Automating Joiners

Lihi Zelnik-Manor*
Dept. of Electrical Engineering
Caltech

Pietro Perona†
Dept. of Electrical Engineering
Caltech

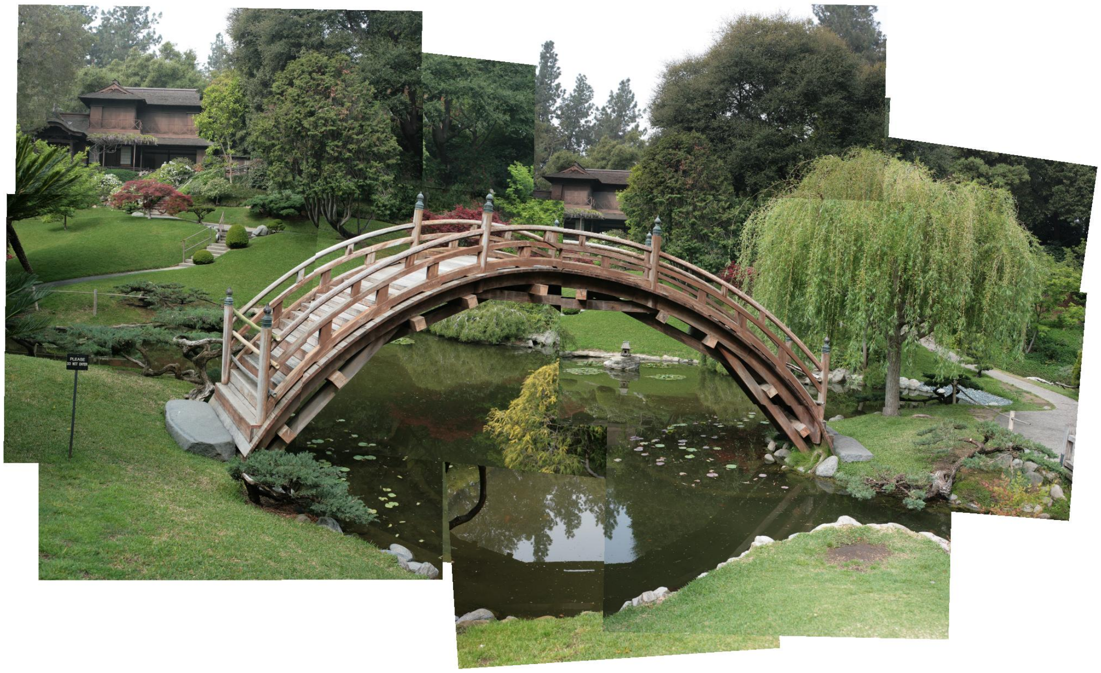


Figure 1: An impossible bridge, but a good memory: Multi-viewpoint joiner of 15 pictures generated by our program.
Project webpage: http://www.vision.caltech.edu/lihi/Demos/AutoJoiners.html

Abstract

Pictures taken from different view points cannot be stitched into a geometrically consistent mosaic, unless the structure of the scene is very special. However, geometrical consistency is not the only criterion for success: incorporating multiple view points into the same picture may produce compelling and informative representations. A multi viewpoint form of visual expression that has recently become highly popular is that of *joiners* (a term coined by artist David Hockney). Joiners are compositions where photographs are layered on a 2D canvas, with some photographs occluding others and boundaries fully visible.

Composing joiners is currently a tedious manual process, especially when a great number of photographs is involved. We are thus interested in automating their construction. Our approach is based on optimizing a cost function encouraging image-to-image consistency which is measured on point-features and along picture boundaries. The optimization looks for consistency in the 2D composition rather than 3D geometrical scene consistency and explicitly considers occlusion between pictures. We illustrate our ideas with a number of experiments on collections of images of objects, people, and outdoor scenes.

NPAR 2007, San Diego, California, August 04–05, 2007.
© 2007 ACM 978-1-59593-624-0/07/0008 $5.00

1 Introduction

A single view may not fully capture a scene as we perceive it. Artists have long known this fact and demonstrated that incorporating multiple view points into the same painting may produce more informative representations than a single viewpoint painting can. For example, see Figure 4.a[1]. This fresco by Paolo Uccello shows the podium as if the viewer is looking upward to it, yet the rider and the horse are painted from a direct side view. For the same reason it has become common among photographers to take multiple pictures of the same scene and compose them into mosaics. When all the pictures are taken from a single view point the geometry of the panorama is well understood [Hartley and Zisserman 2000; Szeliski and Shum 1997; Zelnik-Manor et al. 2005]. Methods for matching informative image features [Lowe 2004] and good blending tech-

Keywords: multiple view point, panorama, mosaic, joiner, composition

*e-mail: lihi@vision.caltech.edu
†e-mail:perona@vision.caltech.edu

[1] The image of the work of art of Paulo Uccello appearing in this paper is in the public domain worldwide. The reproduction is part of a collection of reproductions compiled by The Yorck Project. The compilation copyright is held by The Yorck Project and licensed under the GNU Free Documentation License.

niques [Burt and Adelson 1983; Debevec and Malik 1997; Perez et al. 2003] have made it possible for any amateur photographer to produce automatically smooth panoramas covering very wide fields of view starting from collections of photographs [Brown and Lowe 2003; Szeliski and Shum 1997; Szeliski 2006].

While the painter has the freedom to change the view point smoothly and to select which aspects of the scene will be represented, this is not always possible for the photographer who captures the world in single viewpoint snapshots. When the point of view changes or when objects move in the scene, no consistent photo mosaic may be obtained, unless the structure of the scene is very special. A photo mosaic will, thus, inevitably look 'blocky' because of inconsistencies between photographs. Nevertheless, artists like David Hockney and Gordon Matta-Clark have demonstrated that fragmented compositions can look compelling and informative. Compositions, where photographs are layered on a (digital) 2D canvas, with some photographs occluding others and boundaries fully visible were named "joiners" by Hockney. Such joiners have become popular also among amateur photographers and numerous examples may be found online (e.g. in Flickr.com, search with keywords "joiners", "Hockney", "photocollage").

In spite of their fragmented look, joiners can be compelling because they contain multiple viewpoints and time instants. They have thus become a popular form of visual expression. Besides their artistic value, joiners might be seen as a new way for exploring image collections, where photographs are organized spatially, rather than by file names (see Figure 2 and a preliminary demo at: http://www.vision.caltech.edu/lihi/Demos/AutoJoiners.html). Additionally, there are scenarios in which multi viewpoint mosaics must be used because there is no other option. For example, often one cannot capture the full scene from a single view point due to occlusions, see Figure 5. Changes in view point can also result from people moving while being photographed, see Figure 3. While joiners can be (and are being) composed manually, this is a tedious process, especially when a great number of photographs is involved. We are thus interested in building automated tools for simplifying their construction.

Visual representations incorporating multiple view points have been explored before. Wood et al. [1997] suggested an approach to computerized design of multiperspective panoramas for cel animation where all viewpoints are available apriori. A wide range of approaches have been suggested for constructing multi view panoramas when the input is a video sequence taken by a smoothly moving video camera, e.g., [Rademacher and Bishop 1998; Peleg et al. 2001; Zomet et al. 2003; Li et al. 2004]. Agarwala et al. [2006] generate multi-viewpoint panoramas of long, roughly planar scenes, such as the facades of buildings along a city street. [Rother et al. 2006; Rother et al. 2005; Wang et al. 2006] construct visually appealing collages from collections of independent images.

