Fronto-parallel building facades stitching

Wuxia Yan✉, Sheng Wang, Jun Hu and Chuancai Liu

Parallax handling is a challenging task for image stitching. Existing image stitching methods are quite sensitive to large parallax. One common phenomenon, e.g. is the ghosting. In this Letter, a flexible building facade stitching method that allows large parallax is proposed. Using the rigidity constraints of the man-made environment, each view can be aligned to a reference view. Excellent stitching performance demonstrates that the proposed method is superior to the state-of-the-art ones.

Introduction: Most of the stitching methods can only perform well on images with little or no parallax, which is a big limitation in practical situations. For example, the traditional projective stitching methods [1, 2] tend to warp one image to the other. However, homography cannot account for parallax. In order to solve the problem of image stitching under large parallax, more advanced methods [3, 4] focus on local warping or post-processing optimisation to address ghosting or distortion caused by parallax. As shown in Fig. 1*a*, the representative method shape-preserving half-projective warp (SPHP) combined a local projective transformation on the overlapping region and a local similarity transformation on the non-overlapping region to alleviate the distortion caused by large parallax. Both the traditional projective and advanced image stitching methods share the same projective transformation framework, which results in distortion of projective area and shape under large parallax.

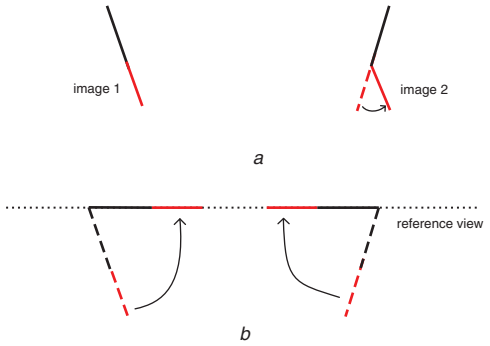

Fig. 1 *Schematic diagram represents of images processing in 3D world coordinate system, black part means non-overlapping region of image and red part is overlapping region of image*

a SPHP method
b Proposed method

On the other way, inspired by the view transformation of image [5–8], we proposed a novel method for fronto-parallel building facades stitching in this Letter. We concentrated on the close-range building image stitching. These images are taken by a hand-held camera so that large parallax is inevitable. To obtain convincing panoramas, we intended on transforming the views of the image pair to the reference view (termed as RV, i.e. the fronto-parallel building facades) before stitching them (see Fig. 1*b*). In this way, the parallax problem in stitching building facade images is solved.

Vanishing points detection and camera intrinsic matrix acquisition: The camera is generally modelled by the usual pinhole. Therefore, the relationship between a 3D point $P_i = [X_i, Y_i, Z_i, 1]^{\mathrm{T}}$ and its image projection $p_i = [x_i, y_i, 1]^{\mathrm{T}}$ is given by

$$\alpha p_i = \boldsymbol{C}[\boldsymbol{R}\ \ \boldsymbol{t}]P_i, \quad \text{with } \boldsymbol{C} = \begin{bmatrix} f_x & \gamma & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \tag{1}$$

where α is an arbitrary scale factor, $\boldsymbol{C}$ is 3×3 intrinsic matrix of the camera, $\boldsymbol{R}$ is 3×3 rotation matrix, $\boldsymbol{t}$ is 3×1 translation matrix, (f_x, f_y) represents focal length, γ is the skew parameter of the focal length and (c_x, c_y) is the pixel coordinate of the principal point.

Ideally, under the assumption of perpendicularity between orientations of the three main vanishing points, sets of parallel lines detected on building facade will intersect on vanishing points at infinity beyond the image boundary [8]. This principle can be used for the detection of vanishing point. In our method, we first detect the initial edges by Canny edge detection and further extracted strong line segments by Hough transformation. Then, under the assumption of Gaussian measurement noise, the vanishing point can be computed from the single intersection point of a set of parallel line segments by the maximum likelihood estimate [9]. Thus, the two points with the highest number of supporting lines on the horizontal and vertical orientation are considered to be the two main vanishing points P_{h} and P_{v}, respectively.

