

“Mind the Gap”: Tele-Registration for Structure-Driven Image Completion

Hui Huang^{**} Kangxue Yin^{*} Minglun Gong[†] Dani Lischinski[‡] Daniel Cohen-Or[§] Uri Ascher^{*} Baoquan Chen^{*○*}
*Shenzhen VisuCA Key Lab / SIAT [†]Memorial University of Newfoundland [‡]The Hebrew University of Jerusalem
[§]Tel-Aviv University *The University of British Columbia ○Shandong University



Figure 1: Given several pieces extracted from the original image and casually placed together (left), our method applies **tele-registration** to align them (middle), and then uses **structure-driven** image completion to fill the gaps (right).

Abstract

Concocting a plausible composition from several non-overlapping image pieces, whose relative positions are not fixed in advance and without having the benefit of priors, can be a daunting task. Here we propose such a method, starting with a set of sloppily pasted image pieces with gaps between them. We first extract salient curves that approach the gaps from non-tangential directions, and use likely correspondences between pairs of such curves to guide a novel *tele-registration* method that simultaneously aligns all the pieces together. A *structure-driven* image completion technique is then proposed to fill the gaps, allowing the subsequent employment of standard in-painting tools to finish the job.

Keywords: image completion, image editing, registration, stitching, texture synthesis

Links: [DL](#) [PDF](#) [WEB](#)

1 Introduction

Image completion is a challenging task, as it attempts to conjure visual detail inside a missing portion of an image. Many completion techniques have been proposed in the computer graphics and image processing literature during the past decade, some of which maturing to the point of being featured in commercial image editing products, such as Adobe Photoshop.

In virtually all of the existing techniques, however, the shape and

*Corresponding authors: Hui Huang (hhzhiyan@gmail.com), Baoquan Chen (baoquan.chen@gmail.com)

position of the missing regions (holes) in the image is fixed, and the focus is on filling them with visually plausible content, while ensuring visual continuity across the boundaries with the known image regions. In this work we consider a significantly more difficult version of the problem, where the exact relative placement of the non-overlapping input image parts is not provided, adding a whole new dimension to consider. An example demonstrating this problem is shown in Figure 1, where the input is an assembly of sloppily pasted image pieces and the result is a complete, natural looking picture. This challenging scenario arises, for example, when creating a digital photomontage [Agarwala et al. 2004] from input images that might not be registered, attempting to copy-and-paste an object (either between two different images or within the same image) [Pérez et al. 2003], or creating a panorama from input images which might not overlap with each other [Poleg and Peleg 2012], or might even depict completely different places on earth.

Different steps of the proposed approach are illustrated in Figures 2–4 using a synthetic example. Given a set of image pieces that are casually pasted together, with gaps left between the pieces, our goal is to first align the pieces relative to each other despite their lack of overlap, and then fill the remaining gaps. We begin with detecting salient curves inside each image piece (Section 3 and Figure 2), and then attempt to find for each curve a matching curve from an adjacent piece, across the gap (Section 4.1, Figure 3). The matched salient curves are used to construct a vector field surrounding all the pieces (Section 4.2). Next, we use this *ambient vector field* to find a similarity transformation (any combination of translation, rotation, and uniform scaling) for each piece, such that the corresponding pairs of salient curves line up (Section 4.3), and construct smooth *bridging curves* that connect such pairs across gaps. Finally, we fill the gaps using structure-driven synthesis (Section 5, Figure 4), while any remaining inside/outside holes are completed using standard inpainting tools, e.g., [Barnes et al. 2009; Darabi et al. 2012].

In summary, the main contributions of our method are two-fold. The first is our novel *tele-registration* method, which simultaneously optimizes the positions of multiple disjoint non-overlapping image pieces with respect to one another. The tele-registration process is automatic, requiring the user only to provide an approximate initial placement of the pieces, which is why we refer to this process as *sloppily pasting*. Our second contribution is the structure-driven image completion technique, which makes use of the bridging curves that were constructed during the tele-registration phase

in order to propagate salient structure across the gaps before filling the remaining missing areas.

2 Related Work

Image completion and inpainting have been investigated in various forms in the computer graphics community over the past decade. The pioneering work by Bertalmio et al. [2000] performs inpainting using diffusion, filling thin missing areas by propagating information from the known boundaries along isophote directions. See also [Chan and Shen 2005] for several mathematically justified techniques. The emergence of non-parametric texture synthesis methods [Efros and Leung 1999; Wei and Levoy 2000; Efros and Freeman 2001] has led to the development of example-based image completion techniques [Drori et al. 2003; Kwatra et al. 2005; Wexler et al. 2007; Barnes et al. 2009; Darabi et al. 2012], which are capable of filling larger missing regions.

These image completion algorithms assume that the missing regions do not contain structures, or salient features that require higher-level understanding to complete correctly. Sun et al. [2005] have introduced a semi-automatic method, where the user assists the completion process by drawing curves that guide the completion of salient structures, which is then followed by standard patch-based texture synthesis. Others have investigated methods to complete the missing parts of a curve or a network of curves automatically using Euler spiral primitives [Kimia et al. 2003; Zhou et al. 2012]. In all of the above works, the basic premise is that the positions and shapes of the holes (or equivalently, the relative positions of the known parts) are known. In the present work we do not make this assumption, allowing the known regions to be disconnected and to move relative to each other in the image plane, thus adding a whole new dimension to explore.

As mentioned earlier, the method proposed in this paper is applicable in a variety of image editing scenarios. One such scenario is the creation of a panorama from a multitude of images [Zelnik-Manor and Perona 2007; Brown and Lowe 2007; Kaneva et al. 2010]. In the line of these works, the assumption is that the individual photographs overlap, thus allowing state-of-the-art registration methods to be applied. A recent exception is the work of Poleg and Peleg [2012], where a panorama is built from images that do not overlap. Their approach to the problem is to first use texture synthesis to extend the input images, and then register the extended images. However, Kopf et al. [2012] show that the visual quality of such extensions may be low near image regions with structure. In contrast, in our work, the extension of the known input image parts is structure-driven, in the sense that we extend salient structures that approach the piece boundaries, and use these extensions both for tele-registration and for subsequent gap completion.

Another application is the generation of photomontage. Pioneering work in this area assumes that the background can be aligned and the foreground objects have limited motion [Agarwala et al. 2004]. Pritch et al. [2011] alleviate these constraints by automatically snapping foreground pieces into a suitable location and blending them with a selected background. Jia and Tang [2008] not only align different pieces, but also deform salient structures in the pieces to allow smooth transition. All of these methods assume the existence of overlap regions, which is not the case in our approach. Similarly to Jia and Tang, we also extract and align salient structures; however, instead of deforming these structures, our strategy is to align the pieces such that a smooth transition can be generated within the gaps.

The problem of finding optimal piece alignment has also been studied for archaeological applications, such as restoring 2D wall paintings [Papaodysseus et al. 2002], 2D pottery [Leitao and S-

tolfi 2002], and 3D sculptures [Huang et al. 2006] from fragments. These works do not assume that fragments overlap each other, but require they share fracture lines or surfaces. Their focus is to determine that the adjacency relationship among fragments through analyzing and matching boundary curves or surfaces. Once the adjacency is found, the position of different pieces can be easily computed since the shared boundaries provide strong constraints. Our approach, on the other hand, does not assume that different pieces touch each other. It can therefore be used to align fragments separated by wider gaps.

