Get Out of my Picture!
Internet-based Inpainting

Oliver Whyte[1]

Josef Sivic[1]

Andrew Zisserman[1,2]

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

[2] Department of Engineering Science,
University of Oxford

Abstract

We present a method to replace a user specified target region of a photograph by using other photographs of the same scene downloaded from the Internet via viewpoint invariant image search. Each of the retrieved images is first geometrically then photometrically registered with the query photograph. Geometric registration is achieved using multiple homographies and photometric registration is performed using a global affine transformation on image intensities. Each retrieved image proposes a possible solution for the target region. In the final step we combine these proposals into a single visually consistent result, using a Markov random field optimisation to choose seams between proposals, followed by gradient domain fusion. We demonstrate removal of objects and people in challenging photographs of Oxford landmarks containing complex image structures.

1 Introduction

Often when we review our holiday photos, we notice things that we wish we could have avoided, such as vehicles, construction work, or simply other tourists. To go back and retake the photograph is impossible, so what can we do if we want to remove these things from our photos? We wish to replace the offending pixels in as convincing a way as possible, a problem often referred to as image completion or inpainting, which has been widely studied recently. However, most existing methods are only applicable to small image regions, or require significant user intervention. We would ideally like to have no limit on the complexity of the image in the regions we are replacing, and would like to be able to fill pixels with whatever would actually have been observed there had the occlusions not occurred, with minimal user interaction. With the growth of online photo sharing websites, we have millions of photos of the world available to us, probably including many of the very same place as our own photo, so how can we use them to fix up our snaps?

In this work we leverage recent advances in viewpoint invariant image search [10, 23] to find other images of the same scene. 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. Taking the search results to be a set of *oracles*, we use them to propose solutions

Figure 1: **An example result from our algorithm.** From left to right: Original query image, target region to replace, result from our system. The replaced region is consistent with the rest of the original image and the boundary between is effectively hidden, producing a convincing result. Note the complexity of image structures on the inpainted facade.

to the filling problem, and solve a labelling problem to decide how best to combine the proposals. Figure 1 shows an example result obtained from our system.

Previous work on the problem of image completion / inpainting is diverse, varying according to the size of the region to be filled, the kind of image surrounding it, and the amount of user interaction required. The filling of small regions has generally been approached by modelling the local behaviour of images, *e.g.* Bertalmío *et al.* [7] and Efros & Leung [13]. Except in some simplified cases, these local approaches fail when the region is larger than a few pixels across, which has led to algorithms such as those of Criminisi *et al.* [11], Sun *et al.* [26], and Bertalmío *et al.* [8] being proposed, which try to preserve both large-scale salient structures across the region and fine textures. However, these algorithms are still constrained to draw their information from the image itself or from simple manually-input guides. Thus there is evidently a limit to the kind of regions they can reconstruct.

The idea of using large regions of other images for image completion was explored by Hays & Efros [16], who proposed using Internet photo collections for the purpose, however they focussed not on using photos of the same scene, but rather images which were semantically similar to that being completed. Thus the results are often convincing, but are not true to the actual scene content. Others have combined images from the same scene [4, 24, 29] under the constraint that the images are all captured in quick succession by the same camera, or by carrying out a full 3D reconstruction [14, 25].

Image completion using images of the same scene from the Internet was independently of our work approached by Amirshahi *et al.* [5, 6], who obtain relevant images using a text-based image search, then combine unoccluded blocks from them to replace the target region using a greedy algorithm. Our novel contributions are the following: (i) candidate images are obtained automatically using viewpoint invariant image search (Section 2), (ii) geometric registration is performed using multiple homographies, which allows registering significantly more complex scenes captured from varying camera viewpoints (Section 3), and (iii) occlusion reasoning and seamless combination of multiple registered oracle images is formulated as a single labelling problem efficiently solved using tree-reweighted belief propagation (Section 4).

