https://doi.org/10.2352/ISSN.2470-1173.2019.11.IPAS-256

Image Stitching by Creating a Virtual Depth

Ahmed H. Eid, Brian E. Cooper, Tomasz J. Cholewo; Lexmark International, Inc.; Lexington, Kentucky, USA

Abstract

A method for image stitching is presented. The approach focuses on images with parallax (depth variation) to create panoramic views with high fidelity. The approach creates the stitching seam at a virtual depth to convert hard stitching problems to simple ones. The virtual depth is created by applying local distortions to the input images at the stitching seam so that the contents visually appear to be located at the same depth. The presented approach targets a wide variety of applications that require generating high (or super) resolution, wide-view images. These applications include tele-presence (or tele-reality) applications such as shopping, touring, conferencing, planning or architecting, learning, inspection, and surveillance. Our results show that the proposed approach provides promising results compared to commercial products that rely on stitching solutions.

Introduction

Capturing a wide field-of-view image at high resolution is difficult to achieve using a single static camera. Image mosaicking techniques are traditionally used to create such wide view images from sequences of high-resolution images that cover the scene while providing enough overlap between images. Applications of image mosaicking include object identification, mapping, tele-presence (e.g. tele-shopping, distant-learning, etc.)

Basic approaches for image stitching use 2D transformations between images to determine the stitching seam. A global 2D image transformation (e.g. homography) can provide perfect stitching results, but only for planar scenes or if the depth variations are relatively small compared to the distance between camera(s) and objects (i.e. objects are located at approximately the same distance from the camera). However, if the scene has relatively large depth variations, as do most of the applications we are targeting, then stitching based on a global 2D homography often introduces objectionable artifacts, by duplicating or omitting details in the scene. This usually happens if the stitching seam intersects objects located at different depths within the individual images.

Figures 1(a) and 1(b) show left and right input images, respectively, with the stitching seam optimized for stitching front objects. The seam is overlaid (in red) on top of the images. The stitched image in Figure 1(c) shows perfect stitching for front objects. However, stitching artifacts (duplicate contents) appear at back objects. On the other hand, Figures 2(a) and 2(b) show images with a stitching seam that is optimized for back objects. The stitched image in Figure 2(c) shows perfect stitching at back objects, but some details of front objects are missing.

Other mosaicking approaches create panoramic images using video sequences with dense sampling to reduce the effect of depth parallax. However, in applications such as tele-operation, dense sampling is often impractical due to bandwidth limitations (especially when several cameras are involved) or due to constraints on the scene or camera(s) locations. These limitations motivated other approaches to estimate a full depth map from sparse images. Dense depth maps of the entire scene can be used later to generate synthetic views in-between the original views to achieve the dense sampling requirement. However, such approaches require extensive computation of dense depth maps of the scene.

(a) (b) (c)

Figure 1: Stitching seam optimized for frontal objects. (a) Left image and (b) right image show the same corresponding seam. (c) Stitched image shows perfect stitching at frontal objects, but duplicated contents for back objects.

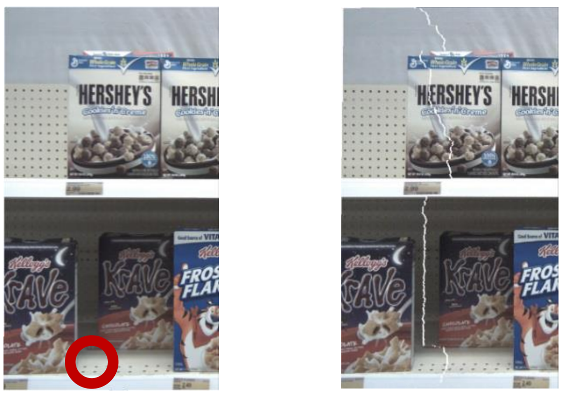
(a) (b) (c)

Figure 2: Stitching seam optimized for back objects. (a) Left image view and (b) right image view show the same corresponding seam. (c) Stitched image shows perfect stitching at back objects, but missing contents for front objects.

Recent approaches, including [1, 2, 3] and ones found in Photoshop [4], search for the optimal stitching seam that minimizes stitching artifacts. Seams that avoid salient features in the image tend to avoid stitching images at areas of different depths, effectively hiding the stitching artifacts. Of course, the stitching seam must pass through areas with negligible details.