The registration of non-overlapping and non-touching parts has been recently applied in 3D [Huang et al. 2012] for connecting two surfaces together in a coherent way. Similarly to that work, the core idea of our approach is to define a custom-designed ambient vector field surrounding each image piece to assist in the alignment process. However, unlike Huang et al. [2012], who optimize a single transformation for registering two pieces, here we simultaneously compute multiple transformations for optimal placement of multiple pieces. By extrapolating salient image features from different pieces into a common ambient vector field, we transform a difficult multi-way registration problem among different pieces into one of optimizing a set of bridging curves across a common ambient field that changes as the pieces are transformed.

3 Salient curve detection

Our multi-part tele-registration algorithm relies on salient curves detected inside the different image pieces to align them. General saliency detection is a difficult problem, which has been studied extensively [Borji and Itti 2013]. In our case it is particularly challenging because we aim to be conservative and extract only a small number of curves that can lead to reliable alignment. Fortunately, we are only interested in curves that cross the gaps between pieces and hence can limit our search to the ones intersecting the pieces’ boundaries, which are known *a priori*.

The extraction process starts by applying an edge-preserving filter to remove small-magnitude gradients from each piece (we use the WLS filter [Farbman et al. 2008]), and then computing a multiscale gradient magnitude map; see Figure 2(b). Following the boundary of each piece, we search for points with local maximal gradient magnitude along the boundary, shown as large blue dots in Figures 2(b-d). These local maxima are then considered as candidate seed points for salient curves, and we attempt to trace a salient curve inward (into the piece) from each such point. Starting from a seed point e , we search for a pixel x with the highest gradient magnitude within a prescribed distance along the expected direction. That is, pixel x needs to satisfy the following two constraints: (i) the angle between vector $x - e$ and the normal to the piece boundary at e is small enough (less than 60° by default); and (ii) the gradient magnitude at x is a local maximum and exceeds a certain threshold (0.2 by default). This process continues until no further pixels satisfying the above constraints can be found. The green curves shown in Figure 2(c) indicate salient curves that were traced in this manner.

Next, we evaluate the quality of the detected curves for our registration purposes, and discard curves that are unlikely to be useful. For each curve s we evaluate $u(s) = \text{len}(s) + \text{grad}(s) + \text{orth}(s) - \text{curv}(s)$, where len is the curve length, grad is the average gradient magnitude along s , orth is the curve orthogonality to the boundary, measured as the average of the cosines between the curve segments and the normal to the boundary, and curv is the average curvature (the smaller the better). Each of these four quantities is normalized to be in the range $[0, 1]$, so $-1 \leq u(s) \leq 3$. Only curves for which $u(s)$ is higher than a threshold (1 by default) are kept for the subsequent correspondence search; see Figure 2(d).

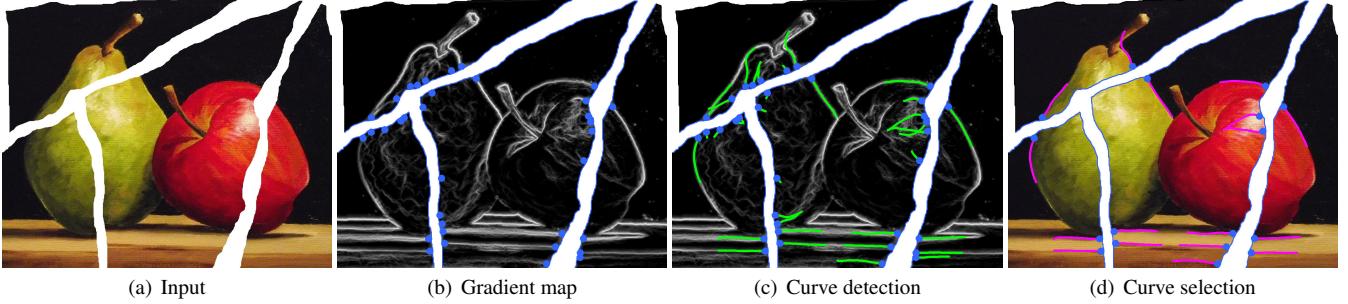


Figure 2: Salient curve detection and selection. Given the input pieces (a) and their multiscale gradient magnitude maps (b), we first identify (blue) seed points, from which we trace (green) salient curves (c). The detected curves are evaluated, and only those which pass the thresholding (pink) are kept (d).

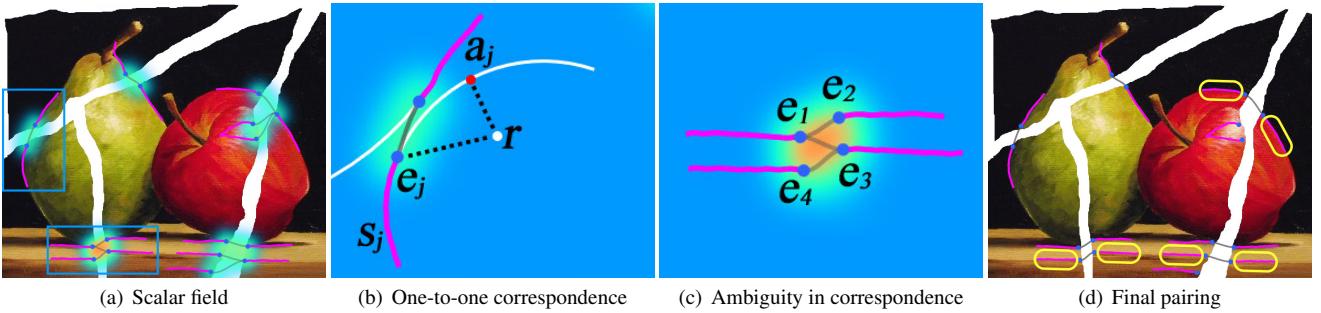


Figure 3: Curve correspondence. In some cases (b), the correspondence can be determined solely using the scalar field (a), whereas in others (c), additional color similarity between regions along the curve (inside yellow ovals in (d)) is needed to resolve the ambiguity.

4 Tele-registration

We seek a similarity transformation for the image pieces such that the corresponding salient curves line up as smoothly as possible. Our first task is to select pairs of matching salient curves from the detected ones for registration (Section 4.1). These matching curve pairs are then used to generate an ambient vector field (Section 4.2), which guides the alignment of different pieces (Section 4.3).

4.1 Curve matching

Let $P = \{p_i\}_{i \in I}$ denote the set of image pieces, $S = \{s_j\}_{j \in J}$ and $E = \{e_j\}_{j \in J}$ denote the detected salient curves and the their starting seed points, respectively. Further, let $g : J \rightarrow I$ indicate the piece that contains each curve. Our goal is to find matching pairs of curves $(s_j, s_{j'})$ such that $g(j) \neq g(j')$. Informally, the idea is to predict the extension of each salient curve outward from the piece boundary, and look for smooth bridging curves between pairs of salient curves that maximize the proximity to the predicted curve extensions.