Our goal, however, is to create compositions of discrete sets of photographs taken from different view points of general scenes. We do assume that the photographs in a collection relate to a spatially continuous experience of a visual scene (i.e., the images form a single connected component). Additionally, we assume that there is a topology to the visual experience of the photographer: despite the fact that, during his exploration, the photographer visited a number of different viewpoints, there is a notion of 'left', 'right', 'up' or 'down' and any pair of images will have relative positions in space. While, we cannot obtain a geometrically consistent composition, we do wish to obtain one which represents this topology.

Our motivation is two fold. For artists: Rather than keeping busy with manual alignment of images we wish to facilitate the artists' work by providing with an organized starting point to be refined manually. For non-artists and data exploration: Generate representations of image collections that are pleasant to the eye and are easily readable. We argue that for both purposes a good solution is one obeys the following principles:

- It is a layering of the pictures on a 2D canvas (a joiner).
- This arrangement should respect and convey the topology of the photographer's visual experience.
- It should show as much information as possible, minimizing redundancy.
- Each photo is an "object" and should be respected as such (no distortions).
- Inconsistencies at photo-to-photo transitions should not be distracting, and should be minimized if possible.

In order to recover the topology and minimize redundancy we exploit correspondences between portions of pictures. To avoid distortions images may undergo only rotation, translation and scaling. Finally, we minimize distracting visual artifacts by layering the pictures in the most consistent order.

We acknowledge that the evaluation of the results presented in this paper is not straightforward. Clearly, the assessment of the quality of a joiner is subjective. We evaluate the results according to the aforementioned goals and request the reader to do the same.

The rest of the paper is organized as follows. We start by outlining the overall framework in Section 2. We then proceed and describe in detail the various steps of the approach in Sections 3,4,5,6,7. We conclude in Section 8. Our ideas are illustrated through experiments which appear throughout the text.

2 Overall Framework

When pictures are taken from different view points there is no globally consistent geometric solution to align all of them. Nevertheless, we wish to obtain a 2D composition that represents the topology of the underlying scene while minimizing redundancies. To achieve that, we exploit correspondences between portions of pictures.

To minimize distracting visual artifacts in the composition our scheme optimizes appearance consistency directly on the 2D composition plane. Note, that appearance consistency can often be achieved even for pictures that are geometrically inconsistent as they may easily blend into each other, e.g., when there is texture or uniform color near the picture boundary. Alignment errors in these cases are more acceptable than when the errors are salient. Furthermore, geometrical and appearance inconsistencies that are hidden from view have little importance, as compared to those that are visible. Our optimization takes this into account.

The suggested framework consists of the following steps:

1. For each pair of images find point-feature correspondences and fit a similarity transformation between them [Umeyama 1991]. Keep only correspondences which can be approximately aligned by the transformation (Section 3).
2. Allow the user to add/remove matches and/or mark important regions (Section 6).
3. Find global alignment of the images in the composition by minimizing distances between correspondences. If importance weights were assigned to the correspondences, incorporate them in the optimization process (Section 3).

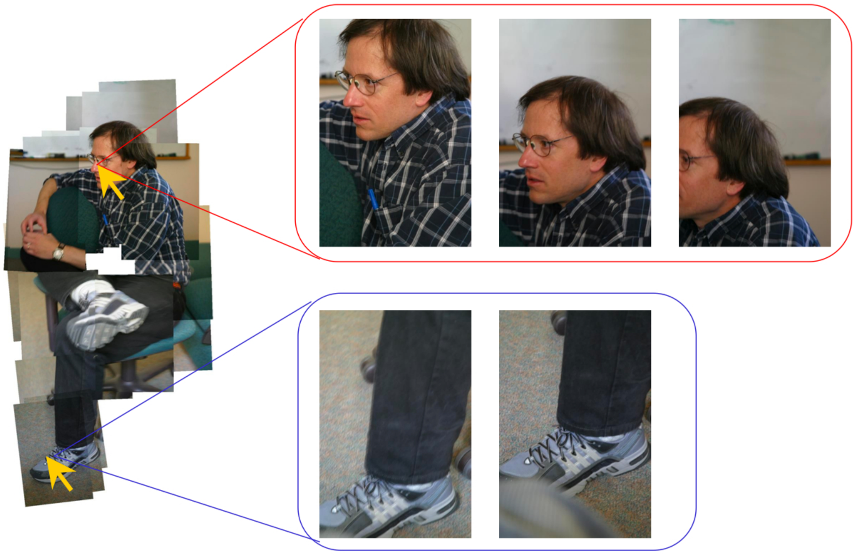


Figure 2: **Organized memories:** *Joiners can be useful for exploring image collections. Clicking on a spot in the joiner will display all pictures overlapping with it.*

4. Find the best layering of the images: search over all possible orders the one which minimizes discontinuities across image boundaries in the composition (Section 4).

5. Assign high weights to correspondences near visible image boundaries and low weights otherwise (Section 5).

6. Repeat steps 2 to 5 until weights and transformations are not updated.

7. If desired, blend images only near visible seams (Section 7).

The user interaction in step 2 is optional and without it the system is fully automatic. In the following sections we describe in detail each of the above steps.

3 Image Alignment

For image alignment we adopt the feature-based technique suggested by Brown & Lowe [Brown and Lowe 2003], with two major differences. In [Brown and Lowe 2003] images were assumed to be taken from a single viewpoint, implying a geometrically consistent panorama. Hence, alignment was obtained on the viewing sphere by solving for the camera rotation at each image and all features had equal contribution. This approach is inadequate for images taken from multiple view points. Rather than optimizing the alignment on the sphere we optimize it directly on the composition canvas. Furthermore, as one cannot expect all feature matches to be nicely aligned they are assigned importance weights. Important features will be well aligned while others are allowed to have larger errors.

We optimize the alignment on the 2D composition canvas by solving for a similarity transformation for each image [Umeyama 1991]. That is, we allow images to translate, scale and rotate. The choice of similarities is motivated by the principles presented in Section 1, which were directed by the beautiful compositions we have found on the web, as well as by our own experience in manually creating joiners.

Following Brown & Lowe, we first extract and match SIFT features [Lowe 2004] between all pairs of images. We then use RANSAC [Hartley and Zisserman 2000] to select a set of inliers that are compatible with a similarity transformation between each pair of images. Next we apply the probabilistic model suggested in [Brown and Lowe 2003] to verify the match. We discard all feature matches which are not geometrically consistent with the transformation between the images (RANSAC outliers). Finally, given the set of geometrically consistent matches, we use bundle adjustment [Brown and Lowe 2003] to solve for all of the transformations jointly.

Unlike the single view point case, when the images are taken from multiple view points one cannot expect all the matches to be nicely aligned. Assigning the same importance to all matches (as was done in [Brown and Lowe 2003] for the single view point case) will result in misalignments distributed across the whole panorama. Instead, one would like "important" matches to be well aligned while allowing other matches to have larger errors. This can be achieved by assigning each feature match with a weight indicating its importance. The decision on which features are "important" and the setting of the weights will be described in Section 5.

The objective function of the optimization process is thus a weighted sum of projection errors.