Figure 3 illustrates this problem, where the left and right images (Figures 1(a) and 1(b), respectively) are stitched with very good quality, except for a small region at the bottom of the image, shown in Figure 3(a). Figure 3(b) shows this stitching seam, optimized to pass through frontal objects while avoiding highly structured objects located at different depths.

However, this approach fails when the overlapping areas of the input images lack featureless regions. This can cause prominent stitching artifacts, as illustrated in Figure 4. Note that the overlapping area in the left and right images (Figures 4(a) and 4(b), respectively) mostly has highly structured contents. As a result, artifacts appear in the stitched image in Figure 4(c). The stitching seam in Figure 4(c) could not avoid passing through highly structured areas located at depths different than the optimal depth.



(a) (b)

Figure 3: Photoshop stitching. (a) Stitched images (from Figures 1(a) and 1(b)) show very good stitching except at a small region (circled) at the bottom of the image. (b) The same stitched image shows the stitching seam passing through a featureless area (lower half of the image) to avoid visible stitching artifacts.

(a) (b) (c)

Figure 4: Photoshop stitching. (a) Left and (b) right images lack a featureless area, forcing the stitching seam to pass through highly structured areas located at depths different than the optimal depth. (c) Stitched image shows duplicated content (circled) in the back objects.

Some parallax-tolerant image stitching methods allow small physical distortions in the image to achieve "perceptual plausibility rather than accurate reconstruction" [5]. Such content-aware image resizing or warping can accommodate parallax more effectively than homography alone [1]. We adopt a similar philosophy, by allowing small, local (linear) image distortions. Such simple expansion or contraction of objects helps create a consistent albeit virtual depth in the stitched image.

In this paper, a novel method for stitching images is described. The approach can stitch images that may contain objects located at different depths regardless of the complexity of the structure of the contents of the scene. The approach does not require a dense sampling of video inputs, as it applies to sparse images as well. Moreover, given a set of sparse images, the approach does not require a full depth estimation of the entire scene. Only one requirement exists: finding the relative depth along an arbitrary seam in the overlapping area of a pair of images to be stitched. This seam can simply be a straight line. The relative depth along that seam can be calculated using well-known approaches of stereo matching that can be easily found in the computer vision literature. For example, area-matching approaches can search the right image to find the corresponding pixels that match pixels on the seam of the left image [4, 6].

This paper is organized as follows. The methods section describes the proposed approach. In the experimental results section we compare the proposed approach to a seam-cut approach, with additional results showing the high quality of stitching at different depths. Finally, we summarize our conclusions and possible future improvements of the proposed approach.

Methods

Without loss of generality, we assume that the input image pairs have a simple stereo configuration [6]. This can be constructed by two cameras shifted in space, or by a single camera moving in one dimension. General or complex, stereo configurations can be reduced to the simple stereo configuration using image rectification approaches (for example, [7, 8, 9]).

Knowing the 2D corresponding points in the left and right images leads to computing the depth of a 3D point in space. For example, the red vertical line (points X_L) overlaid on the left image of the stereo pair (Figure 5(a)) has its corresponding points in the right image (Figure 5(b)) marked in red, but these points no longer belong to a single line. This occurs because the red line in the left image passes through points that have different depths. The corresponding red points (X_R) in the right image should have different x-coordinates to reflect the differences in their depths. For example, points close to the right green line in the right image (Figure 5(b)) have disparity ($d = X_L - X_R$) smaller than points that are closer to the left green line. Therefore, the former points are deeper (have greater depth) than the latter points. (Ignore points in the featureless areas that show some matching errors.)

It is typically trivial to stitch the two images if the points on the red line in the left image have the same depth as the corresponding points in the right image. These points will also form a vertical straight line in the right image, having the same disparity and the same depth. In this hypothetical case, the two images in Figure 5 could be joined trivially, by stitching at the corresponding lines in the two images. However, this is not the case in Figure 5, since some of the corresponding points in the right image do not belong to a single line. This leads to the basic idea of the presented stitching approach:

Move the corresponding points in the right image to make them form a single line, using a relevant transformation (i.e. create a line with virtual depth).

This virtual depth line, X_{vir}, as shown in Figure 6, can be any line in the overlapping area in the right image. However, for better image quality, X_{vir} should be close to the X_R values. Examples of X_{vir} include: Average (X_{Ri}), Median (X_{Ri}), Maximum (X_{Ri}), or Minimum (X_{Ri}).