We generate a sequence of n sample points along each salient curve (spaced 20 pixels apart in our implementation) and fit these sample points with a Bézier curve of degree $n - 1$. This curve is then extrapolated across the piece border using its osculating circle at its seed point (white arcs in Figure 3(b)). We define the likelihood $l(r)$ of a point r to lie on an *actual* extension of a salient curve as a sum of products of Gaussian falloff terms:

$$l(r) = \frac{1}{|J|} \sum_{j \in J} \exp\left(-\frac{|r - a_j|^2}{\sigma_1^2}\right) \exp\left(-\frac{|r - e_j|^2}{\sigma_1^2}\right), \quad (1)$$

where a_j is the projection of r onto the first quarter arc of the osculating circle that extrapolates the curve from the seed point e_j (see Figure 3(b)), and $|J|$ is the number of curves in the set S . The parameter σ_1 in the Gaussian falloff function is set to 0.02 by default.

This definition yields a scalar likelihood field with the highest likelihood areas occurring where the seed points of two or more curves are close to each other, as shown in Figure 3(a). For each seed point e_j we look for a matching seed point $e_{m(j)}$, where $m(j) \in J$ and $g(j) \neq g(m(j))$, such that the smooth Hermite curve connecting the two minimizes the sum of field values $1 - l(r)$ along its path.

We further build a matching candidate set for each seed point e_j , which contains the seed point it maps to, $e_{m(j)}$, and zero or more seed points that map to it, $e_{j'}, m(j') = j$. If the set of e_j contains only one candidate $e_{m(j)}$, while the set of $e_{m(j)}$ contains only e_j , we determine that $(e_j, e_{m(j)})$ is a matched pair. Otherwise, we have an ambiguity as shown in Figure 3(c), where both e_2 and e_3 are in e_1 's candidate set. To resolve such ambiguities, we employ additional color cues by extracting a small 20-pixel width patch covering the curve s_j emanating from e_j and also patches covering the salient curves emanating from e_j 's matching candidates; see areas highlighted by yellow rounded rectangles in Figure 3(d). The similarity between two patches is computed by first warping one to another [Schaefer et al. 2006] to align the underlying salient curves, and then compute the RGB least square distance. The candidate $e_{j'}$ that leads to the highest patch similarity is considered the best. The matching pair $(e_j, e_{j'})$ is then confirmed if e_j is also the best candidate of $e_{j'}$. Figure 3(d) shows the final matching pairs obtained after utilizing both geometry and color cues.

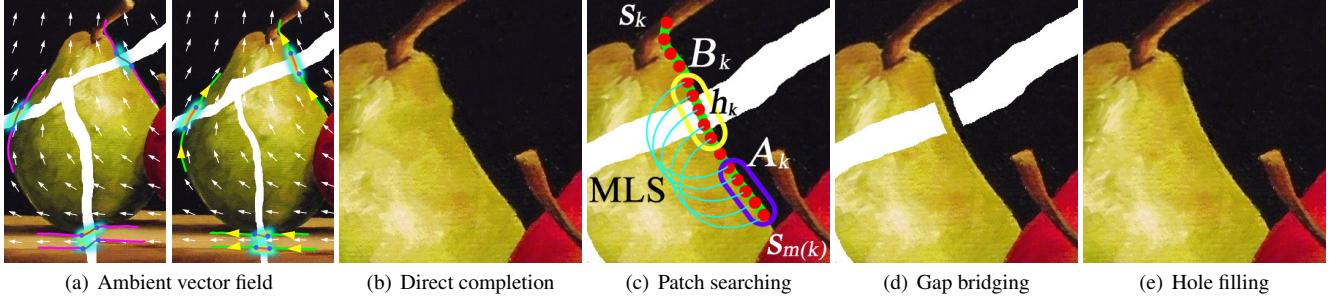


Figure 4: (a) Ambient vector field generated before (left) and after (right) field-guided registration. To fill the gaps and avoid distortion caused by unconstrained inpainting (b), our approach first finds a matching patch along the curve (c) and transfers colors from its neighborhood (d). The remaining holes are filled using the content-aware filling tool embedded in Photoshop (e).

4.2 Ambient vector field generation

Next, a smooth bridging Hermite curve h_k is constructed to connect each pair of matched salient curves $(s_k, s_{m(k)})_{k \in K}$ (where $K \subseteq J$ contains only one curve from each matched pair). Our goal is to align the pieces such that the bridging curve set $\{h_k\}_{k \in K}$ minimizes an energy functional with respect to an ambient vector field.

We construct the ambient vector field using a method similar to that of Xu et al. [2009]. Specifically, we incrementally assign an orientation for each pair of salient curves $(s_k, s_{m(k)})$, so as to obtain the smoothest interpolating vector field. This is done using harmonic interpolation, i.e., by solving Laplace equations with Dirichlet boundary conditions along the curves.

The resulting vector field consists only of directions $\mathbf{v}(r)$, which is not enough to precisely register two curves, especially when they are nearly parallel. Therefore, for each location r in the field we augment the direction with a magnitude $d(r)$ that depends on the distances of r from the extrapolating osculating circles:

$$d(r) = \frac{1}{|K|} \sum_{k \in K} \exp\left(-\frac{|r - a_k|^2}{\sigma_2^2}\right) \exp\left(-\frac{|r - a_{m(k)}|^2}{\sigma_2^2}\right),$$

where $|r - a_k|$ and $|r - a_{m(k)}|$ are the shortest distances from r to the first quarter of the osculating circles extending from the seed points a_k and $a_{m(k)}$, respectively. The parameter σ_2 controls the two Gaussian falloff functions, with a default value of 0.01.

The directions $\mathbf{v}(r)$ and the scalar field $d(r)$ together form the ambient vector field, as shown in Figure 4(a). It naturally extends the salient curve pairs into the ambient space.

4.3 Field-guided registration

With the ambient vector field defined, our goal is now to reposition the image pieces, such that each bridging curve is aligned with the field directions, and also naturally extends the salient curves from its seed point. Specifically, we solve for a set of similarity transformations $T = \{T_i\}_{i \in I}$ defined on the set of image pieces P , that is $T(P) = \{T_i(p_i)\}_{i \in I}$, and given by

$$\begin{aligned} T &= \arg \min_T E(T(P)), \\ E(T(P)) &= \sum_{k \in K} \frac{1}{|h_k|} \int_{h_k(T)} f(r) dr, \\ f(r) &= 1 - \mathbf{t}(r)^T \mathbf{v}(r) + \lambda(1 - d(r)). \end{aligned} \quad (2)$$

Here $|h_k|$ denotes the length of the Hermite bridging curve h_k , and h_k is a function of T since its two endpoint locations are given by $T_{g(k)}(e_k)$ and $T_{g(m(k))}(e_{m(k)})$, respectively. $\mathbf{t}(r)$ denotes the tangential direction of h_k at location r . Note that among the two tangential directions, the one along the orientation of the salient curve pair $(s_k, s_{m(k)})$ is used. The parameter λ balances between following the directions of the vector field $\mathbf{v}(r)$ and minimizing the distances from the salient curve extensions. We found that setting $\lambda = 1$ works well in practice.

To compute the similarity transformations T that include translation, rotation and uniform scaling, we utilize the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm [Nocedal and Wright 2006] and solve (2) using the *fminunc* function from the MATLAB Optimization Toolbox. In each iteration, the pieces are repositioned and the field is therefore updated correspondingly. Note that BFGS, as well as many other non-linear optimization methods, require a good initial guess to guarantee convergence to a local minimum. Our initial sloppily pasted input is assumed to provide a sufficiently accurate initial configuration, leading to an acceptable optimum. The question of automatically deriving such an initial configuration from a general initial state is outside the scope of this article, and is a possible avenue for future work.