Let u_i^k denote the k'th feature in image i and S_{ij} a similarity transformation between images i and j. Given a feature match $u_i^k \leftrightarrow u_j^l$ the corresponding residual is: $r_{ij}^{kl} = u_i^k - S_{ij} u_j^l$ and the assigned weight is denoted by w_{ij}^{kl}. The error function to be minimized is the sum over all images of the weighted residual errors:

$$e = \sum_{i=1}^{n} \sum_{j \in \mathcal{N}(i)} \sum_{k,l \in \mathcal{F}(i,j)} w_{ij}^{kl} f(r_{ij}^{kl}) \tag{1}$$

where n is the number of images, $\mathcal{N}(i)$ is the set of images with feature matches to image i, $\mathcal{F}(i,j)$ is the set of feature matches between images i and j and $f(x)$ is a robust error function:

$$f(x) = \begin{cases} |x| & \text{if } |x| < x_{max} \\ x_{max} & \text{if } |x| \geq x_{max} \end{cases} \tag{2}$$

This robust error function is used to minimize the impact of erroneous matches. As suggested in [Brown and Lowe 2003] we use $x_{max} = \infty$ during initialization and $x_{max} = 5$ pixel for the final solution. This is a non-linear least squares problem which we solve using the Levenberg-Marquardt algorithm[2].

4 Ordering Images

Imperfect alignment will unavoidably result in blurry regions when blending the images. Thus, instead of blending the images we wish to order them into layers such that images placed on top will hide misalignments underneath. This will leave us with visible artifacts only along image boundaries which are not occluded. We will refer to these as "visible image boundaries". Our goal is to find an order of the images which minimizes appearance inconsistencies across the visible image boundaries.

One can adopt two approaches to order the images:

1. Assign each image to a separate layer and find the best order of layers. This is equivalent to what can be easily done in most image editing softwares, e.g., Photoshop.
2. Select a local order of the images separately in each overlap area. For example one could have image A above B, B above C and C above A in different regions of the composition.

The second option is more flexible since decisions can be taken locally at each overlap region, however, as such, it is more likely to produce over-fragmented compositions. Moreover, one may wish to manually refine our automatic result using a standard editing software. We thus chose to focus on the first option and left the second one outside the scope of this paper.

Exact optimization Given an alignment of the images on the composition plane, finding the best order of images can be formulated as a graph problem. Let $G = (V, E)$ be an undirected graph where each node $v_i \in V$ represents an image and edges connect between images that overlap. A valid order of the images can be represented by an acyclic orientation of the graph edges. The set of all acyclic orientations of the edges of G represents all possible orders of the images. It can be found in overall time $O((n+m)\alpha)$ [Barbosa and Szwarcfiter 1999], where n is the number of nodes (images), m is the number of edges and α is the number of acyclic orientations.

We then perform an exhaustive search over all possible orders and select the best one. For each order of the images we compute a cost based on image-to-image consistency measured along visible image boundaries, denoted by $\mathcal{B}$. One can design many such cost functions. We have experimented with three:

1. Sum of gradients across image boundaries: $Cost_{grad} = \sum_{x,y \in \mathcal{B}} P_x^2(x,y) + P_y^2(x,y)$, where P_x and P_y are the horizontal and vertical numerical derivatives of the composition image.
2. Sum of color differences between overlapping images: $Cost_{color} = \sum_{x,y \in \mathcal{N}(\mathcal{B})} (I_{top}(x,y) - I_{scnd}(x,y))^2$, where I_{top} and I_{scnd} are the top and second from top images on one side of the visible boundary $\mathcal{B}$ and $\mathcal{N}(\mathcal{B})$ is a region around the boundary.
3. Quality of curve continuation: We first find curves of length ≥ 5 pixels in all images[3] and project them to the panorama plane. We then find the set of curves C which intersect visible image boundaries and are visible in the panorama (i.e., are not occluded by other images). For each such curve $c \in C$ we find the closest curve $\tilde{c}$ on the other side of the boundary. We fit a line to the last 3 pixel-long bits of both curves. Denote by $L(c, \tilde{c})$ the sum of squared distances between the curves and the fitted line. The curve continuation cost is defined as: $Cost_{curve} = \sum_{c \in C} \min(L(c, \tilde{c}), \tilde{L})$, where $\tilde{L}$ is a penalty for curves whose continuation could not be found.

In our experiments we found that in most cases minimizing $Cost_{grad}$ or $Cost_{curve}$ provided comparable results, better than those using $Cost_{color}$. For consistency in the presentation of the paper, all the presented results were obtained by minimizing the gradient-based cost $Cost_{grad}$.

Approximate solution Clearly, for large datasets the number α of possible orders is too large to test all. To overcome this limitation one has to adopt some heuristics. One possibility is trying just a limited number of random orders and keeping the best one. Alternatively, one can start from a small set of random orders and search around each one by performing a small number of order flips between images. We have experimented with both and found them to often provide good results, yet different executions of the program could result in different outcomes, varying in their quality.

An important observation is that typical image collections aim at covering the scene. Images are usually relatively spread and each one overlaps with only a few others. This implies that many of the order decisions can be taken locally. We thus adopt the following procedure:

- Initialize order according to temporal acquisition order.
- Fix order of all but one image and compute the cost for all relevant orders, i.e., changing only the order-position of the free image. Accept the minimal cost order.
- Iterate over all images until no further updates in order.

The number of orders one needs to consider at each iteration equals the number of images overlapping with the free-to-move image. For example, starting from order $[1, 2, 3, \ldots, N]$ we fix all images but image 1. If image 1 overlaps only with image 2 then we need to consider only the orders $[1, 2, 3, \ldots, N]$ and $[2, 1, 3, \ldots, N]$. All other orders are equivalent with respect to image 1 since it does not overlap with them, i.e., the cost of $[2, 1, 3, \ldots, N]$ and $[2, 3, 1, \ldots, N]$ is the same. In all of our experiments the procedure ended after 2-4 passes over all images. The number of orders that were explored was significantly smaller (often by orders of magnitude) than that of exhaustive search, nevertheless, results are satisfactory. For consistency in presentation, this procedure was used to obtain all the results presented in this paper. Figures 10 and 3 show how appearance inconsistencies can be minimized by layering images according to this framework.

5 Iterative Refinement

The approach we adopted layers the images in the joiner so that parts of the images are occluded. This leaves inconsistency artifacts in the panorama only along visible image boundaries. We thus wish for the alignment to be of high quality along those seams while we can afford it to be sloppier in occluded regions. This is achieved by iterative refinement of the alignment and order of images.

Given an initial alignment and order of images we assign weights to feature matches according to their "importance". Matches near visible image boundaries are assigned high weights while far-from-

[2] Note, that since we solve for a similarities rather than 3D camera rotations, the derivatives of the cost of Eq. (1) with respect to the parameters of the transformations are different from those described in [Brown and Lowe 2003]. Details are omitted due to lack of space.

[3] We used a software written by the Oxford Visual Geometry Group based on Canny edge detection.

