visualized on the left. This allows the system to automatically label 2D regions of interest in images (right).

A more recent example is "Photo Tours" [Kushal et al. 2012], which automatically constructs tours through collections of Internet imagery of a landmark and then uses full 3D transitions between nearby images to give a striking sense of realism and physical presence.

3. DISCUSSION

The construction and continued development of Photo Tourism has been one of the most exciting endeavors in our research careers. We believe that it is a great example of the bridges that can be built between the multimedia, computer vision, and computer graphics communities when advanced (but imperfect) matching and reconstruction techniques are married with beautiful rendering and intuitive navigation techniques to explore the richness of our visual world. In the future, we expect these ideas to percolate into the realm of shared videos and to become a widely used component in larger-scale multimedia search and exploration systems.

REFERENCES

AGARWAL, S., FURUKAWA, Y., SNAVELY, N., SIMON, I., CURLESS, B., SEITZ, S. M., AND SZELISKI, R. 2011. Building Rome in a day. *Comm. ACM 54*, 10, 105–112.

GOESELE, M., SNAVELY, N., CURLESS, B., HOPPE, H., AND SEITZ, S. M. 2007. Multi-view stereo for community photo collections. In *Proceedings of the 11th International Conference on Computer Vision (ICCV'07)*.

KUSHAL, A., SELF, B., FURUKAWA, Y., GALLUP, D., HERNANDEZ, C., CURLESS, B., AND SEITZ, S. M. 2012. Photo tours. In *Proceedings of the 2nd International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT)*. 57–64.

SEITZ, S. M. AND DYER, C. M. 1996. View morphing. In *Proceedings of the ACM SIGGRAPH Conference.* 21–30.

SIMON, I. AND SEITZ, S. M. 2008. Scene segmentation using the wisdom of crowds. In *Proceedings of the 10th European Conference on Computer Vision (ECCV'08)*. 541–553.

SIMON, I., SNAVELY, N., AND SEITZ, S. M. 2007. Scene summarization for online image collections. In *Proceedings of the 11th International Conference on Computer Vision (ICCV'07)*.

SNAVELY, N., GARG, R., SEITZ, S. M., AND SZELISKI, R. 2008. Finding paths through the world's photos. *ACM Trans. Graph. 27*, 3.

SNAVELY, N., SEITZ, S. M., AND SZELISKI, R. 2006. Photo tourism: Exploring photo collections in 3D. *ACM Trans. Graph. 25*, 3, 835–846.

SNAVELY, N., SIMON, I., GOESELE, M., SZELISKI, R., AND SEITZ, S. M. 2010. Scene reconstruction and visualization from community photo collections. *Proc. IEEE 98*, 8, 1370–1390.

SZELISKI, R. AND SHUM, H.-Y. 1997. Creating full view panoramic image mosaics and environment maps. In *Proceedings of the ACM SIGGRAPH Conference.* 251–258.

UYTTENDAELE, M., CRIMINISI, A., KANG, S. B., WINDER, S., HARTLEY, R., AND SZELISKI, R. 2004. Image-based interactive exploration of real-world environments. *IEEE Comp. Graph. Appl. 24*, 3, 52–63.

Received May 2013; accepted June 2013