2 Retrieving Oracle Images

Recent works [10, 17, 23] have demonstrated the feasibility of large-scale image-based search engines, allowing the user to input an image region as a search query, and be presented with other images depicting the same object or scene. In this work we use the "Oxford Buildings Search" online demonstration [3] provided by Philbin *et al*. [23]. The system uses a set of 100,000 images collected from Flickr.com, for which an indexing structure has already been computed. Both the query images and oracle images from which we construct the solution come from this set. Figure 2 shows a typical query image and the top 30 results returned by the search engine. These search results are the *oracles* which are used to replace the target region in a query image.
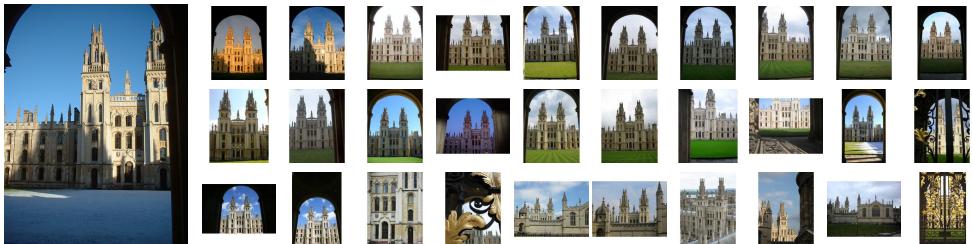

Figure 2: **An example query and the first 30 results returned by the viewpoint invariant image search engine.** For popular landmarks it is reasonable to expect the top retrieved results to contain the building in the query image, and for some to have been taken from a similar viewpoint. These search results will be used to replace the target region of our query image.

3 Geometric and Photometric Registration

Having retrieved a set of oracle images depicting the same place as the query image, we first need to deduce their geometrical and photometric relationship to the query image. We would also like to reason jointly about their registration, in order to be able to combine them into a single, coherent, occlusion-free image. In this section, we first explain the use of homographies to register the oracle images to a query image, and the extension to multiple homographies that is necessary to register scenes with multiple planes. We then find regions which have been registered well by each homography, and use these regions to estimate a global transformation on the colour channels of each oracle. Finally, we group together homographies which are likely to have registered the same scene plane, and for each group compute an unoccluded "median" image, which will be used to guide the region selection process in Section 4.

Registering two images with a homography [15] is valid when the optical centres of the two cameras coincide, or when the scene being imaged is planar, as shown in Figure 3. In outdoor urban scenes, the second condition is approximately satisfied very frequently. Additionally, when we have a large number of photographs of a scene taken by many photographers, we find that the first condition can also sometimes be approximately satisfied. Although we are unlikely to find a photo taken from exactly the same location, we are likely to find many for which the effects of parallax will be small.

Homography estimation: We use the standard method [15] of determining putative point matches between the two images (query and oracle) and estimating the inliers and homogra-

Figure 3: **Pairs of images related by homographies.** From left to right on each row: A query photograph, another photograph of the same scene returned by the image search engine, the second image registered to the first using a homography, the two images overlaid, a close-up of the overlaid images. 1st row: A pair of images with approximately the same camera centre – the transformation applies almost equally well to all parts of the image. 2nd row: A pair of images of a piecewise planar scene – the homography registers the dominant wall, where the variation in depth is small, but is no longer valid for points not on that plane, such as the protruding wall on the right.

phy simultaneously using RANSAC. Here we use Harris-Affine [21] and SIFT [20] interest points (typically 5000-15000 per image) and SIFT descriptors.

Having estimated the homography using the inlying interest points, we check that the number of inliers is sufficient for it to be reliable, discarding any homographies with fewer than 50 inliers. We also check whether the line at infinity in the oracle becomes visible under the homography, and discard it if so, since we deem the transformation to be too extreme in this case. Finally, we retain the oracle only if it covers at least 25% of the target region after warping, since an oracle covering less than this is unlikely to be much use for replacing the region's contents.

Multiple homographies: 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. After estimating a homography between the two images, we remove all the inlying interest points from consideration, and run RANSAC again on the remaining putative matches. We repeat this process until the last estimated homography is rejected by the criteria above. Having estimated multiple homographies between all the oracle images and the query image, we furthermore allow the user to register the (possibly unregistered) ground plane semi-automatically (see Figure 4).