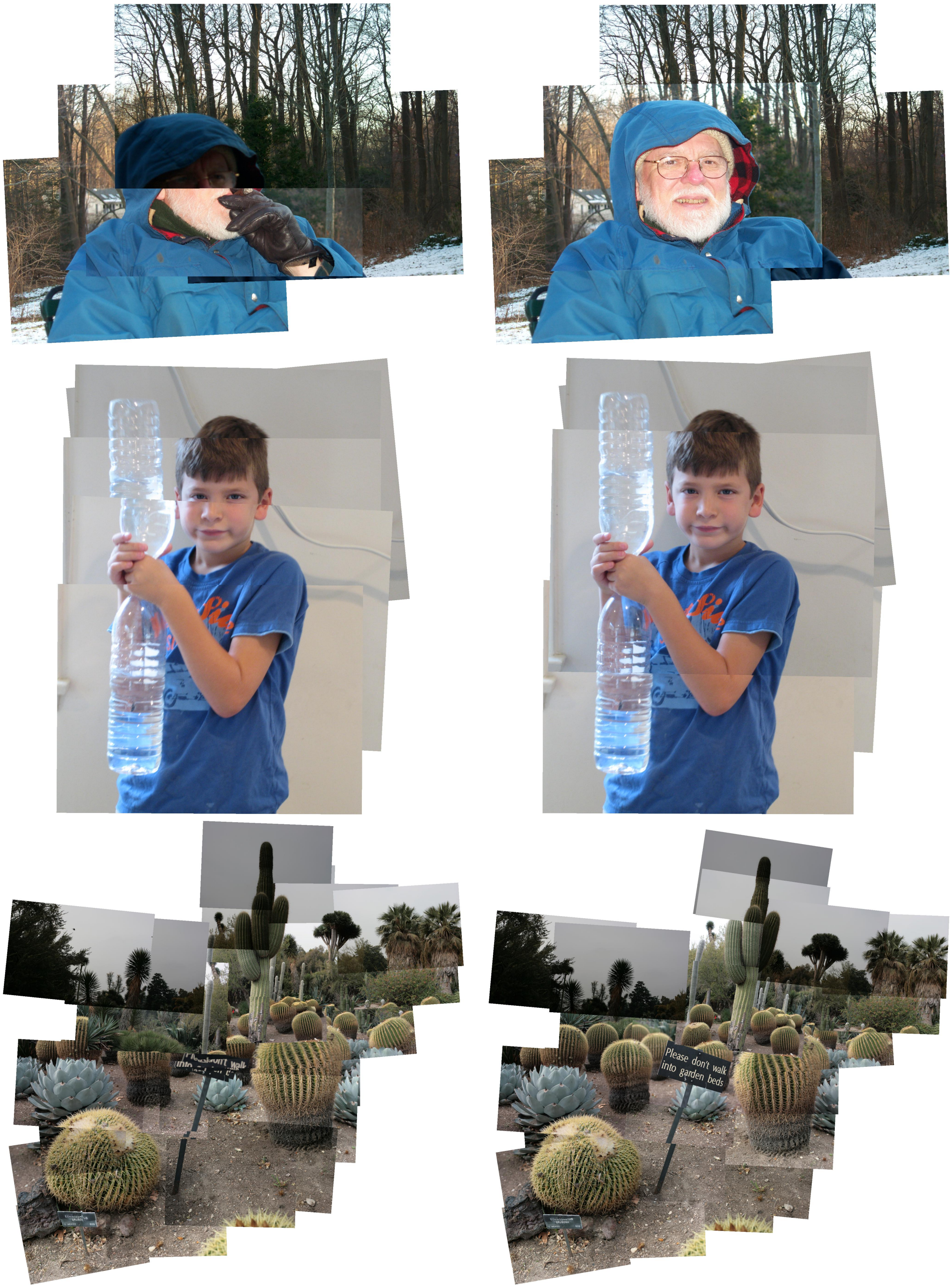


Figure 3: **Iterative refinement:** *Left: Initial alignment of images with equal weights to all features and layered according to acquisition order. The compositions are over fragmented, e.g., the sign in the cacti garden is unreadable and the person's face at the top row is hidden. Right: Final result after iterative refinement of alignment and order. By minimizing inconsistencies better representations of the scene are obtained. The man's face is fully visible and the sign in the cacti garden is readable. For better comparison we propose looking at these on screen in full resolution. These collections include 4, 5 and 33 pictures, respectively.*

boundary matches are given low weights:

$$w_{ij}^{kl} = MAX(\exp^{-MIN(d^2(u_i^k,\mathcal{B}),d^2(u_j^l,\mathcal{B}))/\sigma^2}, \omega) \qquad (3)$$

where $d^2(u_i^k, \mathcal{B})$ is the minimum distance between feature u_i^k and the visible image boundaries $\mathcal{B}$. The parameter σ controls the rate of decay of the exponential function and ω defines the minimum weight of a feature. In all our experiments we used $\sigma = 50$ pixels and $\omega = 0.1$.

We obtain a refined alignment by applying the bundle adjustment procedure of Section 3 while incorporating the assigned weights. Given the new alignment the images are ordered again and weights are reassigned according to the result. This process is iterated until convergence. In our experiments we applied 3 iterations. Figures 10 and 8 compare alignment results with equal weights assigned to all feature matches (i.e., $w_{ij}^{kl} = 1 \ \forall i, j, k, l$) and those obtained with importance weights. The latter minimizes inconsistencies. Figures 4 and 8 show joiners constructed automatically by the proposed iterative process.

6 User Interaction

The automatic approach suggested in the previous sections can successfully join many image collections. Nevertheless, at times it fails. Feature matching is the main difficulty. When seen from highly different view points, feature appearance changes significantly and matching of corresponding features becomes more difficult and often fails [Moreels and Perona 2007]. This can result in too few matches between overlapping images, or even none at all. Another difficulty is that sometimes foreground and background indicate different alignments and the choice between them is subjective. For example, one could chose the align a person standing close to the camera, or the faraway background. Due to parallax one cannot hope to align both. We have thus developed a user interface to allow users to direct and assist the panorama construction in such difficult cases.

Currently, the user interface accepts two types of input. The first lets the user mark manually corresponding points between image pairs. Since we consider only similarity transformation two point matches suffice to align a pair of images. The second allows marking regions of importance. Feature matches within an important region are assigned the maximal weight 1, while matches outside the important region are marked as least important and are assigned to the minimal weight of 0.1. The matches and weights provided manually are used to update the automatically computed ones, and then all are used in the iterative align and order scheme. Alternatively, one could force important regions to be visible. Figures 5,, 6, 9 11 show what can be achieved with user interaction. Minimal interaction (a few mouse clicks) was sufficient.

7 Blending

After aligning and layering the images, artifacts are left only along visible image boundaries. At this point one can choose between three options, depending on individual taste: (i) leave the joiner as is with image boundaries clearly seen, (ii) further emphasize the seams by adding a frame to all the pictures, and (iii) try and remove the visible seams by blending the images. The first two are commonly adopted by artists and require no further effort. Blending is more tricky. Blending all the images, as is done in the single view case [Brown and Lowe 2003], is undesirable since it will make the hidden misalignments appear within a blurred composition (see, for example, Figure 7.a). Alternatively, one could apply a graph-cut approach, e.g., [Agarwala et al. 2004]. This is typically better than global blending and can produce sharp and seamless results, yet, when large misalignments are present it often results in a fragmented composition with irregular boundaries that cut through meaningful objects (see Figure 7.b). As we want to maintain the photographic experience we rejected this option. Instead, we apply blending only along visible image boundaries and use only the top and second from top layers. When the alignment quality is high along the visible image boundaries this removes seams while not introducing blurriness, see Figures 7.c, 4.d and 9. In our experiments we used the multi-band blending approach suggested in [Burt and Adelson 1983].

