Adaptive As-Natural-As-Possible Image Stitching

Chung-Ching Lin, Sharathchandra U. Pankanti, Karthikeyan Natesan Ramamurthy, and Aleksandr Y. Aravkin
IBM Thomas J. Watson Research Center.

The goal of image stitching is to create natural-looking mosaics free of artifacts that may occur due to relative camera motion, illumination changes, and optical aberrations. Parallax error in overlapping areas is one of the main challenges. One of the first approaches that estimates a smooth stitching field is the smoothly varying affine (SVA) stitching proposed in Lin *et al.* [3]. A global affine transform is estimated, which is then relaxed to form a smoothly varying affine stitching field. It is flexible enough to handle parallax while retaining the extrapolation and occlusion handling properties of parametric transforms. However, it fails to impose global projectivity. This drawback is alleviated by the As-Projective-As-Possible (APAP) approach proposed in [4], which estimates a smoothly varying projective stitching field and hence provides excellent alignment accuracy. A simple moving Direct Linear Transformation (DLT) method is used to estimate the local parameters, by providing higher weights to the closer feature points and lower weights to the farther ones.

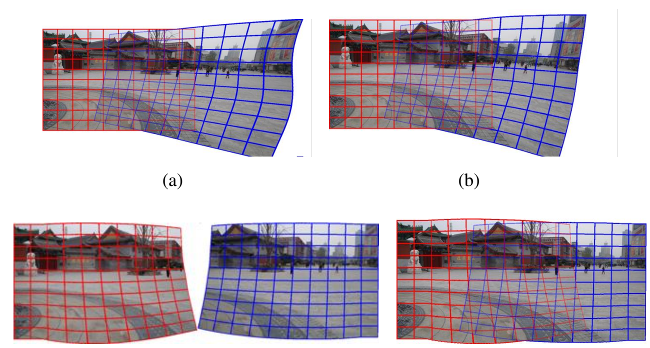
Since APAP extrapolates the projective transform in the non-overlapping regions, it introduces severe perspective distortions in regions far from the boundary. The authors in [1], propose the shape-preserving half-projective (SPHP) warp to preserve shapes in the non-overlapping areas. The stitching provides for shape preservation, but does not guarantee against parallax. Although the combination of SPHP and APAP can be claimed to the state-of-the-art approach, it is sensitive to parameter selection. Furthermore, if the overlapping areas has multiple distinct planes, deriving a single global similarity transformation from the global homography may lead to undesirable and unnatural visual effects in the mosaic.

We propose a new method that incorporating several assumptions to make the panorama look more natural. To mitigate perspective distortion that occurs in APAP, we linearize the homography in the regions that do not overlap with any other image. We then automatically estimate a global similarity transform using a subset of corresponding points in the overlapping regions. Finally, we extrapolate smoothly between the homography and the global similarity in the overlapping regions, and using the linearized homography (affine) and the global similarity transform in the non-overlapping regions. The smooth combination of two stitching fields (homography/linearized homography and global similarity) help us achieve: (a) a fully continuous and smooth stitching field with no bending artifacts, (b) improved perspective in the non-overlapping regions using a global similarity transform, (c) full benefits of the state-of-the-art alignment accuracy offered by APAP.

Within the overlapping areas, a local homography is estimated using moving DLT framework[4], $\mathbf{h}_j = \underset{\mathbf{h}}{\operatorname{argmin}} \|\mathbf{W}_j \mathbf{A} \mathbf{h}\|^2$. Here, the weight matrix, $\mathbf{W}_j = \operatorname{diag}([\omega_{1,j}\ \omega_{1,j}\ \dots\ \omega_{N,j}\ \omega_{N,j}])$, the elements of the observation matrix, $\mathbf{A}$, are functions of the corresponding points, and $\mathbf{h}_j$ is a 9×1 vector containing the parameters of the homography transformation. The weights are obtained using the Gaussian kernel between the points $\mathbf{p}_i$ and $\mathbf{p}_j$: $\omega_{i,j} = \exp(-\|\mathbf{p}_i - \mathbf{p}_j\|^2/\sigma^2)$. Note that the local homography can be computed only in the regions of the target image that overlap with the reference image.