Photometric registration: Due to the diverse origins of the images retrieved from the Internet, they may have very different lighting conditions from the query image. To reduce the effect of lighting variations when combining oracle images with the query, we work in the gradient domain, and furthermore estimate a global linear correction on the gradients of each colour channel of each oracle image. The correction is robustly estimated by taking the median over the *well-registered* pixels of the ratio of gradient magnitudes between the query image and the registered oracle, *i.e.* $\text{median}_{\mathbf{p}} \left(\frac{\|\nabla I^c_{\text{query}}(\mathbf{p})\|}{\|\nabla I^c_{\text{oracle}}(\mathbf{p})\|} \right)$, where ∇I^c is the gradient of image I in colour channel c.
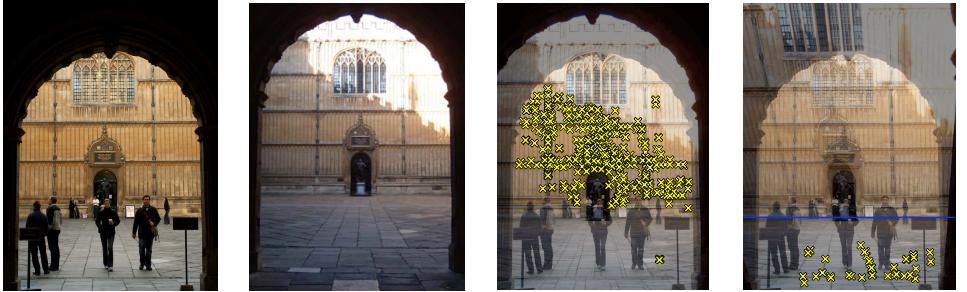
Figure 4: **Semi-automatic ground plane registration.** From left to right: The query image, an oracle image, the first homography extracted with inliers shown, the second homography extracted after the user manually indicates ground plane region (below horizontal line). The first homography extracts the dominant plane, and by manually indicating the ground plane region RANSAC is able to register the ground plane in the second homography.

To determine which regions are well registered in each oracle, in a lighting-invariant way, we compute the normalised cross-correlation (NCC) between the square (15×15) patch around each pixel in the query image with the corresponding square patch in the transformed oracle image. To get a binary mask of well-registered regions, we use a binary MRF with unary costs which are linear in the NCC scores, and a Potts potential on adjacent pixels, and minimise the cost function using graph-cuts. Figure 5 shows typical binary masks indicating the well-registered regions for each oracle.

Grouping homographies: The goal here is to group homographies corresponding to a particular scene plane, and subsequently obtain an unoccluded "median" image for each group (as shown in Figure 5), which will be used to guide selection of unoccluded and well-registered oracles for inpainting the target region (Section 4).

In order to group together homographies which correspond to a particular scene plane, we consider each oracle-homography combination as a node of a graph, and for two homographies having as inliers the sets of interest points S and Q, place an edge between the two nodes if the overlap of these sets is above a threshold: $\frac{|S \cap Q|}{|S \cup Q|} > 0.1$. We consider each connected component of the graph as a group of homographies likely to register the same scene plane.

4 Generating and Combining Proposals

Once we have an oracle geometrically and photometrically registered, we would like to fill the target region using as direct a method as possible. The most direct would be to simply copy the pixels from the oracle into the region, however in practice the variations in lighting mean that this will produce poor quality results, with clear boundaries at the edge of the region. The problem of how best to combine two images whose properties do not necessarily match has been approached in many ways, from methods which aim to conceal boundaries between regions, such as Burt & Adelson's multiresolution spline [9] and Poisson blending [22], to methods that try and find the best place to locate the boundary itself, such as the dynamic programming of Efros & Freeman [12] and the graph-cut technique of Kwatra *et al.* [19].

In this work we use Poisson blending to combine the images, whereby instead of combining pixels from the two images, their gradient fields are combined to form a composite

Group 1:

Group 2:

Figure 5: **Grouping homographies and finding well-registered regions.** From left to right on each row: The query image with each homography group's inlying interest points marked, some of the registered oracles from each group, with the regions considered to be well-registered highlighted, and (far right) the "median" image for each group within the target region. Note that for each group, the median image provides a sharp, unoccluded estimate of the relevant plane, while it is blurry elsewhere. Thus, the difference between a registered oracle and this image will be small where the oracle is well-registered and unoccluded, but large elsewhere.

gradient field, which can then be reconstructed into an image by solving Poisson's equation. The query image provides Dirichlet boundary conditions for the equation around the target region. If the transformed oracle does not span the entire target region, pixels bordering the remaining unfilled region take Neumann boundary conditions, in order to reduce colour artefacts. Figure 6 (4th row) shows some of the individual proposals generated for a query image using Poisson blending.

