WIDE-ANGLE IMAGE STITCHING USING MULTI-HOMOGRAPHY WARPING*

Bin Xu, Yunde Jia

Beijing Laboratory of Intelligent Information Technology
School of Computer Science, Beijing Institute of Technology, Beijing 100081, China
Email: {xubinak47, jiayunde}@bit.edu.cn

ABSTRACT

For wide-angle images with heavy lens distortion, especially those widely used in multi-camera systems, the single global homography cannot satisfy the required warp due to nonlinear distortions, leading to misalignment and shape distortion. This paper proposes a multi-homography warping to stitch wide-angle images with unknown lens distortion, which integrates multiple local homographies with a global homography for accurate alignment and shape preservation. We suggest a solution by conditional sampling to obtain a larger proportion of inliers for more accurate estimation of local and global projective transformations. We introduce an adaptive weighting scheme to combine these transformations for smoothing our warp over the entire target image from the local homographies to the global homography. The experiments evaluate the alignment accuracy and shape preservation of the proposed method.

Index Terms— Wide-Angle Image Stitching, Image Alignment, Multi-Homography Warping

1. INTRODUCTION

Existing stitching methods based on invariant feature matching techniques [1, 2] have already succeeded in aligning overlapping images taken by conventional lenses, and stitching them into seamless and natural photo-mosaics, but can not handle lens distortion well. We focus on dealing with wide-angle images suffering from lens distortion, e.g., radial distortion that seriously affects inlier selection, image alignment, and shape preservation. Panoramic stitching with lens distortion is required for many vision applications such as video surveillance [3, 4], navigation [5], and robotic vision systems [6, 7]. In this paper, we consider situations with unknown lens distortion, for example, images are captured by two cameras with significantly different heavy radial distortions as shown in Fig. 1(a), making it challenging to produce a highly aligned result with less shape distortion.

In order for stitching images with lens distortion, early methods use calibration models [9, 10] to correct the distortion effects, and recently polynomial methods [11, 12, 13] that estimate distortion factors based on Division Model [14] have been well studied. The Division Model is defined as $x' = x/(1 + \lambda||x||^2)$, where λ is the radial distortion coefficient. Jin *et al.* [11] use three image point correspondences to estimate homography accounting for radial distortion and solve it by using Levenberg-Marquardt method. Byröd *et al.* [12] perform homography estimation in the same way and solve the problem by using Gröner basis method. Because they both assume that all images share the same distortion factor, their models can not achieve accurate results when input images suffer from different lens distortions. Ju *et al.* [13] estimate the distortion factors for each image and create an accurate stitched image, however all images are warped according to the corresponding distortion factors, resulting in heavy shape distortion in the non-overlapping region. All the above methods work at the RANSAC [15] stage, also Byröd *et al.* [12] note the decreased performance of distortion free methods in inlier selection, since they conduct random sampling and removes matching feature points close to distorted

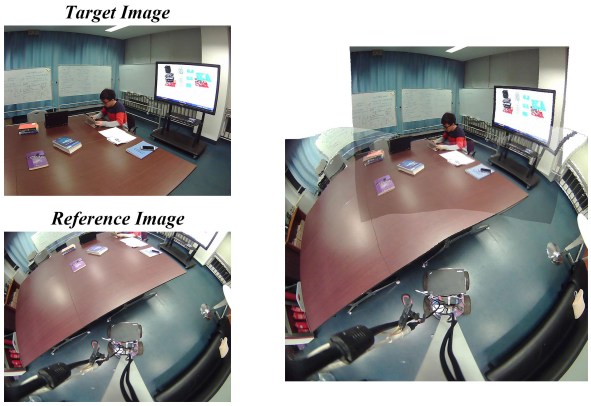


Fig. 1. Illustration of the proposed stitching method for a challenging case. (a) Images captured by our telepresence robot [8] with different heavy radial distortions. (b) Our result with accurate alignment in the overlapping region and shape preservation in the non-overlapping region.

* This research was supported in part by the Natural Science Foundation of China (NSFC) under Grant No.61375044.

edges as outliers.

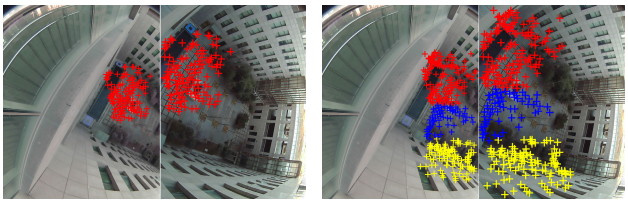
Although much attention has been given to distortion estimation, it is often overlooked that lens distortion brings about the inaccuracy of projective warp parametrised by a 3×3 homography matrix. The single homography projective model is usually too general to handle lens distortion, which might give rise to misalignments and ghosting effects in the overlapping region. Recently, methods based on local warping achieve outstanding performances on aligning images with no distortion. Gao *et al.* [16] propose a dual-homography warping which separates a single scene into a distant plane and a ground plane and estimates two homographies to produce a more seamless image. Moving Direct Linear Transformation (Moving DLT) technique proposed by Zaragoza *et al.* [17] can estimate varying local homographies to allow local deviations, and Lin *et al.* [18] combine Moving DLT with a six-degree-of-freedom similarity transform to create a natural-looking panorama. The similarity transform is also employed in the shape-preserving half-projective (SPHP) warp [19] that smoothly changes from projective to similarity along a rotated co-ordinate axis. The accuracy of local warping methods highly relies on enough inliers, which are especially hard to be chosen from distorted image pairs, furthermore, similarity transform is sufficient only in the presence of no or mild lens distortion.

