

Multi-Viewpoint Panorama Construction with Wide-baseline Images Supplementary Document

Guofeng Zhang¹ Yi He¹ Weifeng Chen¹ Jiaya Jia² Hujun Bao¹

¹State Key Lab of CAD&CG, Zhejiang University

²The Chinese University of Hong Kong

I. MORE DETAILS ABOUT FEATURE ALIGNMENT TERM

To be accurate, each grid of the mesh defines a local homography by four corner vertices, which should be applied to the pixels inside. With a linear approximation, we can represent one point as a weighted sum of the grid corners:

$$p' = (v'_1 \quad v'_2 \quad v'_3 \quad v'_4) \cdot (w_1 \quad w_2 \quad w_3 \quad w_4)^\top,$$

where p' is a point at the original image coordinates, v'_1, v'_2, v'_3 and v'_4 are enclosing vertices of p' . The real position of warped p' is defined by the homography:

$$\begin{aligned} p = Hp' &= H \begin{bmatrix} x'_1 & x'_2 & x'_3 & x'_4 \\ y'_1 & y'_2 & y'_3 & y'_4 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \end{bmatrix} \\ &= \begin{bmatrix} x_1 & x_2 & x_3 & x_4 \\ y_1 & y_2 & y_3 & y_4 \\ z_1 & z_2 & z_3 & z_4 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \end{bmatrix}, \end{aligned}$$

which leads to the inhomogeneous coordinates of p as

$$\begin{aligned} p_x &= \frac{x_1w_1 + x_2w_2 + x_3w_3 + x_4w_4}{z_1w_1 + z_2w_2 + z_3w_3 + z_4w_4}, \\ p_y &= \frac{y_1w_1 + y_2w_2 + y_3w_3 + y_4w_4}{z_1w_1 + z_2w_2 + z_3w_3 + z_4w_4}, \end{aligned}$$

where x_i, y_i, z_i are homogeneous coordinates of v'_i after warping. Note that p is exactly the weighted sum of the warped corners if we use homogeneous coordinates. In our linear approximation, we assume that the weights also hold for the inhomogeneous representation.

$$\hat{p} = \begin{bmatrix} \frac{x_1}{z_1} & \frac{x_2}{z_2} & \frac{x_3}{z_3} & \frac{x_4}{z_4} \\ \frac{y_1}{z_1} & \frac{y_2}{z_2} & \frac{y_3}{z_3} & \frac{y_4}{z_4} \\ z_1 & z_2 & z_3 & z_4 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \end{bmatrix},$$

The introduced error can be calculated as:

$$p - \hat{p} = \left(\begin{bmatrix} \frac{x_1}{\bar{z}} & \frac{x_2}{\bar{z}} & \frac{x_3}{\bar{z}} & \frac{x_4}{\bar{z}} \\ \frac{y_1}{\bar{z}} & \frac{y_2}{\bar{z}} & \frac{y_3}{\bar{z}} & \frac{y_4}{\bar{z}} \\ \bar{z} & \bar{z} & \bar{z} & \bar{z} \end{bmatrix} - \begin{bmatrix} \frac{x_1}{z_1} & \frac{x_2}{z_2} & \frac{x_3}{z_3} & \frac{x_4}{z_4} \\ \frac{y_1}{z_1} & \frac{y_2}{z_2} & \frac{y_3}{z_3} & \frac{y_4}{z_4} \\ z_1 & z_2 & z_3 & z_4 \end{bmatrix} \right) \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \end{bmatrix}, \quad (1)$$

where $\bar{z} = z_1w_1 + z_2w_2 + z_3w_3 + z_4w_4$. We can deduce from Eq. (1) that the error is subtle when z_i is close to \bar{z} , which is true in two conditions: the original vertices are close to each other (with mesh grids dense enough), or the perspective component of H is weak. The assumption is true in most of the typical panorama cases.

Datasets	RMSE	ϵ_1	ϵ_2	ϵ_3	
Desktop1	-APAP	7.63	44.19	28.50	20.46
	-OURS	1.6	30.34	17.05	9.5
Desktop2	-APAP	4.49	27.54	13.99	9.72
	-OURS	1.34	23.48	10.97	7.06
Building1	-APAP	4.15	41.45	15.62	9.74
	-OURS	1.31	38.54	10.10	5.96
Building2	-APAP	5.24	10.79	4.38	2.67
	-OURS	1.49	7.30	2.47	1.35

TABLE I: RMSE and outliers statistics.

II. QUANTITATIVE EVALUATION WITH WIDE-BASELINE IMAGE PAIRS

We follow the method of [1] to evaluate results quantitatively. For pairwise stitching, we quantify the alignment error of the estimated warp $f : R^2 \rightarrow R^2$ by the root mean squared error (RMSE) of corresponding feature points, where $RMSE(f) = \sqrt{\frac{1}{N} \sum_{i=1}^N \|f(x_i) - x_i\|^2}$. We also compare pixel-wise difference quantitatively. Following [2], [1], we define a pixel x as an outlier if there is no similar pixel (intensity difference less than a given threshold η) within the 4-pixel radius of the warped point. We calculate the percentage of outliers in the overlapped area with three thresholds, i.e. $\eta = 10, 20$ and 30 , respectively. The computed corresponding outlier percentages are denoted as ϵ_1, ϵ_2 and ϵ_3 , respectively.