8 Discussion and Conclusions

In this work we have shown that automation of mosaic construction is possible for pictures taken from multiple different view points. Our approach was motivated by the work of photographers and artists who have proved that stitching images taken from multiple view points is not an impossible task. Their compositions are compelling despite inconsistencies at image boundaries. The automatic construction of joiners was achieved by replacing the traditional geometrical consistency requirement with consistency in the image plane. Rather than opting for a globally consistent solution we aimed at minimizing visible artifacts together with hiding large misalignments.

Nevertheless, there are still many open problems. The main difficulty was found to be feature matching. When seen from highly different view points, feature appearance changes significantly and matching of corresponding features becomes more difficult and often fails [Moreels and Perona 2007]. This can result in too few matches between overlapping images, or even none at all. Currently, we have solved this by requesting assistance from the user. Future work will look into refined matching techniques under extreme viewpoint changes.

Acknowledgements

We wish to thank David Lowe and Maneesh Agrawala for reading earlier versions of this paper and providing useful comments, Matthew Brown for his Autostitch software, the Oxford Visual Geometry Group for their edge detection code. This research was supported by MURI award number AS3318 and the Center of Neuromorphic Systems Engineering award EEC-9402726.

References

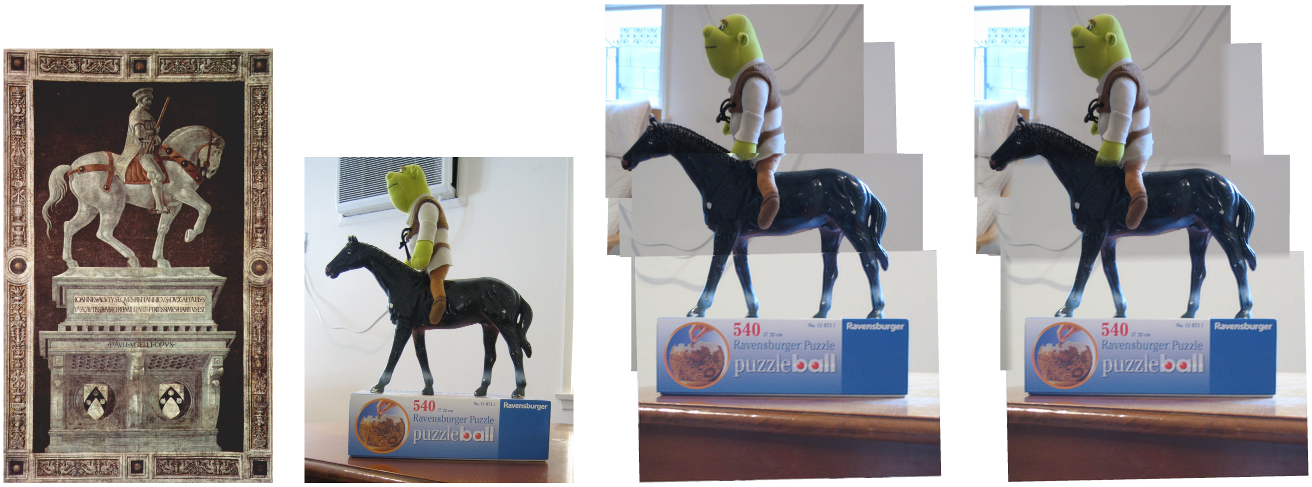
AGARWALA, A., DONTCHEVA, M., AGRAWALA, M., DRUCKER, S., COLBURN, A., CURLESS, B., SALESIN, D., AND COHEN, M. 2004. Interactive digital photomontage. In *Proceedings of SIGGRAPH*.

AGARWALA, A., AGRAWALA, M., COHEN, M., SALESIN, D., AND SZELISKI, R. 2006. Photographing long scenes with multiviewpoint panoramas. In *Proceedings of SIGGRAPH*, 853–861.

BARBOSA, V. C., AND SZWARCFITER, J. L. 1999. Generating all the acyclic orientations of an undirected graph. *Inf. Process. Lett. 72*, 1-2, 71–74.

BROWN, M., AND LOWE, D. 2003. Recognising panoramas. In *Proceedings of the 9th International Conference on Computer Vision*, vol. 2, 1218–1225.

BURT, P. J., AND ADELSON, E. H. 1983. A multiresolution spline with application to image mosaics. *ACM Trans. Graph. 2*, 4, 217–236.



(a) (b) (c) (d)

Figure 4: **Multiple view points:** *(a) "Funerary Monument to Sir John Hawkwood" by Paolo Uccello, 1436. Uccello gave the viewer the impression of standing below the pedestal, thus creating a more monumental effect, but at the same time showed the horse and rider from the side providing a better viewpoint of them (this reproduction is in public domain, see footnote). (b) A picture of our simulation of Uccello's setup represents what can be seen from a single view point. Viewing the "pedestal" from a side view resulted in viewing the head of the character from below and the nasty grill on the wall behind cannot be avoided. (c,d) Our composition result of 5 pictures without and with local blending. As in Uccello's fresco our composition incorporates multiple view points. The "pedestal" is seen from a direct side view, the horse is seen both from below (showing his belly) and from above (showing the top of his mane) while the rider is again seen from a complimenting side view.*

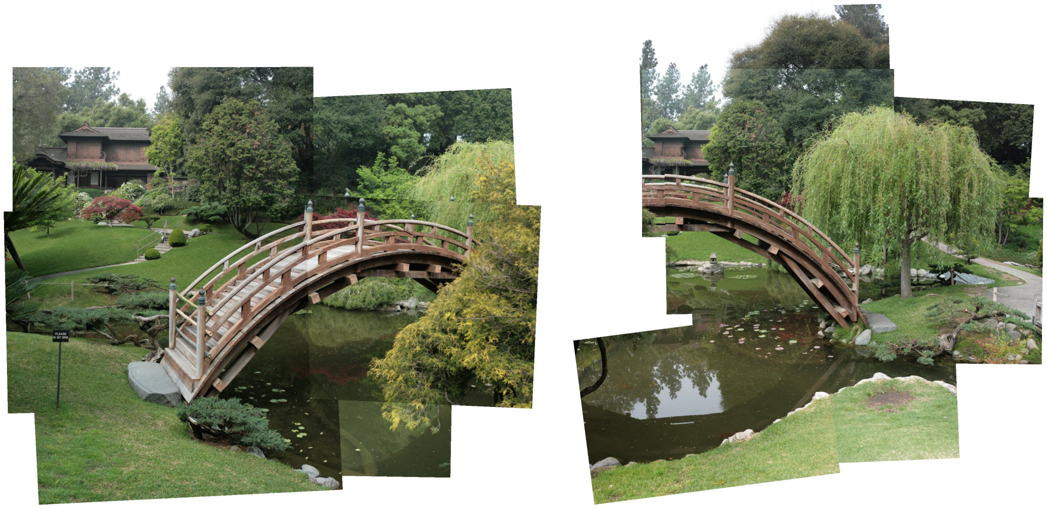


Figure 5: **The impossible bridge:** *Automatically generated joiners of images of a bridge taken from two highly different view points due to a tree occluding part of the bridge (the tree can be seen on the right end of the left joiner). Due to the large change in view point no matches were found across joiners. By manually marking two point correspondences on a single pair of images, all the images were aligned, resulting in the single joiner of Figure .*