4.1   Combining Multiple Proposals

So far we have only described a way of filling an image using a single oracle. However, it may be that individual oracles cannot provide the best solution when taken alone, but may be combined into a single result which could not have been achieved using any single oracle. Reasons for this might be occlusions in the oracles themselves, or the fact that a single homography may not be able to register the whole target region. Figure 6 shows the advantage of combining multiple proposals; a single oracle provides most of the result but requires other oracles to provide some small parts, partly due to the mis-registration of the ground plane.

In order to decide which proposal should be used at which pixel, we want to design and optimise a cost function which encourages each pixel to choose well, but to regularise this with the idea that pixels should agree with their neighbours about what their neighbourhood should look like. We can consider this a labelling problem, where for each pixel $\mathbf{p}$ its label $L_{\mathbf{p}}$ corresponds to which proposal is used there. This can be formulated as a multi-state Markov random field (MRF), where we wish to minimise over the label configuration $\mathbf{L}$ a cost function of the form

$$E(\mathbf{L}) = \sum_{\mathbf{p} \in \mathcal{V}} E_1(\mathbf{p}, L_{\mathbf{p}}) + \sum_{(\mathbf{p},\mathbf{q}) \in \mathcal{E}} E_2(\mathbf{p}, \mathbf{q}, L_{\mathbf{p}}, L_{\mathbf{q}}), \qquad (1)$$

where $\mathcal{V}$ indicates the set of pixels in the region being solved, $(\mathbf{p}, \mathbf{q})$ indicates a pair of neighbouring pixels (4-neighbours), with $\mathcal{E}$ the set of all such pairs in the region being

solved. $E_1(\mathbf{p}, L_{\mathbf{p}})$ is the "cost" of the label $L_{\mathbf{p}}$ at pixel $\mathbf{p}$, encoding our wishes for individual pixels, while $E_2(\mathbf{p}, \mathbf{q}, L_{\mathbf{p}}, L_{\mathbf{q}})$ is the cost of assigning neighbouring pixels $\mathbf{p}$ and $\mathbf{q}$ the labels $L_{\mathbf{p}}$ and $L_{\mathbf{q}}$ respectively, encoding the way we wish neighbouring pixels to agree with each other.

Pixels outside the target region should look similar to the original query image, since they lie outside the region originally specified for replacement. Pixels on the inside of the target region however should be similar to some robust estimate of the unoccluded scene, to avoid inserting new occlusions into the image. To achieve these two goals we choose E_1 to have the form

$$E_1(\mathbf{p}, L_{\mathbf{p}}) = k_{\text{query}} \overline{M}(\mathbf{p}) \| \mathbf{I}_{L_{\mathbf{p}}}(\mathbf{p}) - \mathbf{I}_{\text{query}}(\mathbf{p}) \| + k_{\text{median}} M(\mathbf{p}) \| \mathbf{I}_{L_{\mathbf{p}}}(\mathbf{p}) - \mathbf{I}_{\text{median}(G(L_{\mathbf{p}}))}(\mathbf{p}) \|, \quad (2)$$

where $M(\mathbf{p})$ is a binary mask indicating the target region, and $\overline{M}(\mathbf{p})$ is its logical negation. Outside the target region then, where $M = 0$, the cost depends on the difference between $\mathbf{I}_{L_{\mathbf{p}}}(\mathbf{p})$, the colour of proposal $L_{\mathbf{p}}$ at pixel $\mathbf{p}$, and $\mathbf{I}_{\text{query}}(\mathbf{p})$, the colour of the query image at the pixel. Inside the target region, where $M = 1$, the cost depends on the difference between $\mathbf{I}_{L_{\mathbf{p}}}(\mathbf{p})$ and $\mathbf{I}_{\text{median}(G(L_{\mathbf{p}}))}(\mathbf{p})$, which is the "median" image for that oracle's homography group, $G(L_{\mathbf{p}})$, reconstructed from the median horizontal and vertical gradients of the registered oracles in group, as described in Section 3. This term serves a dual purpose, helping both to avoid inserting new occlusions into the result, and avoid using any proposals outside the region where they are registered, since in both these cases, the deviation $\| \mathbf{I}_{L_{\mathbf{p}}}(\mathbf{p}) - \mathbf{I}_{\text{median}(G(L_{\mathbf{p}}))}(\mathbf{p}) \|$ should be large. Further to this we set $E_1(\mathbf{p}, L_{\mathbf{p}})$ to a large number if proposal $L_{\mathbf{p}}$ does not cover pixel $\mathbf{p}$. This is effectively a hard constraint which prevents a registered oracle being used for regions outside its bounds. The parameters k_{query} and k_{median} weight the terms according to their relative importance.

