Seam-Driven Image Stitching

Junhong Gao[1], Yu Li[1], Tat-Jun Chin[2], Michael S. Brown[1]

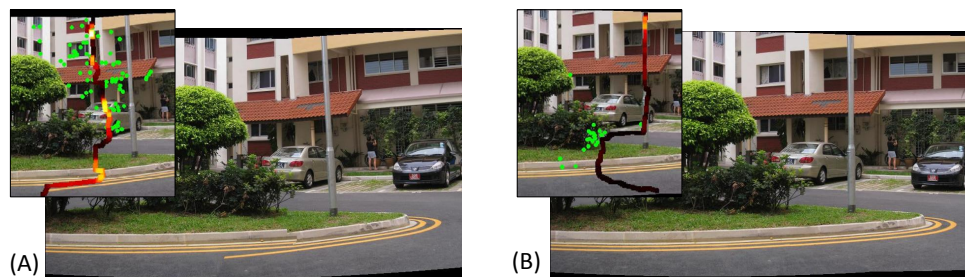[1]National University of Singapore   [2]The University of Adelaide

Figure 1: *(A) A traditional image-stitching result where the images are first aligned using a geometric transform computed based on the best overall fit of 96 matched points followed by seam-cutting. (B) Our seam-driven image stitching result where the geometric transform is less optimal in terms of geometric fit (only 26 matched points used), but the transform gives a perceptually better seam-cut. The insets show the points used to compute the transforms as well as the seam-cut with our computed seam error.*

Abstract

Image stitching computes geometric transforms to align images based on the best fit of feature correspondences between overlapping images. Seam-cutting is used afterwards to to hide misalignment artifacts. Interestingly it is often the seam-cutting step that is the most crucial for obtaining a perceptually seamless result. This motivates us to propose a seam-driven image stitching strategy where instead of estimating a geometric transform based on the best fit of feature correspondences, we evaluate the goodness of a transform based on the resulting visual quality of the seam-cut. We show that this new image stitching strategy can often produce better perceptual results than existing methods especially for challenging scenes.

1. Introduction and Motivation

Constructing panoramas from a set of overlapping images is a well-studied problem (e.g. see [BL07,Sze06]). Virtually all mainstream image stitching methods involve two steps. The first is to compute parametric transforms (usually planar perspective transforms or homographies) to align the images. This is done by matching local features across the images, followed by robust methods to estimate homographies with the best *geometric fit*.

Homography-based alignment assumes that the captured scene is far enough away from the camera to be treated as planar or that the camera is rotated about its center of projection to avoid parallax. Satisfying these conditions is rare in practice. As a result, there are typically misalignment ar-

tifacts that must be removed. To photo-realistically blend the images, a second post-processing step, such as seam-cutting [ADA*04], is applied. Interestingly, it is often the seam-cutting step that is the most crucial in producing perceptually good results.

We introduce the idea of seam-driven image stitching. Instead of selecting homographies based on the best geometric fit of matched feature points, a transform is evaluated based on the perceptual quality of the resulting seam-cut (see Figure 1). To achieve this, we propose a simple, yet effective, method to evaluate the seam cuts produced by different transforms. This seam-driven approach can often produce better results than current state-of-the-art methods for challenging cases where the input image sequence is captured under non-ideal imaging conditions.
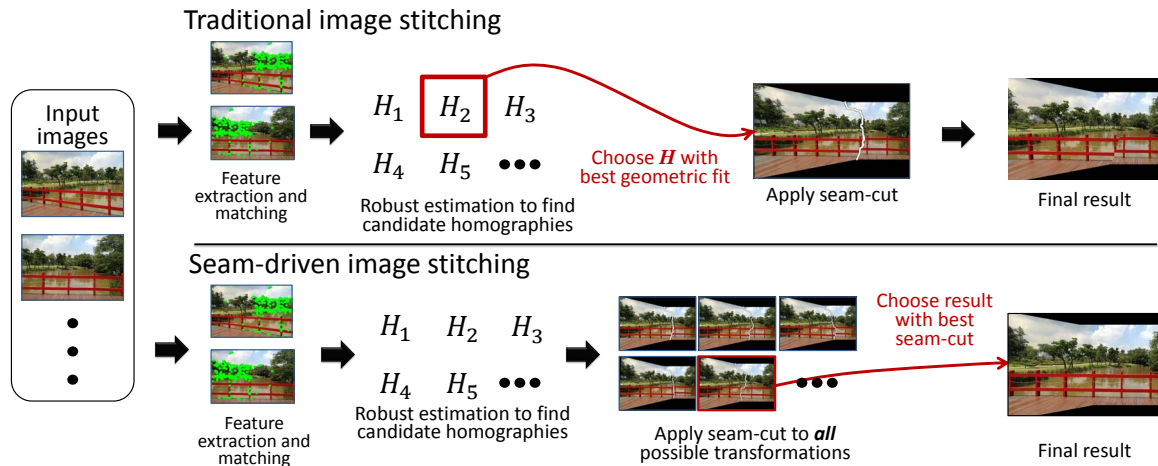
Figure 2: *A comparison between the traditional image stitching process and our seam-driven process.*