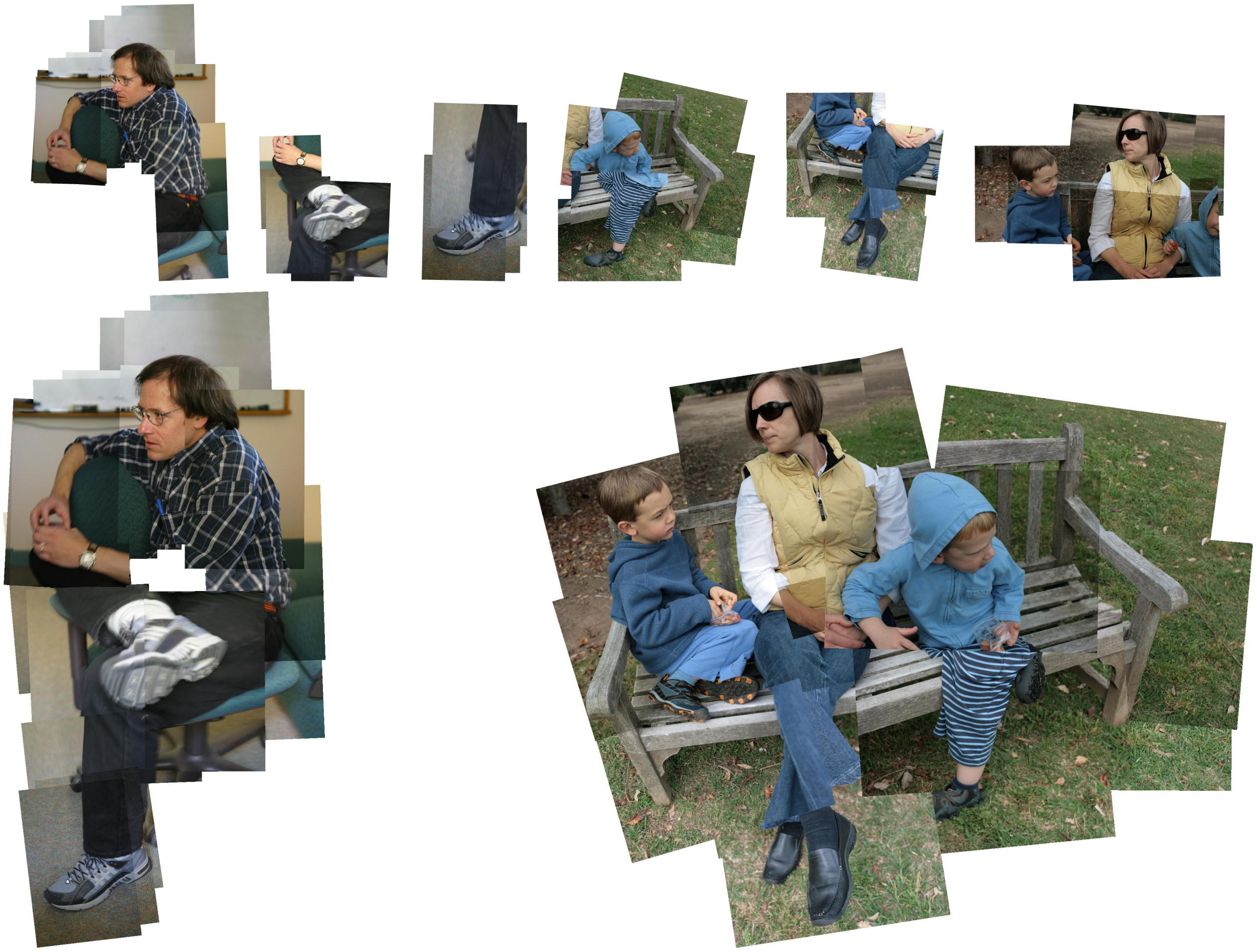
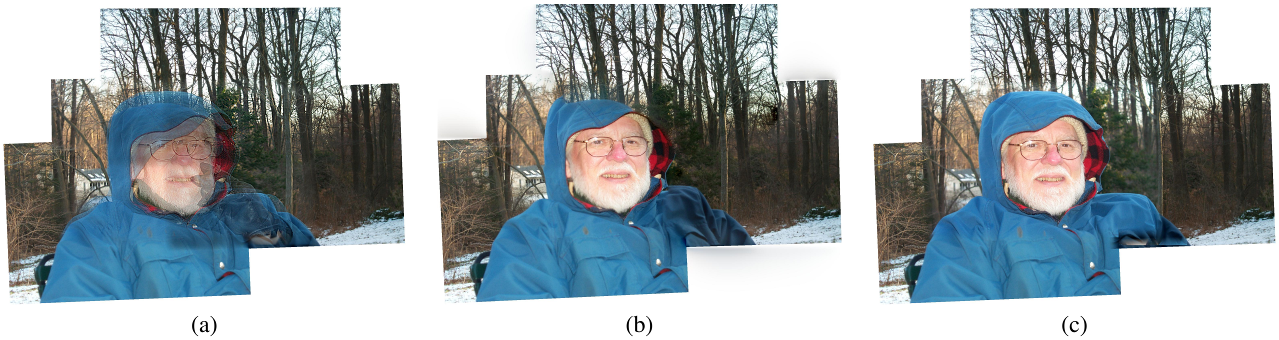


Figure 6: **Incorporating user input:** *Top row: automatically constructed joiners. Each collection got split into three subsets due to lack of feature correspondences. Bottom row: With minimal user interaction, the pieces were merged into single compositions. For each image collection, two point correspondences were manually marked between the left and center joiners, and two between the right and center joiners. The final joiners include 21 and 18 pictures, respectively.*



(a) (b) (c)

Figure 7: **Blending.** *Due to the large changes in view point standard merging techniques often fail. (a) Global multi-band blending results in a blurry joiner. (b) Graph-cuts can cut through meaningful objects (here the hood got cropped) and requires user interaction to produce consistent results. (c) Result with ordering and local blending typically looks best.*