Fig. 2 *Left-hand side column shows original building image pair and right-hand side column represents corresponding fronto-parallel facades*

Table 1: Comparison of RMSE

	[1](L)	[1](R)	[2]	our method
TR	12.76	13.49	13.82	12.26
TE	14.37	14.97	13.95	13.16
Outliers	23.61%	23.73	22.89%	21.04%

TR is training set error and TE is testing set error. [1](L) and [1](R) represent the reference image are the left-hand side and the right-hand side input image, respectively.

Consequently, under the above perpendicularity assumption, we can obtain the last vanishing point P_l using the detection of the two vanishing points P_{h} and P_{v} as follow:

$$P_l = P_{\mathrm{h}} \times P_{\mathrm{v}} \tag{2}$$

Under the condition that the three main vanishing points $P_{\mathrm{h}} = [x_{\mathrm{h}}, y_{\mathrm{h}}]^{\mathrm{T}}$, $P_{\mathrm{v}} = [x_{\mathrm{v}}, y_{\mathrm{v}}]^{\mathrm{T}}$, $P_l = [x_l, y_l]^{\mathrm{T}}$ at infinity, (1) can be rewritten as:

$$\alpha \begin{bmatrix} x_{\mathrm{h}} & x_{\mathrm{v}} & x_l \\ y_{\mathrm{h}} & y_{\mathrm{v}} & y_l \\ 1 & 1 & 1 \end{bmatrix} = \boldsymbol{C}[\boldsymbol{R}\ \ \boldsymbol{t}] \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \tag{3}$$

Then, (3) can be rewritten as (4) under the standard assumptions, including square pixels (i.e. $f_x = f_y$), zero skew (i.e. $\gamma = 0$) and the principle point is located in the image centre (i.e. (c_x, c_y) is known).

$$\begin{bmatrix} \alpha x_{\mathrm{h}} & \alpha x_{\mathrm{v}} & \alpha x_l \\ \alpha y_{\mathrm{h}} & \alpha y_{\mathrm{v}} & \alpha y_l \\ \alpha & \alpha & \alpha \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_x & c_y \\ 0 & 0 & 1 \end{bmatrix} R \tag{4}$$

thus

$$R = \begin{bmatrix} \dfrac{\alpha(x_{\mathrm{h}} - c_x)}{f_x} & \dfrac{\alpha(x_{\mathrm{v}} - c_x)}{f_x} & \dfrac{\alpha(x_l - c_x)}{f_x} \\ \dfrac{\alpha(y_{\mathrm{h}} - c_y)}{f_x} & \dfrac{\alpha(y_{\mathrm{v}} - c_y)}{f_x} & \dfrac{\alpha(y_l - c_y)}{f_x} \\ \alpha & \alpha & \alpha \end{bmatrix} \tag{5}$$

Let $P_0 = [c_x, c_y]^{\mathrm{T}}$, and then using the perpendicularity of the first two columns of $\boldsymbol{R}$ gives:

$$\alpha^2 \left(\frac{(P_{\mathrm{h}} - P_0)(P_{\mathrm{v}} - P_0)}{f_x^2} + 1 \right) = 0 \tag{6}$$

Since $\alpha \neq 0$, f_x is able to be computed through (6). Then we can recover the intrinsic matrix $\boldsymbol{C}$.

Fig. 3 *First and second images are results of traditional projective stitching method with different reference images. Third image is result of SPHP warp and last one is panorama of our method*

Fronto-parallel building facade acquisition: It is known that the vanishing points are affected by the camera rotation rather than the camera position. Given vanishing points P_h and P_v, we can construct a transformation matrix $\hat{\boldsymbol{R}}$ which is utilised to transform the directions of P_h and P_v into RV, as follow:

$$\hat{\boldsymbol{R}} = \begin{bmatrix} \dfrac{(C^{-1}P_h)}{\|(C^{-1}P_h)\|} \\ \dfrac{(C^{-1}P_v)}{\|(C^{-1}P_v)\|} \\ \dfrac{(C^{-1}(P_h \times P_v))}{\|(C^{-1}(P_h \times P_v))\|} \end{bmatrix} \tag{7}$$

Then we can obtain the homography H as follow:

$$H = C\hat{R}C^{-1} \tag{8}$$

Without loss of generality, for any point m_i on one building image and its corresponding point M_i on the other building image, the perspective planar transformation between the two images with overlapping region is defined as follow:

$$\alpha M_i = Hm_i \tag{9}$$

Accordingly, from (9), the fronto-parallel building facades can be obtained.