2. Seam-Driven Image Stitching

Figure 2 shows an example of the traditional image stitching pipeline (top) and the modified seam-driven image stitching pipeline (bottom). In the traditional approach, feature points such as SIFT [Low04] are computed and matched between image pairs resulting in a set of feature correspondences. Due to the non-planar scene geometry and erroneous matches, random sample consensus (RANSAC) [FB81] is used to robustly estimate the homography with the best geometric fit. Specifically, a set of homographies $H_1, H_2, \ldots$ are randomly hypothesized and ranked based on their goodness-of-fit (measured in terms of consensus size). The best homography is used to warp and align the images. Seam-cutting is then applied to produce the final result.

Our seam-driven approach operates in a similar fashion but evaluates and selects the homography differently. This is achieved by modifying the hypothesize-and-test loop in RANSAC, whereby we generate candidate homographies in a manner that encourages diversity, and applying seam-cut to the aligned images by using all candidates. Next, we evaluate the different seam-cuts based on our novel ranking metric that favors results with good perceptual quality.

2.1. Generating Homography Candidates

RANSAC generates model hypotheses by fitting a homography on randomly sampled minimal subsets of the data, where each subset contains four correspondences [BL07]. For each hypothesis, the correspondences that are consistent with it (i.e. the inliers of the hypothesis) are recorded. This random sampling process is repeated n times and the hypothesis with the highest consensus size is chosen.

The number n rarely corresponds to an exhaustive sampling of all four-subsets of the data. Moreover, usually there is a dominant plane in the scene (e.g. the background plane)

that contributes a large number of similar candidate homographies. Traditional image-stitching methods favor the dominant plane as it gives the best consensus; however, there are often several smaller planar structures in the scene whose matched points are treated as outliers to the domain plane. A simple filtering strategy is used to help find these other planes. First, standard RANSAC is applied with a comparatively large n (500 in our implementation) to fit the dominant plane. After this, we remove the inliers of the dominant plane and apply a sequence of RANSAC instances each with a smaller n (50 in our implementation) on the remaining matched points. This helps to find homographies that align other smaller planar structures which may result in better seam-cuts.

2.2. Computing the Seam-Cut

Seam cutting is applied to the overlapping regions of pairs of images (I_1 and I_2) aligned with the candidate homographies. The seam computation can be formulated as a labeling problem on a Markov Random Field (MRF) which minimizes a global energy with the following form:

$$E = \sum_p E_d + \lambda \sum_{(p,q) \in \mathbf{N}} E_s, \qquad (1)$$

where E_d is the data-cost energy reflecting the saliency of a pixel, p, with label l_p. The smoothness energy, E_s, measures the discontinuity of adjacent pixels, p and q, defined over a 4-connected neighborhood $\mathbf{N}$. The label l_p decides which image, I_1 or I_2, will appear in the overlapped region at each pixel p.

Following the formulation introduced by [ADA*04], the data-cost of each pixel is defined to be the gradient at that location:

$$E_d(p, l_p) = -\nabla I_{(l_p)}, \qquad (2)$$

where l_p decides which image gradient (i.e. ∇I_1 or ∇I_2) to
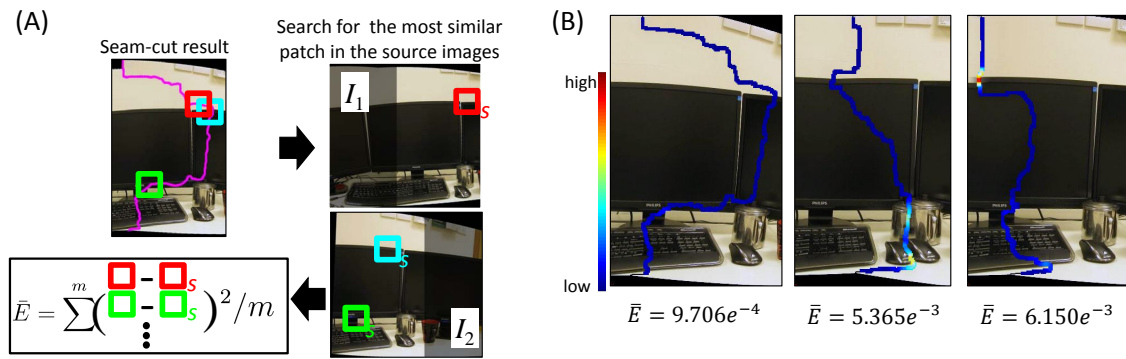
Figure 3: *(A) This illustrates how seams are evaluated. For each pixel along the seam, the squared difference of the most similar patch about this pixel found in either overlapping image is used as the mismatch error. (B) Shows a hot map of energy from different cuts using our error metric. The accumulated seam error, $\bar{E}$, is used to rank the cuts.*