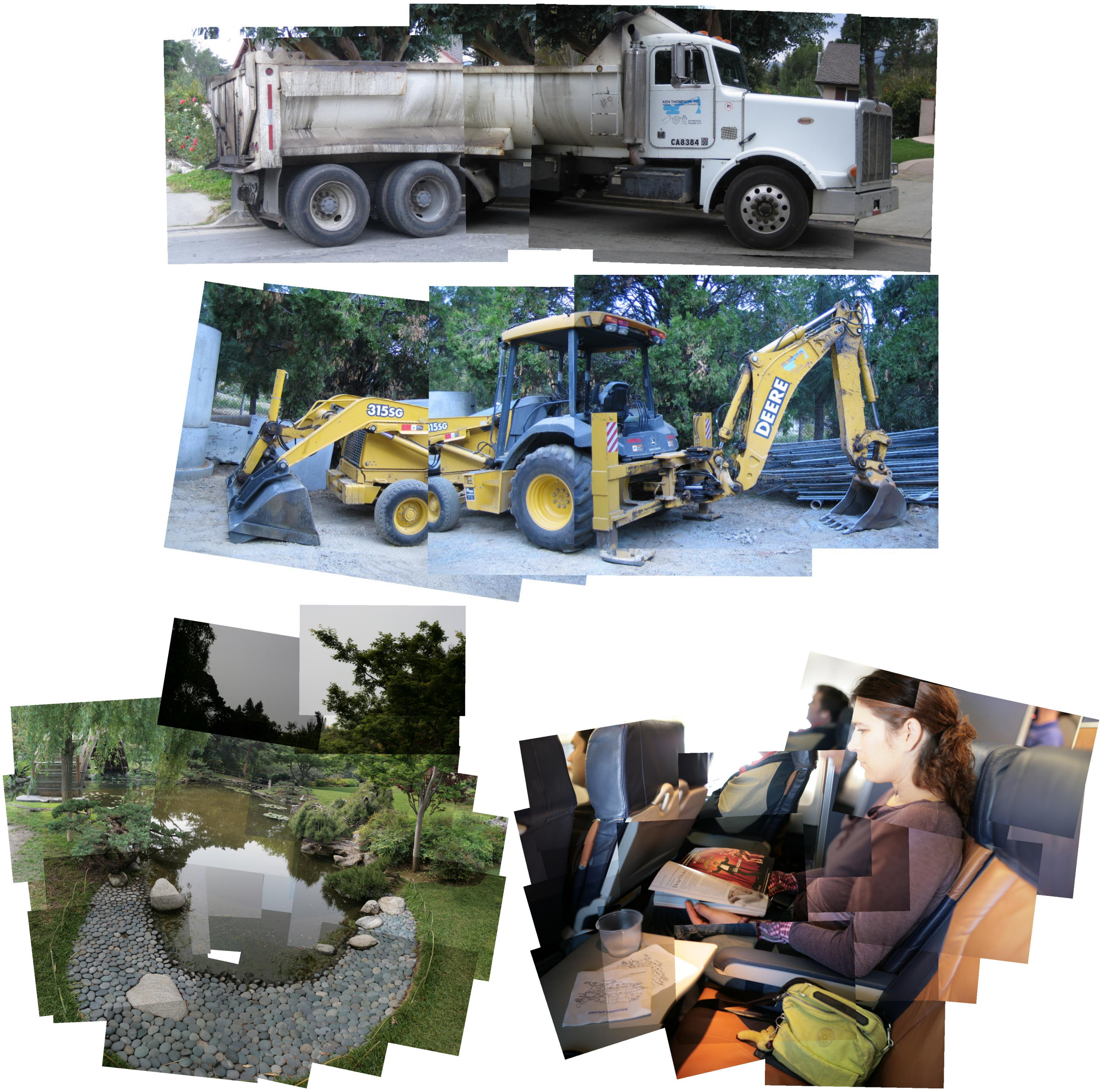


Figure 8: **Automatically generated joiners.** The truck joiner combines 7 pictures, the tractor 5, the pond 37, and the reading lady 22 pictures.

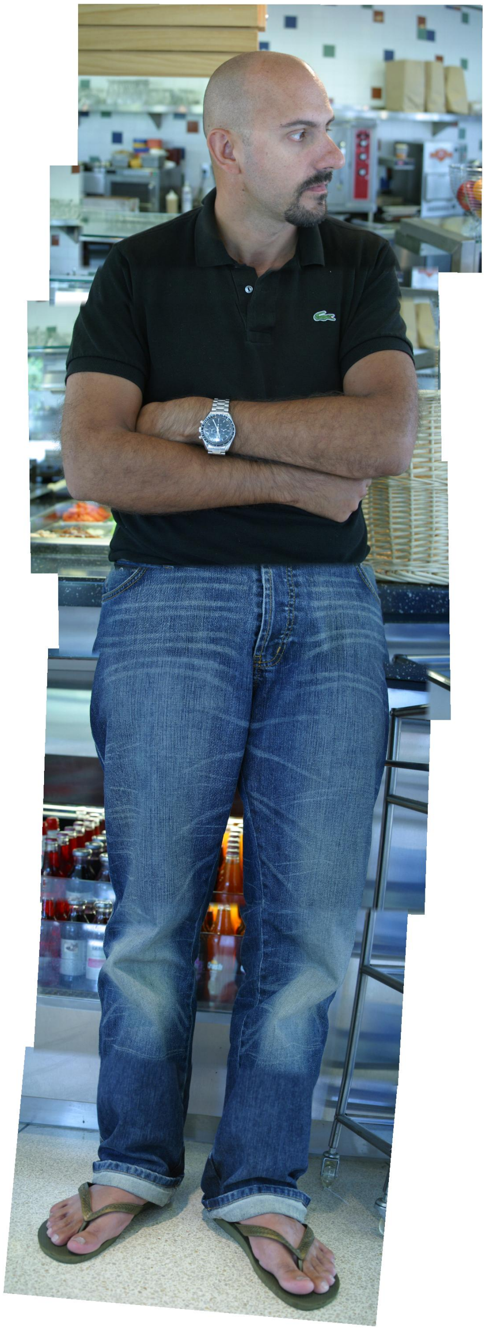


Figure 9: **Homage to David Hockney's LA Visitors series.** *This one required some user interaction due to lack of matches. The final result (after local blending) is almost seamless. The large changes in viewpoint are noticeable on the background. The joiner combines 7 pictures.*

DEBEVEC, P. E., AND MALIK, J. 1997. Recovering high dynamic range radiance maps from photographs. In *Proceedings of SIGGRAPH*.

DEBEVEC, P. E., TAYLOR, C. J., AND MALIK, J. 1996. Modeling and rendering architecture from photographs. In *Proceedings of SIGGRAPH*.

HARTLEY, R. I., AND ZISSERMAN, A. 2000. *Multiple View Geometry in Computer Vision*. Cambridge University Press.

LI, Y., SHUM, H. Y., TANG, C. K., AND SZELISKI, R. 2004. Stereo reconstruction from multiperspective panoramas. *IEEE Trans. on Pattern Analysis and Machine Intelligence 26* (January), 45–62.

LOWE, D. G. 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision 60*, 2, 91–110.

MOREELS, P., AND PERONA, P. 2007. Evaluation of features detectors and descriptors based on 3d objects. *International Journal of Computer Vision 73*, 3 (July).

PELEG, S., BEN-EZRA, M., AND PRITCH, Y. 2001. Omnistereo: Panoramic stereo imaging. *IEEE Trans. on Pattern Analysis and Machine Intelligence 23* (March), 279–290.

PEREZ, P., GANGNET, M., AND BLAKE, A. 2003. Poisson image editing. In *SIGGRAPH '03: ACM SIGGRAPH 2003 Papers*, ACM Press, New York, NY, USA, 313–318.

RADEMACHER, P., AND BISHOP, G. 1998. Multiple-center-of projection images. In *Proceedings of SIGGRAPH*.

ROTHER, C., KUMAR, S., KOLMOGOROV, V., AND BLAKE, A. 2005. Digital tapestry. In *IEEE Conference on Computer Vision and Pattern Recognition*.

ROTHER, C., BORDEAUX, L., HAMADI, Y., AND BLAKE, A. 2006. Autocollage. In *SIGGRAPH '06: ACM SIGGRAPH 2006 Papers*, ACM Press, New York, NY, USA, 847–852.

SZELISKI, R., AND SHUM, H. 1997. Creating full view panoramic image mosaics and environment maps. *Computer Graphics 31*, Annual Conference Series, 251–258.

SZELISKI, R. 2006. Image alignment and stitching: A tutorial. *Foundations and Trends in Computer Graphics and Computer Vision 1*, 2 (December).