For our wide-baseline images pairs, since the number of the matched features is already small, we use all matches and evaluate the whole RMSE and outlier percentage. For fair comparison, we select the first frame as reference, same as that in [1]. In this case, most prior constraints are unnecessary. So we only use feature alignment and regularization terms to construct the energy function, i.e. $E(V) = E_A(V) + \lambda_R E_R(V)$. Figure 1 shows the stitched image pairs by APAP and our method with the same feature matches. Table I shows the computed RMSE and outlier percentages on four image pairs. Our method yields obviously lower errors than APAP [1] in these wide-baseline image pairs. APAP uses a fixed interpolation weight for multi-homography representation, which is easily affected by uneven feature distributions. As shown in Figure 1 (top row), some regions near image boundaries have significant alignment errors. In contrast, our method does not have this problem. The reason is that our regularization term can better propagate the warp transformation to the neighboring regions, which reduces the influence of uneven feature distributions.

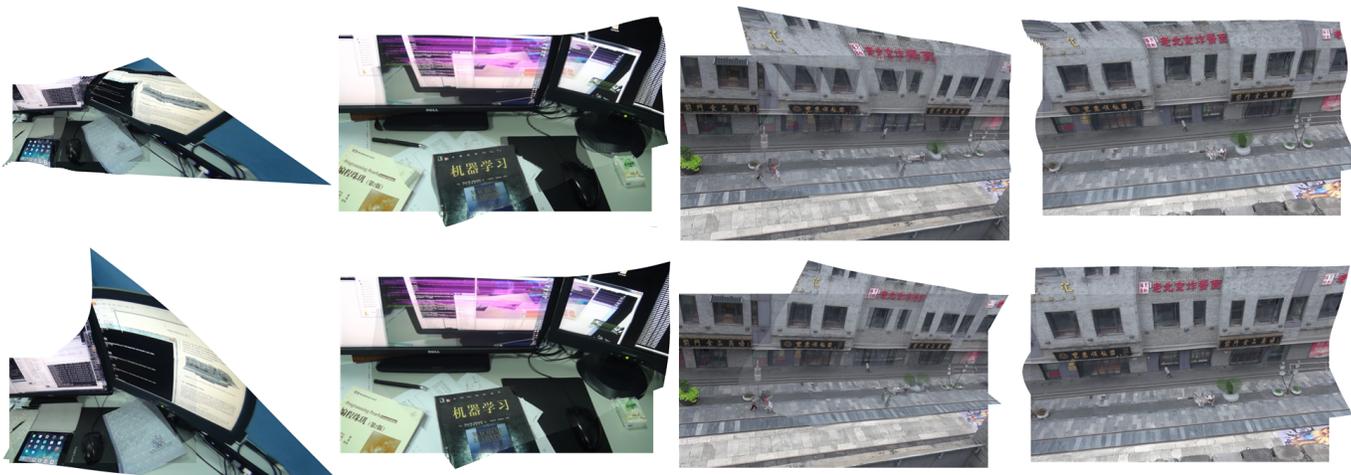


Fig. 1: Image pair alignment. Top: the stitched image pairs by APAP. Bottom: the stitched image pairs by our method.

III. MORE RESULTS

Figure 2 shows the stitched panoramas by Autostitch and APAP [1] for the example shown in Figure 1 in our paper. Besides panoramic mosaics, our approach can also be applied to texture unfolding for simple objects. Figure 3 shows an example where the desktop globe is unfolded to a world map. We captured images along the equator for every 24 longitudes, and stitched them with the loop closure term. Due to the spherical nature of the globe and the limited field of view, polar regions are not captured. It was almost impossible to find correspondences beyond 60° latitude due to large distortion. Autostitch result is shown in Figure 3(b) while our stitching result contains smaller misalignment even at the 45° latitude. APAP fails for this example due to significant accumulated errors and large distortion.

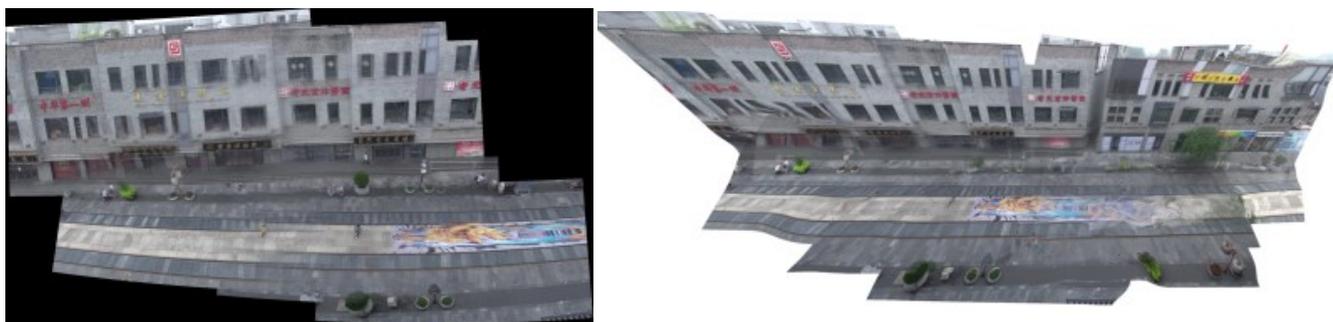


Fig. 2: Reconstructed urban panoramas by Autostitch (left) and APAP (right).