5 Structure-driven completion

The tele-registered image pieces usually contain small gaps among them, but are linked with bridging curves h_k . These curves provide structural information inside the missing areas, allowing us to perform image completion in a similar manner to the structure propagation approach of Sun et al. [2005].

The completion is first performed along the bridging curves, which link between salient curves and thus connect regions containing high image gradient magnitude. Unconstrained inpainting in such areas is likely to blur or distort the high contrast edges that run across the gaps (see Figure 4(b)), and a special treatment is therefore required. Intuitively, to fill the unknown region along a given bridging curve h_k , we infer color information from a region along the corresponding salient curves s_k and $s_{m(k)}$. Since the curvature may vary along these curves, geometric distortion may be required for matching the two regions.

Thus, we proceed as follows. Each connected chain $(s_k, h_k, s_{m(k)})$ is first discretized into a set of ordered points. A *bridging subset* B_k is then selected so as to contain both points from h_k (with unknown color) and points from s_k and $s_{m(k)}$ (with known colors); see yellow oval in Figure 4(c). Next, our task is to find another subset A_k of points from s_k and $s_{m(k)}$ such that, after warp-aligning A_k to

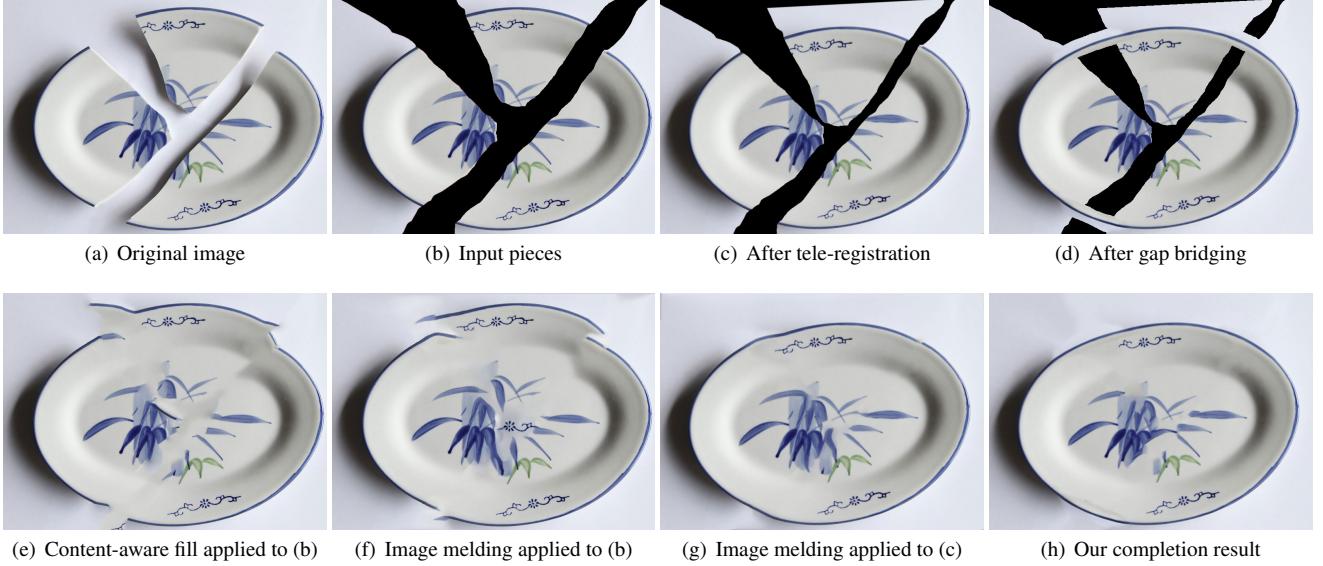


Figure 5: Mending a broken plate. Attempting to complete the image in (b) using content-aware fill (e) or image melding (f) before tele-registration fails to produce a satisfactory result. Image melding after tele-registration produces a better result, but the plate outline is still not smooth (g). Our method applies content-aware fill after structure-driven gap bridging to obtain a more plausible result (h).

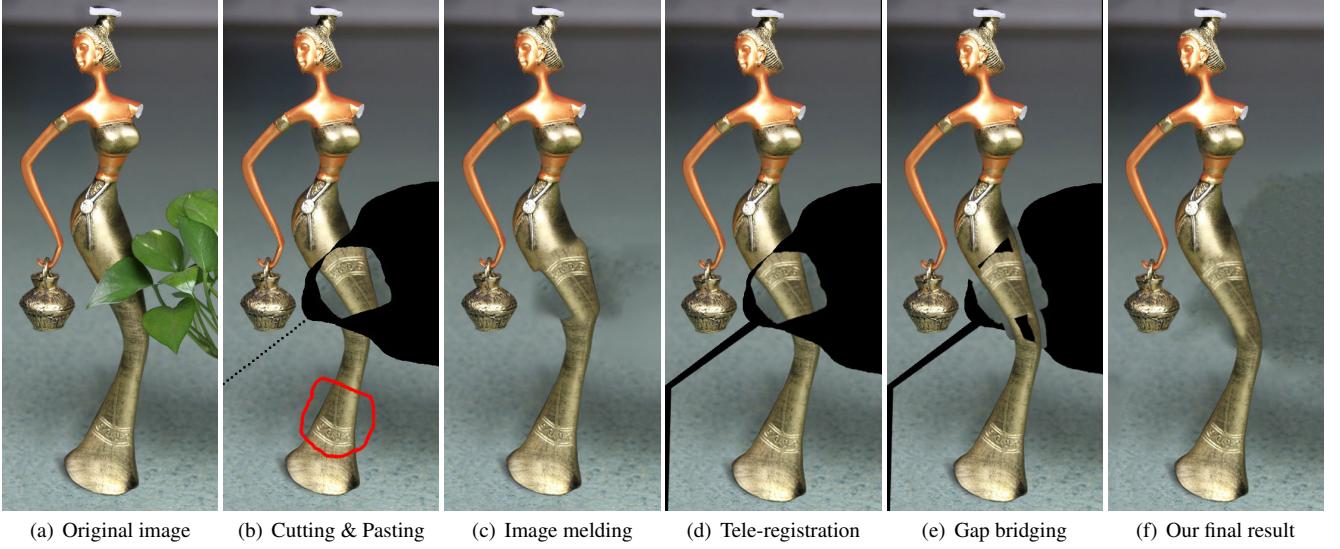


Figure 6: Reconstructing an occluded part of a statue. The lassoed part in (b) is sloppily pasted inside the hole. Image melding fails to produce a plausible result (c). Tele-registration (d), followed by gap bridging (e) yields a more plausible reconstruction (f).

B_k , the similarity between the neighborhoods of the corresponding known points in B_k and A_k is maximized. Once such a subset A_k (purple oval in Figure 4(c)) is found, we transfer color from the neighborhood of A_k to B_k . Here the patch warping is performed using moving least squares (MLS) [Schaefer et al. 2006] with rigid transformation and control points from the salient curves. Following the transfer with warping, a seamless cloning is obtained with Poisson editing [Pérez et al. 2003]; see Figure 4(d).

Once the gaps are completed properly along the bridge curves, the remaining holes are usually contained inside smooth regions. Many existing inpainting tools work well under such conditions. We have used the content-aware filling tool [Barnes et al. 2009] embedded

in Photoshop in all our experiments; see in particular Figure 4(e).