The purpose of E_2 is to encourage a few large regions to be combined instead of many small ones, and to ensure that boundaries between regions from different proposals occur at places where they will be least obvious. For this we use the "gradient" cost function suggested by Agarwala *et al.* [4], where $E_2 = 0$ if $L_{\mathbf{p}} = L_{\mathbf{q}}$, and otherwise

$$E_2(\mathbf{p}, \mathbf{q}, L_{\mathbf{p}}, L_{\mathbf{q}}) = k_{\text{grad}} \left(\| \nabla \mathbf{I}_{L_{\mathbf{p}}}(\mathbf{p}) - \nabla \mathbf{I}_{L_{\mathbf{q}}}(\mathbf{p}) \| + \| \nabla \mathbf{I}_{L_{\mathbf{p}}}(\mathbf{q}) - \nabla \mathbf{I}_{L_{\mathbf{q}}}(\mathbf{q}) \| \right) \quad (3)$$

where $\nabla \mathbf{I}$ is the image gradient in all colour channels, *i.e.* a 6D vector. This cost penalises boundaries between regions using different proposals where the two proposals' image gradients differ. By encouraging these boundaries to fall in places where the image gradients in the two proposals match well, the transitions between regions are hidden. For the results in this paper we used $k_{\text{query}} = 1$, $k_{\text{median}} = 1$, and $k_{\text{grad}} = 10$, and for computational purposes we used at most 10 proposals from each homography group, ranked by the number of homography inliers.

Finally, in order to achieve a good transition between the region that has been filled and the query image surrounding it, the query image itself is included as a proposal, and the MRF is solved over a larger region than the original target region. By generating the proposals such that they extend outside the target region, the optimisation may choose the best place in this boundary region at which to switch from the query image to the proposals. To optimise the MRF described above, we use tree-reweighted belief propagation [18, 27, 28], using the software made available online [2] by the authors of [27].
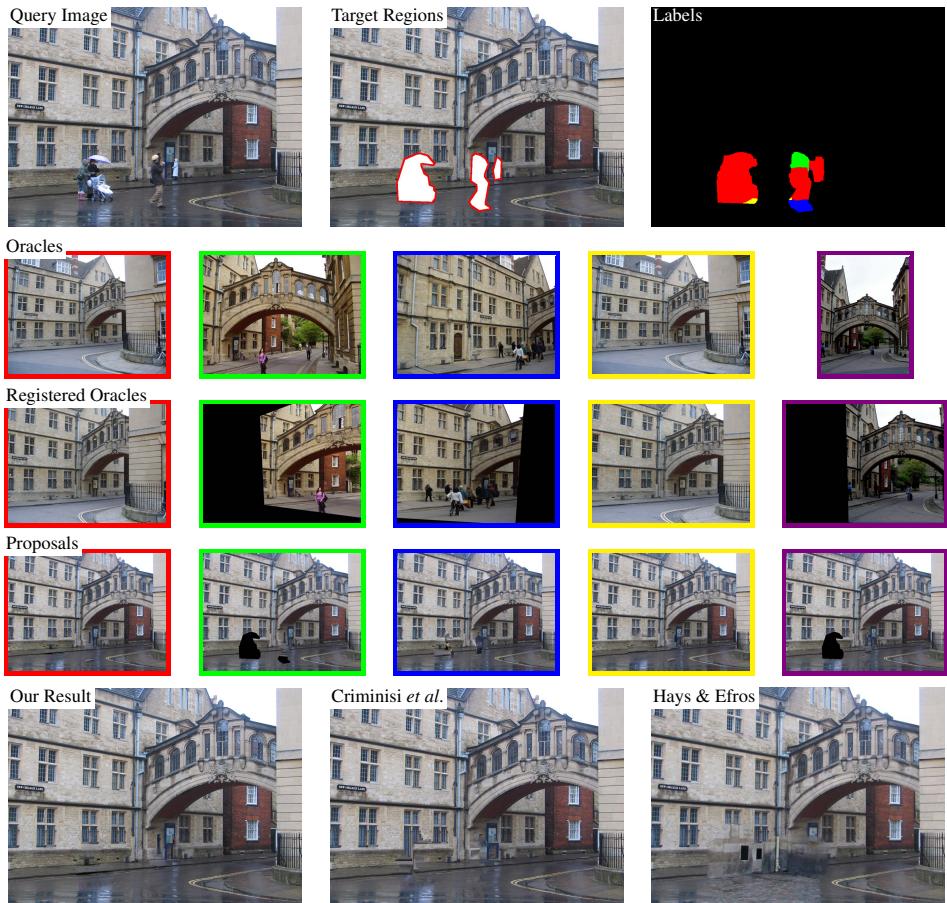
Figure 6: **Combining multiple proposals.** 1st row: The inputs to the system and the output labels showing the combination of proposals in the final result (colours correspond to the borders of the images below). 2nd row: The top 5 original (unregistered) automatically retrieved oracle images used in the result. 3rd row: The geometrically registered oracles. 4th row: The proposals generated by each oracle. Note that none of the individual proposals covers the entire target region and provides a satisfactory result. Last row: Our final result, the result using the algorithm of Criminisi et al. [11] and the result using the method of Hays & Efros [16]. In our result, we obtain a much better image by combining regions from several proposals, and choosing the boundaries to be as inconspicuous as possible. The result of Criminisi et al. propagates strong edges into the target region, but cannot reason about where best to connect them or insert details that are hidden (e.g. the grate at the base of the wall). The method of Hays & Efros produces a result that fits well with the surrounding image, but does not correspond to the actual underlying scene.