To determine the areas in the right image to be modified, we need to determine the boundaries of these areas. Figure 6(a) shows examples of such areas (the four regions G_1, G_2, G_3 and G_4). The left boundaries of these areas should be the x-coordinates, $X_{Ri,}$ of the corresponding points. The right boundary should be equal to or greater than X_{max} = Maximum (X_{Ri}). The widths W_{Ri} of these areas (illustrated in Figure 6(a)) are below. (N is the number of regions.)

$$W_{Ri} = X_{max} - X_{Ri}, \qquad 1 \le i \le N$$

The height of each area can be as low as one, assuming X_R can vary for each individual row. Since our goal is to move pixels from locations X_{Ri} to a virtual location X_{vir}, the target width W_{vir} of the new areas (labeled G_1', G_2', G_3' and G_4' in Figure 6(b)) can be calculated as:

$$W_{vir} = X_{max} - X_{vir} + 1, \qquad 1 \le i \le N$$

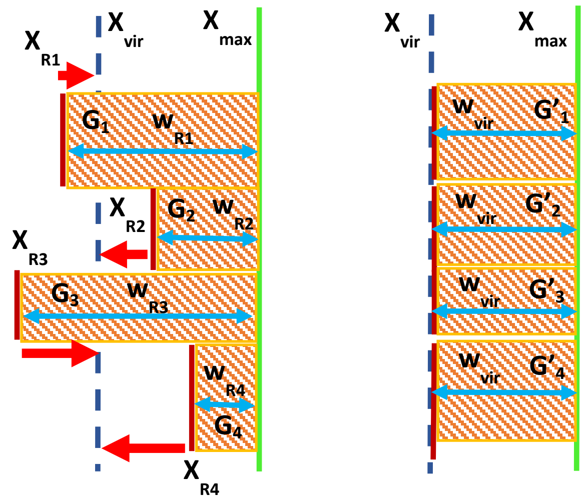
So, the transformation needed to move X_R to X_{vir} can be, for example, a horizontal scaling S_i

$$S_i = \frac{W_{vir}}{W_{Ri}}, \qquad 1 \le i \le N,$$

where this scaling factor varies according to the depth. For example, points with larger X_R (right of X_{vir}), i.e. with larger depth, will have scaling factors ≥ 1. This means that these far objects need to be expanded. On the other hand, points with smaller X_R (left of X_{vir}), i.e. with smaller depth, will have scaling factors ≤ 1. This means that these near objects need to be shrunk.

(a) (b)

Figure 5: Depth estimation. (a) Left image shows a vertical stitching seam (red). (b) Right image marks the corresponding points in red. Values of X_R must fall between the green lines. Larger X_R values (closer to the right green line) have a larger depth (smaller disparity), whereas smaller X_R values (closer to the left green line) have a smaller depth (larger disparity).



(a) (b)

Figure 6: Creating a virtual depth for stitching. (a) Point correspondences X_R in the right image appear at various positions (red lines), corresponding to different depths. (b) A geometric transformation moves X_R to the virtual stitching seam X_{vir}.

This is consistent with the fact that with central (perspective) projection, far objects look small, and near objects look large. That is why stitching at front (near) objects (see Figure 1) provides good results only for front objects. Since far (back) objects are smaller in size, parts of these objects appear twice, once from the left image and again from the same region in the right image, as shown in Figure 1(c). On the other hand, stitching based on far object (see Figure 2) provides good results only for far objects. Since near objects are larger in size, parts of these objects are missed (Figure 2(c)). Based on these facts and observations, the presented method creates a virtual depth by applying a relevant local transformation to expand far objects and shrink near objects to appear as if they were at the same depth, hence permitting the trivial stitching.

As might be noticed, the local transformations we use to create the virtual depth are non-rigid. This means some local distortion may occur. However, the local distortion may not be noticed when the images are close together. For images that are farther apart, we can diffuse this distortion within a larger area. For example, the X_{max} line can be shifted to the right, so the scaling can apply to a larger region. Mathematically, shifting X_{max} with a positive shift D leads to:

$$S_i = \frac{W_{vir}+D}{W_{Ri}+D} = \frac{(W_{vir}/D)+1}{(W_{Ri}/D)+1},$$

where S_i approaches one, as D becomes large. Practically, moderate values of D could be enough to make the distortion unobjectionable. An example of the effect of D on the visibility of the added distortion is shown in Figure 7, where using small D may show some distortion as in Figure 7(a). A moderate value of D may reduce the visibility of such distortion as shown in Figure 7(b).