6 Results

In this section we present a number of results demonstrating several applications of our technique, and compare our results to alternative approaches. The default parameter sets are applied throughout all the presented experiments except for the relief restoration example in Figure 8, where we used smaller Gaussian parameters $\sigma_1 = 0.01$ and $\sigma_2 = 0.005$ to relieve curve matching ambiguities and to limit the shifts of image pieces during the tele-registration. Note that once the user provides a rough initial alignment for image pieces,



Figure 7: Removing a butterfly from the scene. Labeling the area covered by the butterfly as unknown (black) and casually placing the lassoed parts inside (b) provide us the input. Our algorithm then automatically aligns the parts (c) and completes the gaps (d). In comparison, directly inpainting using content-aware filling tool in the unknown area in (b) without the lassoed parts yields artifacts (e).

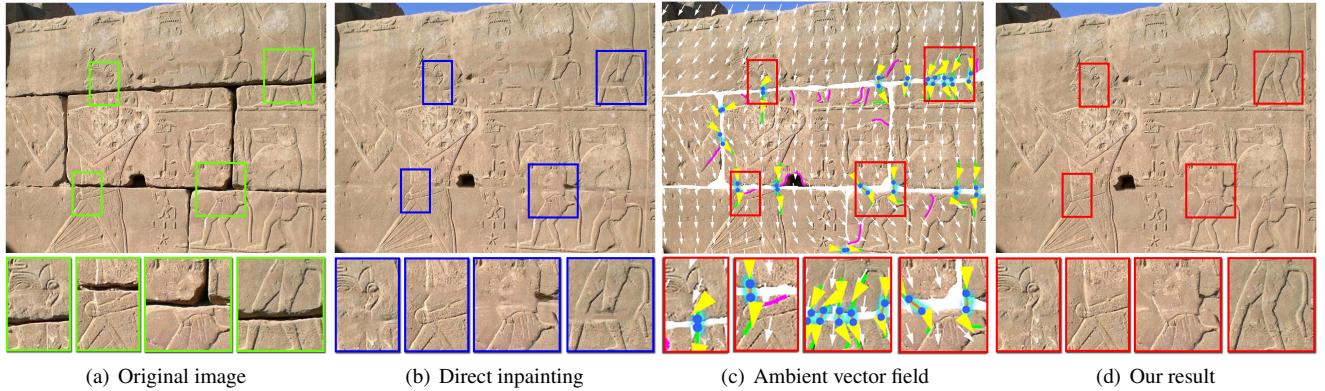


Figure 8: Aligning the large stones guided by the ambient vector field (c) and then completing their gaps for the Temple of Amon. A better result (d) is obtained than by just using direct content-aware inpainting (b). Note that pink curves in (c) indicate those salient curves detected from the borders but not paired, and so they are discarded when computing the ambient vector field.

the remaining operations are automatic. Nevertheless, some user corrections on salient curve detection or matching might be needed in complex or extreme cases.

Image completion. We start by demonstrating how better image completion results can be obtained through aligning pieces into optimal locations. In the following examples the input image pieces originate from the same image, and a good registration can be achieved without scaling. Hence, here we compute an optimal rigid transformation for each input piece.

Figure 5(a) shows an image of a plate shattered into three pieces. Labeling the gaps between the pieces as unknown (black color) yields the input for our method; see Figure 5(b). Attempting to fill the unknown areas using state-of-the-art image completion methods [Barnes et al. 2009; Darabi et al. 2012] fails to produce satisfactory results (Figure 5(e-f)), since the pieces are not positioned well. Applying our tele-registration to the same input repositions the pieces as shown in Figure 5(c). When applied to this input, image melding generates a much better result, however, the rim of the plate still wiggles a bit. Our method first extends the salient structures across the remaining gaps (Figure 5(d)), and then fills the remaining holes using Photoshop’s content-aware fill, yielding the visually plausible result in (Figure 5(h)).

Figure 6 shows a statue parts of which are obscured by the leaves of a plant. Our method can be used to remove the occlusion by erasing the occluding leaves from the image, selecting a suitable replacement for the missing part of the statue from another part (red lasso in Figure 6(b)), and sloppily pasting it inside the hole. Applying image melding [Darabi et al. 2012] or the content-aware filling tool

in Photoshop (not shown) fails to produce a satisfactory result because of the inaccurate position of the pasted part. Tele-registration followed by gap bridging and completion yields the much more plausible result depicted in Figure 6(f).

A somewhat similar scenario is shown in Figure 7, where the objective is to remove the foreground butterfly. A suitable replacement for the uncovered background may be found from elsewhere in the same image. Again, sloppily pasting these replacement pieces, followed by our tele-registration and structure-driven completion, yields a highly convincing result.

Painting/relief restoration. As a variant of the image completion problem, next we show how corrupted paintings and reliefs can be restored from their fragments using our technique. Figure 8(a) shows a photo of a wall with a bas-relief from an ancient temple in Egypt. Over the years, the huge stone blocks comprising the wall may have shifted slightly resulting in misalignments of the depicted shapes. Direct inpainting to fill the gaps between the blocks results in blurring, as seen in Figure 8(b). In contrast, a better result is produced by separating the blocks from each other, letting our method re-align the blocks, and then filling the gaps (Figure 8(d)). Here, the curve matching is challenging and the registration needs to be very accurate as even a small erroneous drift may lead to noticeable artifacts. Our ambient vector field shown in Figure 8(c) plays an essential role and its robustness has been demonstrated clearly in such a difficult scenario.

In Figure 9, the reconstruction starts from a set of torn pieces of an oil painting. The user roughly positions the pieces (Figure 9(b)), and our algorithm takes over from there. The ambient vector field

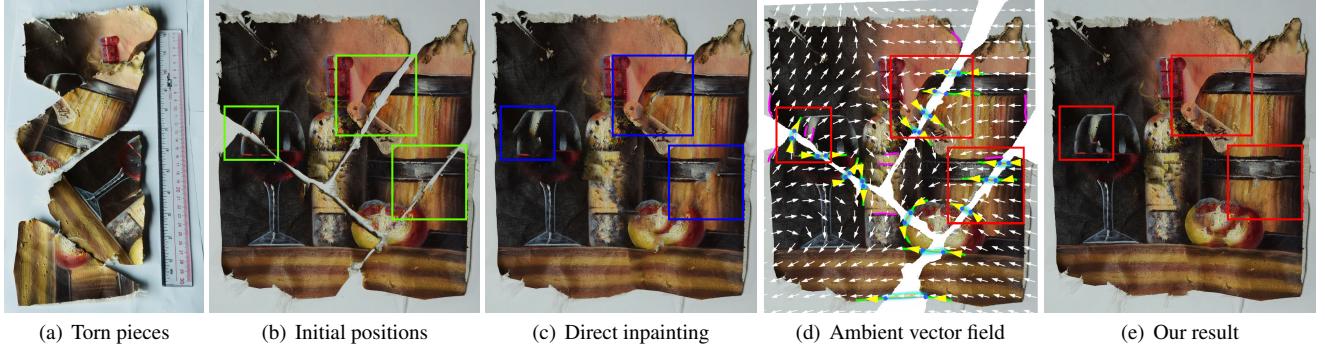


Figure 9: Putting together an oil painting from its torn pieces. Areas worthy of closer inspection are highlighted using boxes.