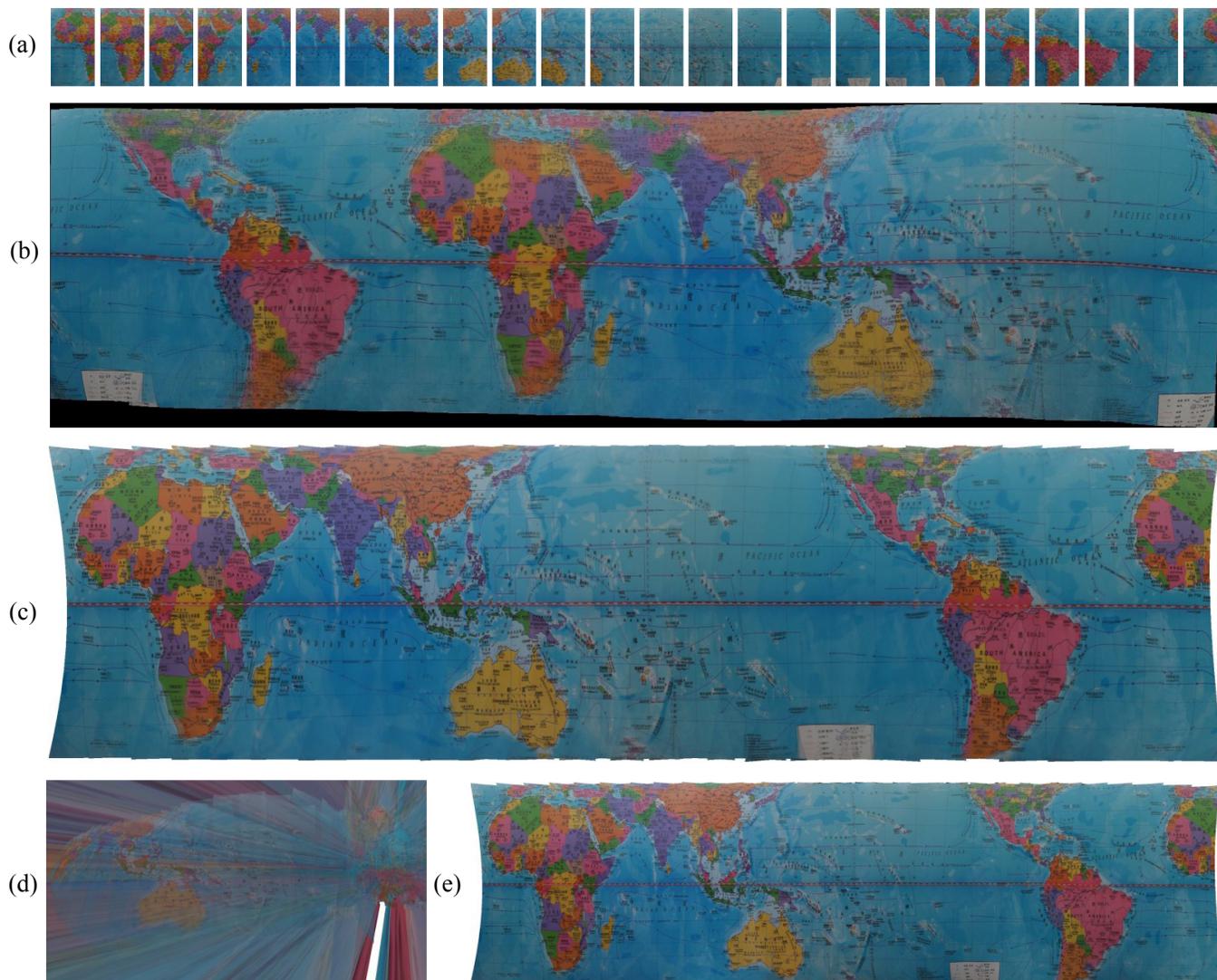


Fig. 3: A world map unfolded from desktop globe. (a) Input images. (b) The result generated by Autostitch. (c) The average of the stitched images by our approach. (d) The average of the stitched images by APAP. (e) Our final result with seamless composition.

REFERENCES

- [1] J. Zaragoza, T. Chin, Q. Tran, M. S. Brown, and D. Suter, "As-projective-as-possible image stitching with moving DLT," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 7, pp. 1285–1298, 2014.
- [2] W.-Y. Lin, L. Liu, Y. Matsushita, K.-L. Low, and S. Liu, "Aligning images in the wild," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2012.