5   Results & Discussion

In this section, we demonstrate our method on several query images from the Oxford Buildings image set, with various difficulties which our method is able to overcome. More results can be seen online at [1]. In Figure 1, the region to be replaced is spanned almost entirely by a single scene plane. In this case, the image search returned 49 photos of the same scene,
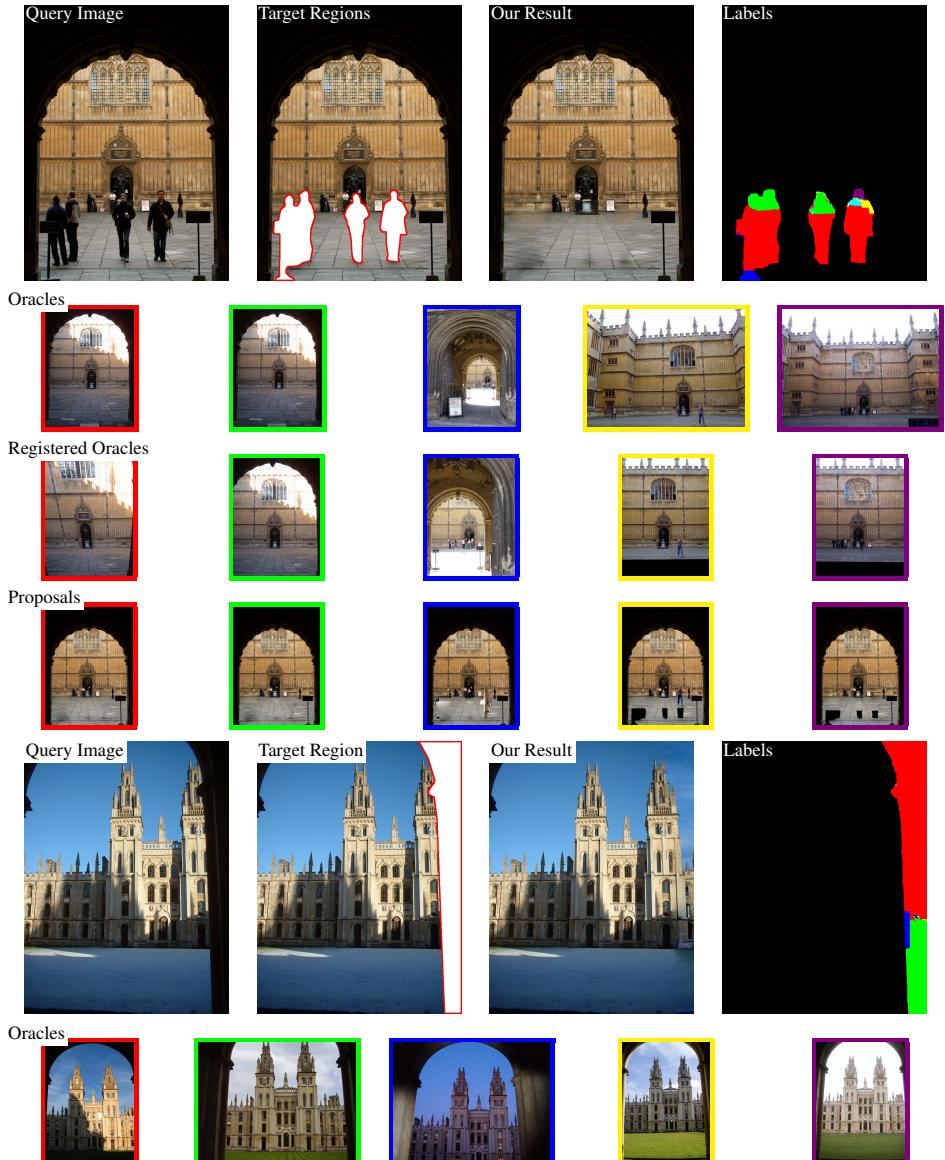
Figure 7: **Results.** Two example queries and the corresponding results from our system. In the first example, the MRF optimisation correctly chooses oracles registered to the ground plane where relevant, and combines other oracles registered to the wall in order to complete the result. In the second example, our method successfully completes the image with the unoccluded building facade, however it cannot reproduce the environmental effects particular to the query image, such as the snow and strong shadows.