The presented approach uses common stereo matching techniques to find corresponding points in a stereo pair, so false matching can happen in areas with less distinct features as shown at the bottom of Figure 4(b). This limitation does not affect the image quality of the presented approach, as stitching at false depth in featureless areas typically does not produce noticeable artifacts. As described earlier, seam-cut approaches (including Photoshop) rely on stitching images in such areas to avoid showing objectionable artifacts.

(a) (b)

Figure 7: Increasing X_{max} (shifting by D) to improve stitching quality. (a) A small shift could cause some visible distortion: shrinkage at front objects and expansion at back objects. (b) A moderate shift diffuses the distortion to a larger area, making it less visible.

To make stitching less visible, image pairs may need to be blended. A blending function can be applied to the whole area of overlap or to a small area around the stitching seam. Regional blending is preferred if images are far apart to eliminate ghosting artifacts that may appear due to differences in views (due to large parallax) in the overlapped areas. In our experiments we use local linear blending of the left and right images in the overlapped areas.

Experimental Results

Comparing the proposed approach to seam-cut techniques, we use Adobe Photoshop. Figure 8 shows an example, where the left and right images of a bookshelf show depth variations. First, we estimate the depth as shown in Figure 8(a) and 8(b). The stitching result of the proposed approach is shown in Figure 8(c), while the result of Adobe Photoshop is shown in Figure 8(d). Both approaches provide good results, with less noticeable artifacts.

Another example is shown in Figure 9, where the corresponding points are overlaid on the left and right images of Figure 9(a) and 9(b), respectively. The proposed approach provides good stitching results, although some shrinkage appears (circled) in Figure 9(c). However, the seam-cut approach (using Adobe Photoshop) shows visible artifacts, circled in Figure 9(d).

To study the effect of stereo matching accuracy on the results of the proposed approach, we downscaled versions of the input images. Although downscaling input images can expediate the stereo matching process, it may introduce ghosting in the blending area due to less accuracy in estimating the depth. Figure 10(a) shows the corresponding points (in red) estimated using images scaled by 0.1, while Figure 10(b) shows the corresponding points estimated using images scaled by 0.5. As the corresponding points estimated at 0.1 scaling factor are smoother, the stitching results show ghosting artifacts due to less accuracy in the estimated virtual depth. Figure 10(c) shows this behavior. On the other hand, more accurate results are shown in Figure 10(d) due to accurate depth estimation, corresponding to Figure 10(b).

To show the effect of lighting conditions on the accuracy of the stitching results of the proposed approach in a large set of images, a moving camera with translational motion captured a set of 400 images under normal room lighting. The result of stitching such images is shown in Figure 11(a). The scene was next recaptured using an additional lighting source attached to the camera. The results are shown in Figure 11(b). Comparing Figures 11(a) and 11(b), the stitching results remain accurate when providing an additional light source due to the accuracy of depth estimation even though reflective surfaces could complicate stereo matching approaches to find the depth. The darker areas are affected the most, with the small bottles and jars toward the right of Figure 11(a).

Conclusions and Future Work

In this paper we proposed a new approach to stitch images captured with cameras with translational motion. The approach creates a virtual common depth to facilitate stitching images with parallax. Our experiments show promising results in stitching images with parallax compared to seam-cut approaches, especially for images with highly structured contents at different depths.

In future work, we will investigate variants for the presented approach. For example, similar geometric transformations can be applied to some regions around the stitching seam in the left image as well as the right image. This is desirable in some cases to make distortion less objectionable if it is split between images. Other variants could permit X_{max} and/or X_{vir} to be a general 1-D curve, rather than a straight line.