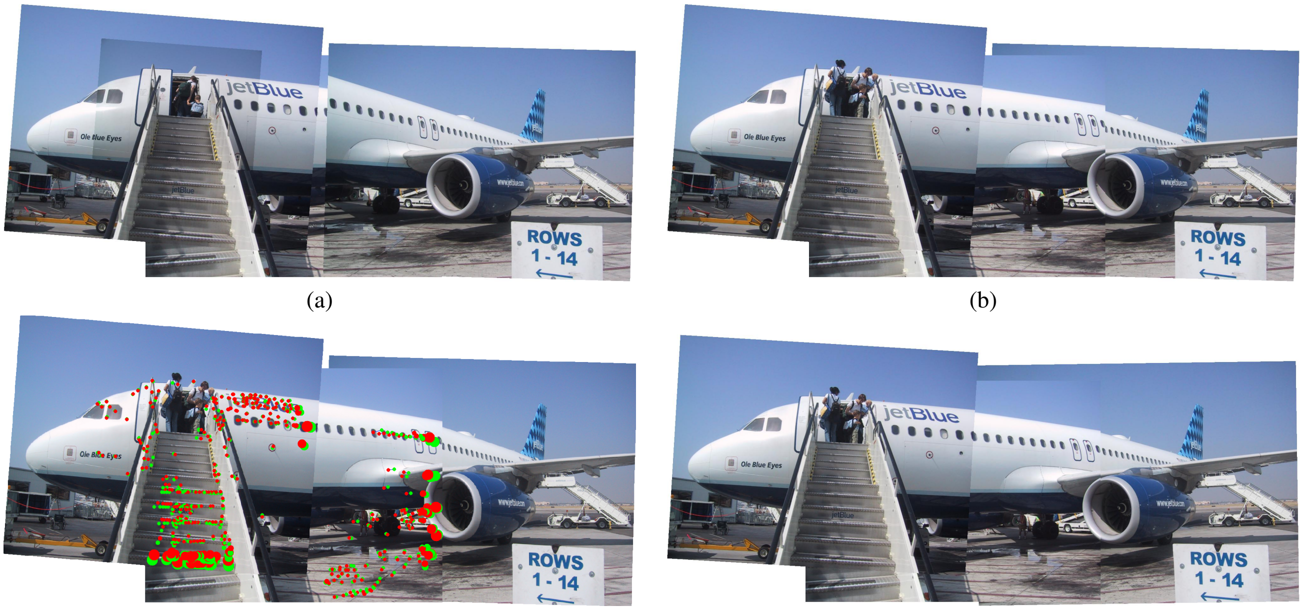
UMEYAMA, S. 1991. Least-squares estimation of transformation parameters between two point patterns. *IEEE Trans. Pattern Anal. Mach. Intell. 13*, 4, 376–380.

WANG, J., SUN, J., QUAN, L., TANG, X., AND SHUM, H. 2006. Picture collage. In *IEEE Conference on Computer Vision and Pattern Recognition*.

WOOD, D. N., FINKELSTEIN, A., HUGHES, J. F., THAYER, C. E., AND SALESIN, D. H. 1997. Multiperspective panoramas for cel animation. In *Proceedings of SIGGRAPH*.

ZELNIK-MANOR, L., PETERS, G., AND PERONA, P. 2005. Squaring the circle in panoramas. In *Tenth IEEE International Conference on Computer Vision (ICCV'05)*, vol. 2, 1292–1299.

ZOMET, A., FELDMAN, D., PELEG, S., AND WEINSHALL, D. 2003. Mosaicing new views: The crossed-slits projection. *IEEE Trans. on Pattern Analysis and Machine Intelligence* (June).



(a)

(b)

(c)

(d)

Figure 10: **Panorama construction phases:** *(a) Initial alignment of 4 pictures with all feature-matches having equal weight. Images are layered according to acquisition order resulting in large inconsistencies. (b) Layering the images by minimizing the gradient-based cost is already more consistent, yet still, small misalignments remain, e.g., the engine. (c) Same as (b) with feature matches marked in green and red. The radius of the markers is proportional to the feature weight: high weights are assigned to features near visible boundaries and low weights to those far. (d) Final result after realigning the images using the assigned importance weights. Inconsistencies are further redcued.*

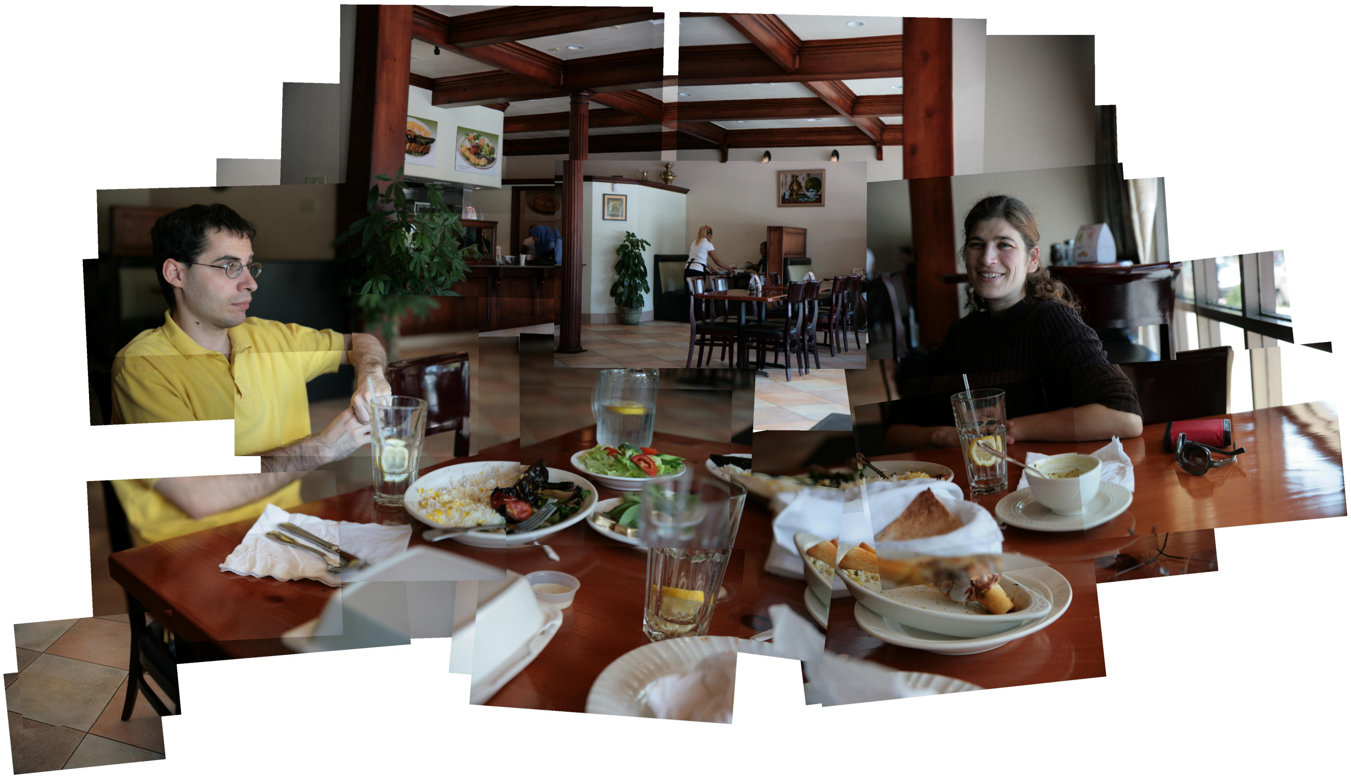


Figure 11: **Semi-Automatic** *A joiner of 57 pictures. The construction required approximately 2 minutes of user interaction. Manually assembling the same pictures into a joiner in photoshop took about 40 minutes.*