The remaining steps of fronto-parallel building facade stitching share a usual stitching pipeline. First, SIFT feature points are extracted and matched on the overlapping regions. Then, mismatches are removed by random sample consensus. The remaining inliers are used to estimate the transformation function. Finally, the image pair is aligned and composited using the transformation function to get the final panorama.

Experimental results: The experiments were performed on a PC with 2GB of RAM and a 2.93 GHz Intel CPU.

In our experiment, the input images and the corresponding rectified images are shown in Fig. 2. In order to align the x-axis to the horizontal direction, the rotation angles of the two images are 19.0569° and 47.0011° around the y-axis, respectively.

To quantify the accuracy of a transform f, we computed the root mean squared error (RMSE) on a set of corresponding points $\{m_i, M_i\}_{i=1}^N$, i.e. $\mathrm{RMSE}(f) = \sqrt{(1/N)\sum_{i=1}^N \|f(m_i) - M_i\|^2}$. Further, we obtained the SIFT feature matches from the rectified image pair, and randomly partitioned them into a training set and a testing set, respectively. The training set was used to learn a warp, and the RMSE was evaluated over both sets. Table 1 depicts the results. It is obvious that the proposed method provides the lowest errors.

Fig. 3 demonstrates panoramas generated using different image stitching methods from a pair of building images with large parallax. We can see that the traditional projective stitching result suffered from the heavy projective distortion on sizes, shapes and orientations. Meanwhile, the SPHP warp produced distortion on the transition region that from the overlapping region to non-overlapping region. This is because that these two methods are based on the projective transformation framework which cannot account for parallax. It is clearly that the proposed method is relatively stable in the presence of parallax.

Moreover, more experimental results are available at https://github.com/rukia007/Panorama/blob/master/all2.pdf.

Conclusion: In this Letter, the possibility to obtain the panoramas of close-range building images with large parallax was demonstrated. This method can work well without constraints of input images (e.g. the input images should be taken from the same view or the scene should be roughly planar). Experiment proves that this method offers a high degree of flexibility on large parallax for practical use.

Acknowledgments: This work has been supported by The Project of Ministry of Industry, Information Technology of PRC (grant no. E0310/1112/02-1)

Submitted: *29 April 2016* E-first: *5 August 2016*
doi: 10.1049/el.2016.1530
One or more of the Figures in this Letter are available in colour online.

Wuxia Yan, Jun Hu and Chuancai Liu (*School of Computer Science and Engineering, Nanjing, People's Republic of China*)

✉ E-mail: wxyan127@outlook.com

Sheng Wang (*Institute of Image Processing and Pattern Recognition, Kaifeng, People's Republic of China*)

References

1 Brown, M., and Lowe, D.G.: 'Automatic panoramic image stitching using invariant features', *Int. J. Comput. Vis.*, 2007, **74**, (1), pp. 59–73
2 Song, T., Jeon, C., and Hanseok, K.: 'Image stitching using chaos-inspired dissimilarity measure', *Electron. Lett.*, 2015, **51**, (3), pp. 232–234
3 Chang, C.H., Sato, Y., and Chuang, Y.Y.: 'Shape-preserving half-projective warps for image stitching'. Proc. of IEEE CVPR, Columbus, Ohio, 2014, vol. 6, pp. 3254–3261
4 Zhang, F., and Liu, F.: 'Parallax-tolerant image stitching'. Proc. of IEEE CVPR, Columbus, Ohio, 2014, vol. 6, pp. 3262–3269
5 Wolberg, G.: 'Image morphing: a survey', *Vis. Comput.*, 1998, **8**, pp. 360–372
6 Seitz, S.M., and Dyer, C.R.: 'View morphing'. Proc. of Computer Graphics, Los Angeles, 1999, vol. 8, pp. 21–30
7 Rother, C.: 'A new approach to vanishing point detection in architectural environments', *Image Vis. Comput.*, 2002, **20**, pp. 647–655
8 Hartley, R.I.: 'Theory and practice of projective rectification', *Int. J. Comput. Vis.*, 1999, **35**, pp. 1–16
9 Hartley, R., and Zisserman, A.: 'Multiple view geometry in computer vision' (Cambridge university press, 2003)