of which 15 were accurately geometrically registered to the query image using single homographies. The result comes mostly from a single oracle, which was automatically selected by our MRF formulation since it provided a consistent, occlusion free replacement for the target region.

Figure 6 shows the advantage of our method's ability to combine multiple proposals. The search returned 48 photos of the same scene, of which 9 were registered to the query image, providing 16 proposals due to multiple homographies. Most of the result comes from a single proposal, but the other proposals are used to fill some regions where this fails demonstrating the benefit of using multiple oracles. The final composite obtained by our method is significantly better than results obtained by the approaches of Criminisi *et al.* [11] and Hays & Efros [16] (see the last row of Figure 6).

Figure 7 shows two additional results. In the first example, the search returned 36 results, of which 18 were registered automatically and 2 homographies registering the ground plane were obtained semi-automatically (see Section 3). Thanks to the homography grouping combined with the MRF optimisation most of the occluders were convincingly removed. In the second result, the search returned 47 images of the same scene, 17 of which were accurately registered. Note the successful removal of the occluding arch.

Although we do not show any significant failures of our system here, they tend to occur when we have only a limited number of oracles available. In this case, the median images used to guide the final solution can be misleading, since the assumptions that they will be blurry in unregistered regions, or be free of occlusions are no longer valid. As a result, the final composite image may contain new occlusions or include mis-registered planes.

6   Conclusion

We have demonstrated an inpainting method which is able to combine images taken from different viewpoints and under different lighting conditions to convincingly replace large regions of the query photograph containing complex image structures, thereby significantly extending the capabilities of local patch-based inpainting methods [11, 13].

The approach is mainly applicable to tourist snapshots of frequently visited landmarks, which have many pictures taken by other people available on the Internet. Although results are visually pleasing, in some cases under a close inspection subtle artefacts remain in the final composites. These are mainly due to photometric and resolution issues such as differences in the length of shadows and different image resolution/focus.

Acknowledgements

References

[1] URL http://www.di.ens.fr/willow/research/inpainting/.

[2] URL http://vision.middlebury.edu/MRF/code/.

[3] URL http://www.robots.ox.ac.uk/~vgg/research/oxbuildings/.

[4] A. Agarwala, M. Dontcheva, M. Agrawala, S. Drucker, A. Colburn, B. Curless, D. Salesin, and M. Cohen. Interactive digital photomontage. *ACM Trans. Graphics (Proc. SIGGRAPH 2004)*, 23(3):294–302, 2004.

