

Get Out of my Picture! Internet-based Inpainting

Oliver Whyte¹

Josef Sivic¹

Andrew Zisserman^{1,2}

¹INRIA, WILLOW Project,
Laboratoire d'Informatique de l'Ecole Normale Supérieure, Paris,
France

²Department of Engineering Science,
University of Oxford

Often when we review our holiday photos, we notice things we wish we could have avoided, such as vehicles, construction work, or simply other tourists. We cannot go back and retake the photo, so what can we do if we want to remove these things from our photos? We want to replace these sections of the image in a convincing way, preferably with what would really have been seen there without the occlusions. Previous work on this problem, often referred to as “inpainting”, are mainly applicable to small image regions, and rely largely on models of the local behaviour of natural images, *e.g.* [3, 5]. Recently, replacing large occlusions in photographs has been approached using images from the Internet [2, 7] or by combining several images captured at approximately the same time [1, 9].

In this paper we leverage recent advances in viewpoint invariant image search [8] to find other images of the same scene on the Internet. Beginning with a *query image* containing a *target region* to be replaced, we first use an online image search engine to retrieve images of the same scene, and take these to be a set of *oracles*. Since these images may have significant variations in viewpoint and lighting, we register each oracle to the query image using multiple homographies and a simple global photometric correction. We then use each oracle to propose a solution, by copying image data into the target region using Poisson blending. Finally, we use a Markov random field (MRF) formulation to combine the proposals into a single, occlusion-free result. Figure 1 shows the main stages of our system, and compares our result to those of two other methods.

Registration: To estimate a homography between the query image and an oracle, we use the standard method [6] of putatively matching interest points between the two images, and estimating the inliers and homography simultaneously using RANSAC. We discard any homography which has too few inliers to be reliable, which distorts the oracle too extremely, or for which the oracle covers too little of the target region after warping. For a scene containing multiple planes, a single homography will in general be insufficient to register an oracle to the whole query image. Thus we allow multiple homographies to be detected for each oracle, by repeatedly running RANSAC on the remaining unused interest points. In a final step, we allow the user to register the ground plane semi-automatically, which is often unregistered due to a lack of distinctive points.

To reduce the effect of lighting variations when combining oracle images with the query, we estimate a global photometric correction for each oracle image. To do this, we find regions of the oracle (outside the target region) which have been *well-registered*, and estimate a global linear correction on the gradients of each colour channel using those regions.

In order to guide the final solution, we aim to compute a robust “average” estimate of the unoccluded image from the registered oracles. To avoid averaging together oracles which have been registered to different scene planes, we group the homographies such that each group corresponds to one scene plane, following the idea that homographies registering the same scene plane should use many of the same interest points from the query image.

Generating and combining proposals: Once we have registered an oracle, we use Poisson blending to combine it with the query image, whereby the two images’ gradient fields are combined to form a composite gradient field, which can then be reconstructed into an image by solving Poisson’s equation.

The final stage in our system is to take the multiple proposals of what might replace the target region, and to combine parts of them to produce the best result, free from any occlusions or any badly registered regions. To do this, we set up a multi-label MRF over the target region, where the label at a pixel corresponds to which proposal is used there, and where we wish to minimise over the label configuration a cost function which encourages the solution to follow the unoccluded estimate provided by the median images, while hiding the seams between regions. We use a unary cost that around the edge of the target region increases as the result devi-



Figure 1: **An overview of our system.** Top row: The inputs to the system and the output labels showing the combination of proposals in the final result. Middle rows: The top 5 oracles used in the result and their proposals. Oracles are obtained automatically from the Internet using viewpoint invariant search. Last row: Our result, the result using the algorithm of Criminisi *et al.* [4] and the result using the method of Hays & Efros [7].

ates from the query image, while inside the target region, increases as the result deviates from the robust average of the oracles computed earlier. As our pairwise cost, we use the “gradient” cost of Agarwala *et al.* [1], which encourages boundaries between regions with different labels to move to places where the gradients of the two proposals are similar. Finally, we optimise the MRF using tree-reweighted belief propagation.

The details of our implementation are described more fully in the paper, along with example results showing that our system is able to produce convincing results on a range of images, and that the results correspond well to the images that would have been observed without the occlusions.

- [1] A. Agarwala, M. Dontcheva, M. Agrawala, S. Drucker, A. Colburn, B. Curless, D. Salesin, and M. Cohen. Interactive digital photomontage. *ACM Trans. Graph. (Proc. SIGGRAPH 2004)*, 23(3), 2004.
- [2] H. Amirshahi, S. Kondo, and T. Aoki. Photo completion using images from internet photo sharing sites. In *Proc. MIRU*, 2007.
- [3] M. Bertalmío, G. Sapiro, V. Caselles, and C. Ballester. Image inpainting. In *SIGGRAPH '00*, 2000.
- [4] A. Criminisi, P. Pérez, and K. Toyama. Object removal by exemplar-based inpainting. In *Proc. CVPR*, 2003.
- [5] A. A. Efros and T. K. Leung. Texture synthesis by non-parametric sampling. In *Proc. ICCV*, 1999.
- [6] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge: CUP, second edition, 2004.
- [7] J. Hays and A. A. Efros. Scene completion using millions of photographs. *ACM Trans. Graph. (Proc. SIGGRAPH 2007)*, 26(3), 2007.
- [8] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman. Object retrieval with large vocabularies and fast spatial matching. In *Proc. CVPR*, 2007.
- [9] M. Wilczkowiak, G. J. Brostow, B. Tordoff, and R. Cipolla. Hole filling through photomontage. In *Proc. BMVC*, 2005.