DOI:10.1145/1400181.1400201

Technical Perspective
Computational Photography on Large Collections of Images

By Marc Levoy

THIS PAPER WILL strike a familiar chord with anyone who has ever taken a picture. The problem is easy to understand—replacing unwanted parts of a photograph. The authors start with an interesting twist, by making a distinction between data that "should have been there" but was obscured by a telephone pole, and data that "could have been there," meaning that it constitutes an incorrect but plausible picture. This is a difficult problem, because faked pictures are relatively easy to spot, as recent scandals in photojournalism have proven. Nevertheless, Hays and Efros obtain impressive results; check out Figure 1 online (http://graphics.cs.cmu.edu/projects/scene-completion/scene-completion.pdf); it looks seamless no matter how closely you zoom into it.

The paper represents the confluence of several noteworthy trends in computing. First, it exemplifies a new application area, *computational photography*, which refers broadly to sensing strategies and algorithms that extend the capabilities of digital photography. Representative techniques include high-dynamic-range imaging, flash-noflash and coded aperture imaging, panoramic stitching, digital photomontage, and light field imaging. ACM SIGGRAPH is at the forefront of this new area. Indeed, of the 108 papers at its 2007 conference, 20 were arguably about computational photography. This paper fits squarely in that group.

Second, the authors exemplify the ongoing convergence of several formerly isolated research communities. To find a suitable replacement for the unwanted part of a photograph, the authors search a collection of images using "gists," an image summarization technique pioneered in the cognitive science community by Aude Oliva and Antonio Torralba. They then find the best seam along which to insert the matching content using graph-cuts, an algorithm first applied to images by Yuri Boykov, Vladimir Kolmogorov, and Ramin Zabih in the computer vision community. Finally, they smooth the seam between new and old imagery using gradient domain blending, a technique introduced into the graphics literature by Raanan Fattal (2003) for tone mapping of high-dynamic-range images.

Third, this paper provides evidence of the notion, gaining credence in many application domains, that simple machine learning algorithms often outperform more sophisticated ones if trained on large enough databases. Natural language translation algorithms work dramatically better if trained on millions of documents than on thousands. Image classification and segmentation algorithms do, too, as the authors argue here. Want to remove a garbage truck from your snapshot of an Italian piazza? Start with a database containing lots of Italian piazzas.

How far can one push this data-centric approach to image matching? While it's impossible to collect all possible images of the world, the authors make the conjecture that one could collect all "semantically differentiable scenes." I'm doubtful. It is well known that features in natural scenes form a heavy-tailed distribution, meaning that while some features in photographs are more common than others, the relative occurrence of less common features drops slowly. In other words, there are many unusual photographs in the world.

A closely related question is: What is meant by data that "could have been there?" The authors define this as all "semantically valid" scenes, but semantic validity is maddeningly difficult to pin down. Indeed, to evaluate their results quantitatively, Hays and Efros resort to a human study: How well can naïve viewers distinguish an algorithmically completed image from a real photograph? In the end this question may prove to be "AI-complete," that is, it's as hard as making computers as intelligent as people.

Regardless of whether we ever answer this question, it is clear that large collections of images are useful. Aside from completing photographs, they can be used to build 3D models of urban monuments (as in Snavely et al.'s Photo Tourism, http://phototour.cs.washington.edu/) or to synthesize textures from image exemplars (as in Kopf et al.'s Solid Texture Synthesis; http://www.johanneskopf.de/publications/solid/index.php). Maybe collections of images can even help me take better pictures! If we sort the collection geographically, as Hays and Efros do in a follow-on paper (in CVPR 2008), then my camera could query a central database to help it decide what color balance to use for the shot I'm now taking of the Grand Canyon. Closer to home, what color balance should the camera use to photograph my wife? It ought to know, since my online albums contain hundreds of photographs of her.

Finally, while the future of computational photography is exciting because of papers like this, the future is also murky because it's not obvious who will use this stuff. When I show this paper to someone outside computing, their first question is: Why? How many people (other than revisionist dictators) would remove a person or large object from a photograph? Tied to this question is the debate about how much effort can we expect consumers to exert after they've taken a photograph. No matter what position you take on these questions, nobody doubts that the coming years will be interesting ones for the computer graphics and vision communities. This paper helps light the way. C

Marc Levoy (levoy@cs.stanford.edu) is a professor of computer science and (jointly) electrical engineering at Stanford University.