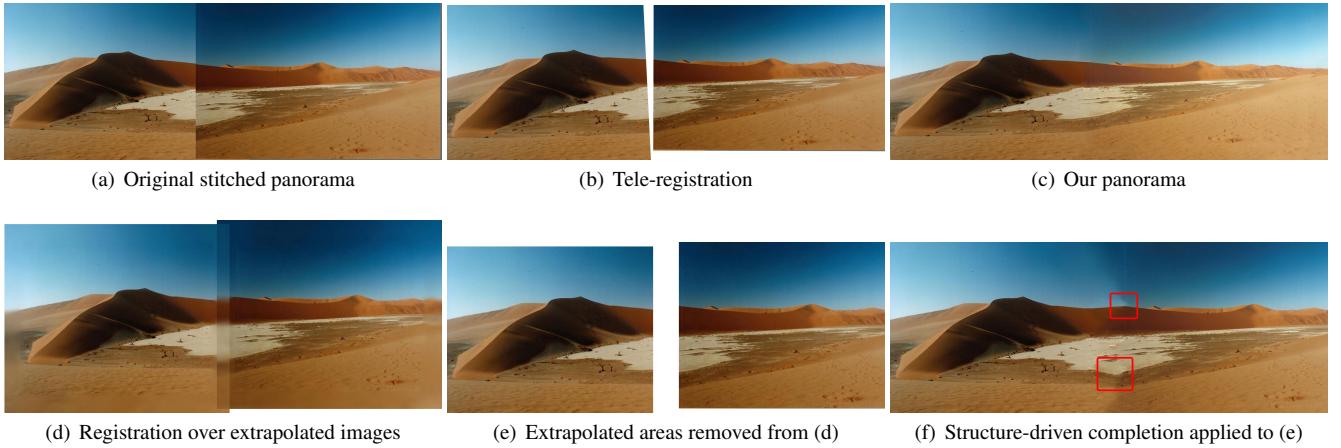


Figure 10: Fixing a poorly stitched panorama (a) found through Google image search for Namib desert. Separating the two photos and feeding them to our method results in better alignment (b) and a more seamless panorama (c). Note that the right piece in (b) is scaled to allow for smoother alignment of salient curves. In comparison, Poleg and Peleg [2012] first extrapolate the two photos to create overlap, which yields the alignment in (d). Removing the extrapolated areas (e) reveals that the registration result does not fully respect the salient curves in the scene. Applying structure-driven completion over (e) does not fully overcome the problem.

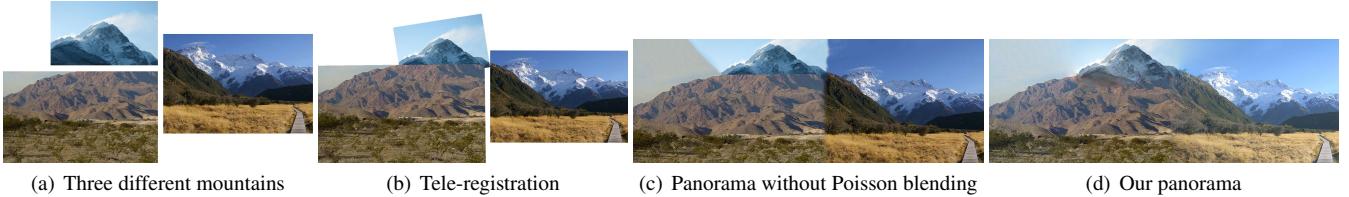


Figure 11: Generating an unusual yet natural looking mountain panorama from images of three different mountains. Note that in the tele-registration result (b), the two left pieces overlap. Compared to the direct inpainting result (c), the additional Poisson blending operation provides much smoother transition among image pieces of large color difference (d).

(Figure 9(d)) is again very effective and our result is again visibly superior to what content-aware inpainting can do.

Panorama creation. Figures 10 and 11 demonstrate the creation of panoramas from non-overlapping images. Unlike in our previous examples, here the input pieces originate from different images and may differ in scale. Hence, the optimal solution is searched from the space of similarity transformations, i.e., any combination of translation, rotation, and uniform scaling. Figure 10 shows how our approach is applied to fix a poorly stitched panorama. Although the two input photos capture the same desert scene, tradi-

tional panorama stitching algorithms [Brown and Lowe 2007] can hardly apply due to lack of overlap, as well as lack of feature points. Our approach is able to naturally align the main feature curves (Figure 10(b)) and generate the more natural-looking panorama in Figure 10(c). For comparison, we also show the result generated using the approach of Poleg and Peleg [2012], which performs registration of extrapolated images, as shown in Figures 10(d-e). It may be seen that such registration is less successful at aligning the salient curves, resulting in a more noticeable transition between the two photos after completion (Figure 10(f)).



Figure 13: Swapping the heads between two animal statues (a). Using the same input (b), the results obtained under both similarity transformation (c) and rigid transformation (d) are shown. In both cases, the contours of the resulting foreground objects are smooth.

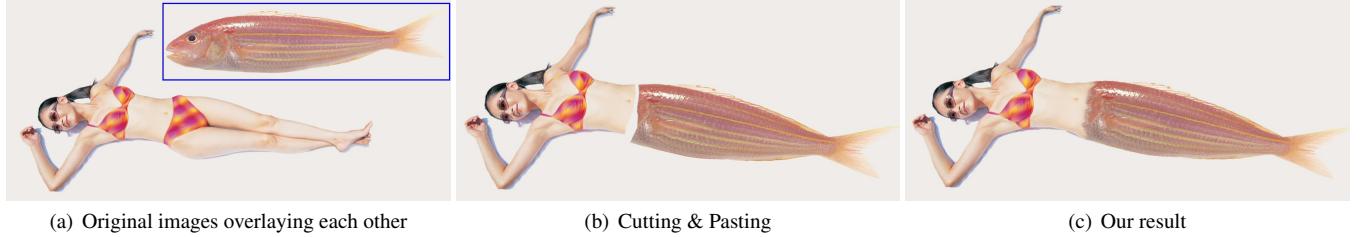


Figure 14: A composition of pieces originating from different images.

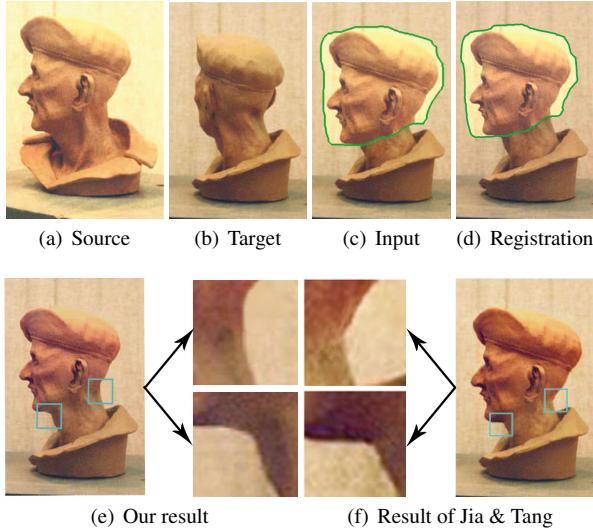


Figure 12: Comparison with Jia & Tang [2008] on a photomontage application. Given source (a) and target (b) images, casually overlaying a region from the source (the area enclosed by green curve in (c)) over the target image serves as our input. Our registration approach raises the source region higher and scales it smaller to achieve local optimal alignment among paired salient curves (d). The final result blended with Poisson editing (e) offers a realistic looking image. Since our approach optimizes the smoothness of salient curves, the resulting silhouette of the foreground model is smoother than the one generated by Jia and Tang (f).