[5] H. Amirshahi, S. Kondo, and T. Aoki. Photo completion using images from internet photo sharing sites. In *Proc. MIRU*, 2007.

[6] H. Amirshahi, S. Kondo, K. Ito, and T. Aoki. An image completion algorithm using occlusion-free images from internet photo sharing sites. *IEICE Trans. Fundamentals of Electronics, Communications and Computer Sciences*, E91-A(10):2918–2927, 2008.

[7] M. Bertalmío, G. Sapiro, V. Caselles, and C. Ballester. Image inpainting. In *SIGGRAPH '00*, pages 417–424, 2000.

[8] M. Bertalmío, L. Vese, G. Sapiro, and S. Osher. Simultaneous structure and texture image inpainting. *IEEE Trans. Image Processing*, 12(8):882–889, 2003.

[9] P. J. Burt and E. H. Adelson. A multiresolution spline with application to image mosaics. *ACM Trans. Graphics*, 2(4):217–236, 1983.

[10] O. Chum, J. Philbin, J. Sivic, M. Isard, and A. Zisserman. Total recall: Automatic query expansion with a generative feature model for object retrieval. In *Proc. ICCV*, 2007.

[11] A. Criminisi, P. Pérez, and K. Toyama. Object removal by exemplar-based inpainting. In *Proc. CVPR*, pages 721–728, 2003.

[12] A. A. Efros and W. T. Freeman. Image quilting for texture synthesis and transfer. In *SIGGRAPH '01*, pages 341–346, 2001.

[13] A. A. Efros and T. K. Leung. Texture synthesis by non-parametric sampling. In *Proc. ICCV*, pages 1033–1038, 1999.

[14] M. Goesele, N. Snavely, B. Curless, H. Hoppe, and S. M. Seitz. Multi-view stereo for community photo collections. In *Proc. ICCV*, 2007.

[15] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge: CUP, second edition, 2004. ISBN 0521540518.

[16] J. Hays and A. A. Efros. Scene completion using millions of photographs. *ACM Trans. Graphics (Proc. SIGGRAPH 2007)*, 26(3), 2007.

[17] H. Jégou, M. Douze, and C. Schmid. Hamming embedding and weak geometric consistency for large scale image search. In *Proc. ECCV*, 2008.

[18] V. Kolmogorov. Convergent tree-reweighted message passing for energy minimization. *IEEE PAMI*, 28(10):1568–1583, 2006.

[19] V. Kwatra, A. Schödl, I. Essa, G. Turk, and A. Bobick. Graphcut textures: Image and video synthesis using graph cuts. *ACM Trans. Graphics (Proc. SIGGRAPH 2003)*, 22(3):277–286, 2003.

[20] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60:91–110, 2004.

[21] K. Mikolajczyk and C. Schmid. Scale & affine invariant interest point detectors. *IJCV*, 60(1):63–86, 2004.

[22] P. Pérez, M. Gangnet, and A. Blake. Poisson image editing. *ACM Trans. Graphics (Proc. SIGGRAPH 2003)*, 22(3):313–318, 2003.

[23] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman. Object retrieval with large vocabularies and fast spatial matching. In *Proc. CVPR*, 2007.

[24] C. Rasmussen and T. Korah. Spatiotemporal inpainting for recovering texture maps of partially occluded building facades. In *Intl. Conf. Image Proc.*, volume 3, pages 125–128, 2005.

[25] A. Shahrokni, C. Mei, P. H. S. Torr, and I. D. Reid. From visual query to visual portrayal. In *Proc. BMVC*, 2008.

[26] J. Sun, L. Yuan, J. Jia, and H.-Y. Shum. Image completion with structure propagation. *ACM Trans. Graphics (Proc. SIGGRAPH 2005)*, 24(3):861–868, 2005.

[27] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, and C. Rother. A comparative study of energy minimization methods for Markov random fields. In *Proc. ECCV*, volume 2, pages 16–29, 2006.

[28] M. J. Wainwright, T. S. Jaakkola, and A. S. Willsky. MAP estimation via agreement on trees: Message-passing and linear-programming approaches. *IEEE Trans. Information Theory*, 51(11):3697–3717, 2005.

[29] M. Wilczkowiak, G. J. Brostow, B. Tordoff, and R. Cipolla. Hole filling through photomontage. In *Proc. BMVC*, pages 492–501, 2005.