To mitigate perspective distortion on non-overlapping areas, the local transformation is obtained using the linearized homography transformations. For a set of R anchor points $\{\mathbf{p}_i\}_{i=1}^R$ at the boundary with possibly different local homographies, the weighted combination of linearizations is given as $\mathbf{h}^L(\mathbf{q}) = \sum_{i=1}^R \alpha_i \left(\mathbf{h}(\mathbf{p}_i) + \mathbf{J}_\mathbf{h}(\mathbf{p}_i)(\mathbf{q} - \mathbf{p}_i)\right)$, where $\mathbf{J}_\mathbf{h}(\mathbf{p})$ is the Jacobian of the homography $\mathbf{h}$ at the point $\mathbf{p}$. We assume α_i to be a function of $\|\mathbf{q} - \mathbf{p}_i\|$, and in particular we consider the Gaussian weighting or the Student's t-weighting where $\alpha_i = \left(1 + \frac{\|\mathbf{q} - \mathbf{p}_i\|^2}{\nu}\right)^{\frac{-(\nu+1)}{2}}$. Student's t-weighting is more robust since that tail of the distribution decays slowly compared to Gaussian and hence when $\mathbf{q}$ is far from anchor points, all the anchor points are given similar weighting. However, if Gaussian weighting is chosen, the tail should be made flat at the offset parameter γ to avoid "wavy" effects (e.g. see Fig. 1(a)).

The keypoint matches are grouped using recursive RANSAC [2]. The global similarity transformation $\mathbf{S}$ is obtained by the group of matches that can calculate the similarity transformation with smallest rotation angle. Then, we gradually update the local transformations of entire target image to the global similarity transformation using the following equation: $\hat{\mathbf{H}}_i^{(t)} = \mu_h \mathbf{H}_i^{(t)} + \mu_s \mathbf{S}$, where $\mathbf{H}_i^{(t)}$ is i^{th} local homography, $\hat{\mathbf{H}}_i^{(t)}$ is updated local transformation, $\mathbf{S}$ is the global similarity transformation, μ_h and μ_s are weighting coefficients, and the superscript (t) refers to the target image. μ_h and μ_s are computed as: $\mu_h(i) = \langle \overrightarrow{\kappa_m p(i)}, \overrightarrow{\kappa_m \kappa_M} \rangle / \left|\overrightarrow{\kappa_m \kappa_M}\right|$, $\mu_s(i) = 1 - \mu_h(i)$, where κ is the projected point of warped target image on the $\overrightarrow{o_r o_t}$ direction. o_r and o_t are the center points of the reference image and the warped target image. κ_m and κ_M are the points with smallest and largest value of $\langle \overrightarrow{o_r p(i)}, \overrightarrow{o_r o_t} \rangle$ respectively. Here, $p(i)$ is the location of the i^{th} location in the final panorama. Then, the local transformation of the reference image can be obtained as $\hat{\mathbf{H}}_i^{(r)} = \hat{\mathbf{H}}_i^{(t)} (\mathbf{H}_i^{(t)})^{-1}$.



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

Figure 1: Illustration of proposed algorithm. (a) Warp after applying moving DLT with Gaussian weighting. (b) Extrapolation of non-overlapping areas using homography linearization and Student's t-weighting, (c) Proposed final warps after integrating global similarity transformation, and (d) Final stitched result.

[1] Che-Han Chang, Yoichi Sato, and Yung-Yu Chuang. Shape-preserving half-projective warps for image stitching. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 3254–3261. IEEE, 2014.

[2] M.A. Fischler and R.C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and aarticleutomated cartography. 1981.

[3] Wen-Yan Lin, Siying Liu, Yasuyuki Matsushita, Tian-Tsong Ng, and Loong-Fah Cheong. Smoothly varying affine stitching. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 345–352. IEEE, 2011.

[4] Julio Zaragoza, Tat-Jun Chin, Michael S Brown, and David Suter. As-projective-as-possible image stitching with moving dlt. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 2339–2346. IEEE, 2013.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage. This work is sponsored in part by Defense Advanced Research Projects Agency, Microsystems Technology Office (MTO), under contract no. HR0011-13-C-0022. The views expressed are those of the authors and do not reflect the official policy or position of the Department of Defense or the U.S. Government. This document is: Approved for Public Release, Distribution Unlimited.