References

[1] F. Zhang and F. Liu, "Parallax-tolerant image stitching," in IEEE Conference on Computer Vision and Pattern Recognition, pp. 3262–3269, 2014.
[2] J. Gao, Y. Li, T-J.Chin, and M.S.Brown, "Seam-driven image stitching," in Eurographics (Girona, Spain), pp. 45-48, 2013.
[3] V. Kwatra, A. Schödl, I. Essa, G. Turk, and A. Bobick, "Graphcut textures: image and video synthesis using graph cuts," in ACM SIGGRAPH (San Diego, CA), Vol. 22, No. 3, pp. 277-286, 2003.
[4] A. Agarwala, M. Dontcheva, M. Agarwala, S. Drucker, A. Colburn, B. Curless, D. Salesin, and M. Cohen, "Interactive digital photomontage," in ACM SIGGRAPH (Los Angeles, CA), Vol. 23, No. 3, pp. 294-302, 2004.
[5] F. Liu, H. Jin, M. Gleicher, and A. Agarwala, "Content-preserving warps for 3D video stabilization," in ACM SIGGRAPH (New Orleans, LA), Vol. 28, No. 3, pp. 44:1-44:9, 2009.
[6] E. Trucco and A. Verri, Introductory Techniques for 3-D Computer Vision, Upper Saddle River, NJ: Prentice-Hall, 1998.
[7] Open Source Computer Vision Library, www.opencv.org.
[8] R.I. Hartley, "Theory and practice of projective rectification," Intl. J. Computer Vision, Vol. 35, No. 2, pp. 115-127, 1999.
[9] C. Loop and Z. Zhang, "Computing rectifying homographies for stereo vision." in IEEE Conference on Computer Vision and Pattern Recognition (Fort Collins, CO), pp. 125–131, 1999.

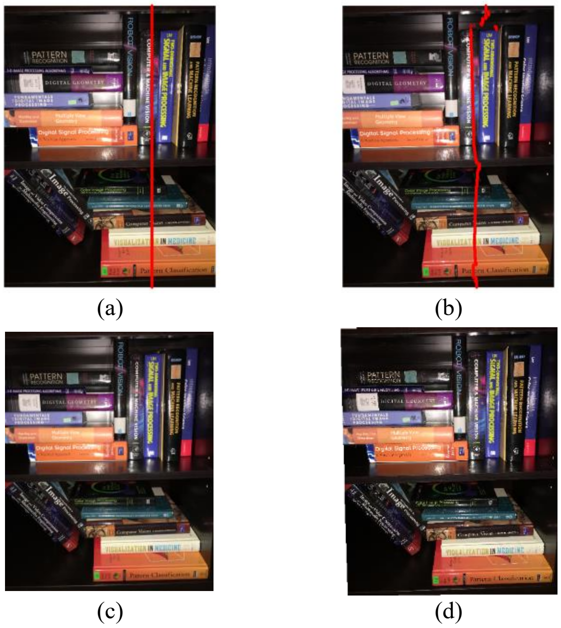


Figure 8: Stitching comparison of bookshelf. (a) Left and (b) right images show the same corresponding stitching seam, with stitched results from (c) the proposed approach and (d) Photoshop.

Figure 9: Stitching comparison of store shelf. (a) Left and (b) right images show the same corresponding stitching seam, with stitched results from (c) the proposed approach and (d) Photoshop.

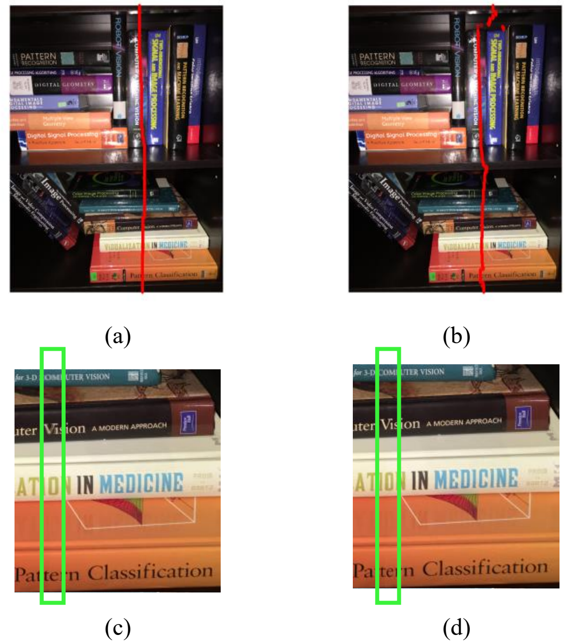


Figure 10: Stitching comparison on downscaled images. The stitched images downscaled at scale factors of (a) 0.1 and (b) 0.5 appear comparable at a distance. However, the corresponding zoomed images (c) and (d) reveal differences. The former image (0.1 scaling factor) shows stitching ghosts due to less accurate depth estimation. The latter image (0.5 scaling factor) provides better stitching quality due to more accurate depth estimation.

Figure 11: Stitching comparison with different lighting. (a) The stitched image with normal room lighting has darker areas (right) that are more challenging for depth estimation. (b) The stitched image with an additional light source attached to the camera performs better.