use at position p. The smoothness cost between two pixels p and q is defined as:

$$E_s(p, l_p, q, l_q) = |l_p - l_q| \cdot (D(p) + D(q)), \quad (3)$$

which represents discontinuities between each pair of neighboring pixels. If $l_p = l_q$, the smoothness cost is 0; if $l_p \neq l_q$, the smoothness cost is defined as the difference D of the overlapped pixels, where D is:

$$D(v) = \|I_1(v) - I_2(v)\|^2 + \alpha \|\nabla I_1(v) - \nabla I_2(v)\|^2, \quad (4)$$

where $\alpha = 2$. Graph-cut optimization is used to assign the labels to our MRF [BVZ01].

2.3. Evaluating the Cut

While the seam-cut energy minimizes image gradient between I_1 and I_2, it is ineffective when used to rank the perceptual quality of different cuts. We introduce a more discriminating error measurement. For each pixel, p, along the seam, we estimate an error value, $E(p)$ by extracting a 17×17 patch, P, centered at p, and searching for its most similar patch in either I_1 or I_2. This can be expressed as:

$$E(p) = \min_{S_i \in I_1, I_2} ||P - S_i||^2, \quad (5)$$

where S_i represents all image patches in the overlapping regions of I_1 and I_2. The idea is that a patch along the seam is perceptually plausible if it resembles a patch found in either I_1 or I_2. If a patch along the seam cannot be found in either source images, it is likely to be an artifact and therefore assigned a larger error. The total error along a seam of size m, is computed as $\bar{E} = \sum_p^n E(p)/m$.

Figure 3(A) and (B) shows a diagram of this procedure as well as an example with three different cuts with the corresponding per pixel error, $E(p)$ shown as a hot map and the total error $\bar{E}$. The minimum error is indicative of our perceptual ranking.

3. Results and Discussion

Figure 4 shows our results and those obtained using and Photoshop CS6 which is based on [ADA*04]. The four input images contain parallax and therefore are difficult to align. Photoshop uses a traditional image-stitching pipeline to select the best fit transform. There are still noticeable artifacts even after seam-cutting. In contrast, our seam-driven strategy is able to obtain a perceptually better result. Also see accompanying supplemental material for this paper.

Inline with the short paper theme, our seam-driven image stitching idea is a work in progress. While the error measurement defined in Section 2.3 allows us to rank different seam-cut results, there may be better ways to formulate this. Also, traditional image stitching methods perform a bundle adjustment step where the collection of homographies are adjusted to provide a global fit to the matching feature points. Applying bundle adjustment within our framework is not as straight forward and our approach is therefore limited to being applied in an incremental fashion. This could result in our method selecting a homography that provides a good seam-cut locally, but has adverse affect on subsequent homography estimation in other overlapping images. These are interesting areas that warrant further investigation.

References

[ADA*04] AGARWALA A., DONTCHEVA M., AGRAWALA M., DRUCKER S., COLBURN A., CURLESS B., SALESIN D., COHEN M.: Interactive digital photo-montage. *ACM Transactions on Graphics (SIGGRAPH)* (2004). 1, 2, 3

[BL07] BROWN M., LOWE D.: Automatic panoramic image stitching using invariant features. *International Journal of Computer Vision (IJCV)* (2007). 1, 2

Figure 4: *Comparison of panoramas constructed based on a traditional image stitching pipeline and our seam-driven approach. There are noticeable visual artifacts in the traditional results, while our results appear more seamless.*

[BVZ01] BOYKOV Y., VEKSLER O., ZABIH R.: Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)* (2001). 3

[FB81] FISCHLER M. A., BOLLES R. C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM* (1981). 2

[Low04] LOWE D. G.: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision (IJCV)* (2004). 2

[Sze06] SZELISKI R.: Image alignment and stitching: a tutorial. *Foundations and Trends in Computer Graphics and Vision 2* (2006). 1