Figure 11 goes one step further and demonstrates how several photos depicting different landscapes may be combined together to generate a panorama of a landscape that does not exist in reality. Note that in this example, the optimal alignment found overlaps the

two images on the left. In such situations, we automatically and dynamically erode the paired salient curves so that they do not intersect with each other. Furthermore, since there is no gap between the overlapping pieces, the structure-driven completion step is no longer needed. However, when the colors or texture from different images vary largely, direct content-aware filling would yield unnatural color or texture transitions among pieces (Figure 11(c)). Thus here we first choose one input piece (the top one in this example) as a main theme and apply Poisson blending [Pérez et al. 2003] with the chosen boundary constraints and mixing gradients to slightly adjusting other original pieces. The ensuing content-aware filling can then lead to a rather smooth color and texture transition between image pieces (Figure 11(d)).

Digital photomontage. As demonstrated by Jia and Tang [2008], synthesizing images by sloppily pasting pieces from different images with different objects is another interesting and practical application. Using the same image set, Figure 12 compares the result generated by our approach with the one reported in their paper. It shows that our approach yields an arguably more natural composition with smoother object contours. A similar example is shown in Figure 13, where the heads of a dog and a cock models are swapped seamlessly. The figure also shows that users have the freedom of either enabling or disabling the scaling of the pieces. Figure 14 further shows how a mermaid is constructed by composing a fish tail with a sunbathing girl. Our method ensures that the body contour of the girl is continued in a smooth fashion by that of the fish.

Stress test. Attempting to establish how robust our method is to the initial arrangement of the pieces, we performed a stress test whose results are reported in Figure 15. Two pieces cut from the same image are used so that the ground truth registration is known. The result shows that our method is generally more robust to translation of pieces away from their correct position than to rotation away from their correct orientation: while we were able to move the smaller piece 100 pixels away and still converge to a reasonable result, it was difficult to obtain a good result with rotations exceeding 10 degrees; see Figure 15(d-f).

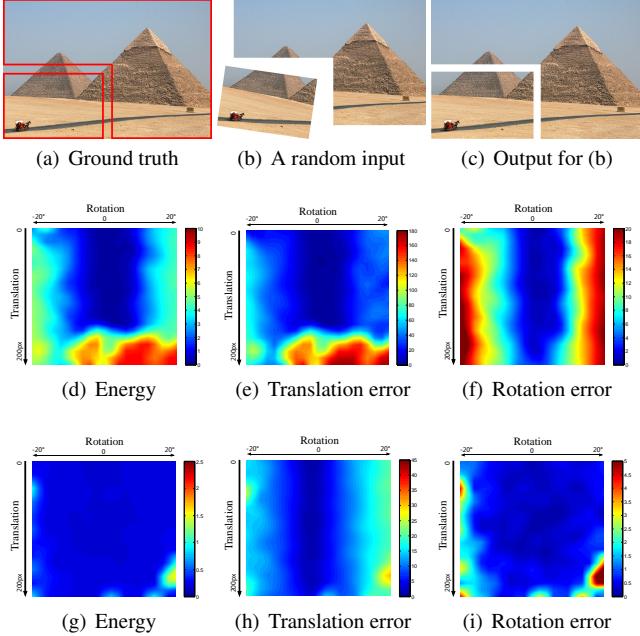


Figure 15: Stress test performed using two pieces cut from the same image (a). For a random input (b) where the small piece is translated 60 pixels and rotated 8 degrees, our algorithm properly moves the piece back to the correct position (c). Energy and translation/rotation error with (g-i) and without (d-f) random initial search are plotted for different amounts of initial translation (vertical axis) and rotation (horizontal axis). Note that each plot uses its own color scale (the maximum value of the bottom one is a quarter of the corresponding one above).

The robustness to initial arrangement can be improved by randomly perturbing the piece to be registered. For example, if we randomly move the smaller piece within [-40px, 40px] and [-8°, 8°] to generate 50 initial arrangements, run our tele-registration, and pick the registration result with the smallest energy, both the final energy and translation/rotation error can be greatly reduced; see Figure 15(g-i). Note that although random perturbation is found useful for increasing the robustness of the registration, in practice, we found the users’ casually provided initial arrangements are accurate enough. Hence, none of the example shown earlier uses this feature.

7 Conclusions

Given a set of several sloppily pasted image pieces with or without gaps among them, our system can re-align the pieces together and produce a plausible completed image without the benefit of overlaps or priors. We have proposed a novel tele-registration method using extracted salient curves, followed by a structure-driven image completion technique that is topped by standard inpainting. We have demonstrated our system in action for several different image editing scenarios, and showed that the best existing tools alone cannot achieve comparable results. In the context described for such problem instances, our method is clearly superior to any alternatives we are aware of.

Limitations. As designed, our algorithm depends on a successful salient curve extraction and matching. Given that, the field-guided tele-registration is robust enough to handle many challenging images as demonstrated in the previous section. However, if the input

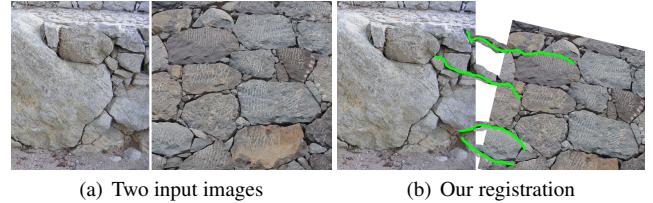


Figure 16: Limitation: with distracting texture patterns or complicated edge detail, correct detection and matching of salient curves from different pieces can be very challenging, yielding undesirable registration results.



Figure 17: Limitation: while the salient curves from different pieces are smoothly connected in the registered result, the buildings in the area highlighted by the red box are tilted, resulting in an unrealistic panorama.

image pieces contain distracting texture patterns (Figure 16) or cluttered scenes (Figure 17), our salient curve detection and matching steps might not work well and user intervention may be needed. Also, we assume the input image pieces are clean enough to detect and pair salient curves. In the presence of low image quality issues, such as heavy noise and edge blurriness, some image enhancement operations need to be performed first.

Another issue is that our approach aims at aligning paired salient curves among different pieces without accounting for the content inside each individual piece. Hence, for applications such as panorama creation, the registration result may appear unnatural even though the salient curves are smoothly connected, e.g., the buildings are tilted in Figure 17. In such situations, additional user intervention is needed to provide the direction of the horizon and restrict the search space for optimal transformations accordingly.

Figure 11 shows an interesting example, where the texture and colors from the three input image pieces are not really compatible. Even with the Poisson blending, one can still observe some color shifts and blurry artifacts in the new panorama generated (Figure 11(d)). Meanwhile, the gaps among the image pieces must be narrow enough, or the missing structure across the gaps must be sufficiently smooth and unsurprising (e.g., linear), to allow any reasonable reconstruction without additional a priori information. If such conditions do not hold, our method can be made to fail.

Future work. Parts of our algorithm rely on a set of heuristics in order to overcome uncertainty in the data in the context of our inverse problem. We intend to model such uncertainties using a statistical (probabilistic) framework which may have to be learned.

Acknowledgements

The authors would like to thank all the reviewers for their valuable comments. This work was supported in part by grants from NSFC (61379090, 61103166, 61232011), Guangdong Science and Technology Program (2011B050200007), Shenzhen Innovation Program (KQCX20120807104901791, JCYJ20130401170306810, CXB201104220029A, ZD201111080115A), NSERC (293127 and 84306), the Israel Science Foundation and the US-Israel Binational Science Foundation.

References

- AGARWALA, A., DONTCHEVA, M., AGRAWALA, M., DRUCKER, S., COLBURN, A., CURLESS, B., SALESIN, D., AND COHEN, M. 2004. Interactive digital photomontage. *ACM Trans. on Graph (Proc. of SIGGRAPH)* 23, 3, 294–302.
- BARNES, C., SHECHTMAN, E., FINKELSTEIN, A., AND GOLDMAN, D. B. 2009. Patchmatch: a randomized correspondence algorithm for structural image editing. *ACM Trans. on Graph (Proc. of SIGGRAPH)* 28, 3, 24:1–24:11.
- BERTALMIO, M., SAPIRO, G., CASELLES, V., AND BALLESTER, C. 2000. Image inpainting. *Proc. of SIGGRAPH*, 417–424.
- BORJI, A., AND ITTI, L. 2013. State-of-the-art in visual attention modeling. *IEEE Trans. Pat. Ana. & Mach. Int.* 35, 1, 185–207.
- BROWN, M., AND LOWE, D. G. 2007. Automatic panoramic image stitching using invariant features. *Int. J. Comp. Vis.* 74, 1, 59–73.
- CHAN, T., AND SHEN, J. 2005. *Image Processing and Analysis: Variational, PDE, Wavelet and Stochastic Methods*. SIAM.
- DARABI, S., SHECHTMAN, E., BARNES, C., GOLDMAN, D. B., AND SEN, P. 2012. Image melding: Combining inconsistent images using patch-based synthesis. *ACM Trans. on Graph (Proc. of SIGGRAPH)* 31, 4, 82:1–82:10.
- DRORI, I., COHEN-OR, D., AND YESHURUN, H. 2003. Fragment-based image completion. *ACM Trans. on Graph (Proc. of SIGGRAPH)* 22, 3, 303–312.
- EFROS, A. A., AND FREEMAN, W. T. 2001. Image quilting for texture synthesis and transfer. *Proc. of SIGGRAPH*, 341–346.
- EFROS, A., AND LEUNG, T. 1999. Texture synthesis by non-parametric sampling. In *Proc. IEEE Conf. on Comp. Vis. and Pat. Rec.*, 1033–1038.
- FARBMAN, Z., FATTAL, R., LISCHINSKI, D., AND SZELISKI, R. 2008. Edge-preserving decompositions for multi-scale tone and detail manipulation. *ACM Trans. on Graph (Proc. of SIGGRAPH)* 27, 3, 67:1–67:10.
- HUANG, Q.-X., FLÖRY, S., GELFAND, N., HOFER, M., AND POTTMANN, H. 2006. Reassembling fractured objects by geometric matching. *ACM Trans. on Graph (Proc. of SIGGRAPH)* 25, 3, 569–578.
- HUANG, H., GONG, M., COHEN-OR, D., OUYANG, Y., TAN, F., AND ZHANG, H. 2012. Field-guided registration for feature-conforming shape composition. *ACM Trans. on Graph (Proc. of SIGGRAPH Asia)* 31, 6, 171:1–171:11.
- JIA, J., AND TANG, C.-K. 2008. Image stitching using structure deformation. *IEEE Trans. Pat. Ana. & Mach. Int.* 30, 4, 617–631.
- KANEVA, B., SIVIC, J., TORRALBA, A., AVIDAN, S., AND FREEMAN, W. T. 2010. Infinite images: Creating and exploring a large photorealistic virtual space. In *Proc. the IEEE*, vol. 98, 1391–1407.
- KIMIA, B. B., FRANKEL, I., AND POPESCU, A.-M. 2003. Euler spiral for shape completion. *Int. J. Comp. Vis.* 54, 1-3, 157–180.
- KOPF, J., KIENZLE, W., DRUCKER, S., AND KANG, S. B. 2012. Quality prediction for image completion. *ACM Trans. on Graph (Proc. of SIGGRAPH Asia)* 31, 6, 131:1–131:8.
- KWATRA, V., ESSA, I., BOBICK, A., AND KWATRA, N. 2005. Texture optimization for example-based synthesis. *ACM Trans. on Graph (Proc. of SIGGRAPH)* 24, 3, 795–802.
- LEITAO, H. C. G., AND STOLFI, J. 2002. A multiscale method for the reassembly of two-dimensional fragmented objects. *IEEE Trans. Pat. Ana. & Mach. Int.* 24, 9, 1239–1251.
- NOCEDAL, J., AND WRIGHT, S. J. 2006. *Numerical Optimization* (2nd ed.). New York: Springer-Verlag.
- PAPAODYSSSEUS, C., PANAGOPOULOS, T., EXARHOS, M., TRIANTAFILLOU, C., FRAGOULIS, D., AND DOUMAS, C. 2002. Contour-shape based reconstruction of fragmented, 1600 b.c. wall paintings. *IEEE Trans. on Signal Processing* 50, 6, 1277–1288.
- PÉREZ, P., GANGNET, M., AND BLAKE, A. 2003. Poisson image editing. *ACM Trans. on Graph (Proc. of SIGGRAPH)* 22, 3, 313–318.
- POLEG, Y., AND PELEG, S. 2012. Alignment and mosaicing of non-overlapping images. In *Proc. IEEE Int. Conf. on Computational Photography*, 1–8.
- PRITCH, Y., POLEG, Y., AND PELEG, S. 2011. Snap image composition. In *Proc. the 5th int. conf. on Computer vision/computer graphics collaboration techniques*, 181–191.
- SCHAFFER, S., MCPHAIL, T., AND WARREN, J. 2006. Image deformation using moving least squares. *ACM Trans. on Graph (Proc. of SIGGRAPH)* 25, 3, 533–540.
- SUN, J., YUAN, L., JIA, J., AND SHUM, H.-Y. 2005. Image completion with structure propagation. *ACM Trans. on Graph (Proc. of SIGGRAPH)* 24, 3, 861–868.
- WEI, L.-Y., AND LEVOY, M. 2000. Fast texture synthesis using tree-structured vector quantization. *Proc. of SIGGRAPH*, 479–488.
- WEXLER, Y., SHECHTMAN, E., AND IRANI, M. 2007. Spacetime video completion. *IEEE Trans. Pat. Ana. & Mach. Int.* 29, 3, 463–476.
- XU, K., COHEN-OR, D., JU, T., LIU, L., ZHANG, H., ZHOU, S., AND XIONG, Y. 2009. Feature-aligned shape texturing. *ACM Trans. on Graph (Proc. of SIGGRAPH Asia)* 28, 5, 108:1–108:7.
- ZELNIK-MANOR, L., AND PERONA, P. 2007. Automating joiners. In *Proc. Int. symp. on Non-photorealistic animation and rendering*, 121–131.
- ZHOU, H., ZHENG, J., AND YANG, X. 2012. Euler arc splines for curve completion. *Computers & Graphics* 36, 6, 642–650.