We propose a multi-homography warping to bring overlapping images with unknown lens distortion into better alignment. Instead of estimating distortion factors, we first use residual sorting information to obtain a larger proportion of inliers, then generate a global projective transform for global shape preservation, and estimate multiple local projective transforms for accurate alignment. Finally the global homography and multiple local homographies are combined according to our weighting scheme to smooth the warp over the entire target image. Our method contributes to both accurate alignment in the overlapping region and shape preservation in the non-overlapping region, as presented in Fig. 1(b).

2. INLIER SELECTION

The goal of inlier selection is to obtain a noise-free set of matching feature points between the target image I and the reference image I' for homography estimation. RANSAC generates hypotheses of a given geometric model, e.g., the homography estimated from four point correspondences in our case, and each hypothesis is satisfied by a randomly chosen minimal subset of the input data. Most existing methods rank the homography hypotheses, and typically select inliers from all points according to their corresponding residuals to the optimal hypothesis.

In situations with large lens distortion, the non-linear effects at the RANSAC stage are notable but often ignored. The input data of corresponding points between wide-angle images are usually seriously contaminated with outliers. Even



(a) (b)

Fig. 2. (a) Inliers selected by 500 random sampling iterations. (b) Inliers selected from three portions by our method (10 random and 100 conditional sampling iterations for each portion) with the same outlier threshold.

points close to image borders which are actually more important to the quality of the final result, are removed as outliers because of their large residuals. In Fig. 2(a) we notice that the random sampling is only capable of finding matching points in the central, less distorted portion of an image pair, in fact, the true inliers spread out over the images. Certainly more points can be retained by setting a higher outlier threshold, but it increases the probability of accepting mismatching ones.

Intuitively, points close to the same portion of an image pair are more likely to satisfy the same homography hypothesis that better respects the local distorted structure. That is, given a minimal subset of points chosen from the same portion, we can generate the local homography hypothesis to recover as many inliers as possible in the portion. Based on Multi-GS [20], we use residual sorting information to perform a conditional sampling instead of pure random sampling, because the probability $f(P_i, P_j)$ of two points arising from the same homography can be encoded by residual sorting [20].

As an example, a fixed number of homography hypotheses $\{H_i\}_{i=1}^{M}$ are generated under random sampling from the input data $\{p_i, p_i'\}_{i=1}^{N}$, for each datum $P_i = \{p_i, p_i'\}$ we rank M hypotheses in terms of the residuals to them and form the permutation $h^i = \{h_1, ..., h_M\}$. The higher H_* is ranked in h^i, the more likely P_i is an inlier to it. Moreover, there will be many common hypotheses shared by P_i and P_j at the top of their list h^i and h^j if they are inliers from the same portion. Multi-GS constructs the conditional inlier probability distribution to guide the selection of the next datum according to the residual sorting information of datums that have been chosen. The probability is defined as

$$f(P_i, P_j) = \frac{1}{a} |h^i_{1:a} \bigcap h^j_{1:a}|, \tag{1}$$

where $h^i_{1:a}$ is the vector with the first-a hypotheses of h^i and $|h^i_{1:a} \bigcap h^j_{1:a}|$ denotes the number of shared hypothesis.

Instead of estimating distortion factors by random sampling, we choose conditional sampling to accelerate the generation of the local homography hypothesis, hence can recover as many local inliers as possible. By repeating the procedure and removing inliers obtained previously from the input

Algorithm 1 Multi-Homography Inlier Selection

Input: the input data P, size of a minimal subset $p > 0$, outlier threshold ε_o and ε_i ($\varepsilon_o > \varepsilon_i$), required number of homography hypotheses η_r and η ($\eta_r < \eta$).

Output: a set Θ of inliers.

remove the mismatches from P using RANSAC with ε_o;
repeat
  generate a set Π of η_r hypotheses from P by random sampling;
  repeat
    purely randomly select the first datum s_1 from P;
    select $p-1$ datums based on Π using Multi-GS;
    $\Pi = \Pi \cup \{\text{new hypothesis fitted on } s_{1:p}\}$;
  until (size of $\Pi > \eta$)
  find the optimal hypothesis with the largest number of inliers I whose residuals $< \varepsilon_i$;
  $\Theta = \Theta \cup I$; remove I from P;
until (size of $P < p$)
return Θ.

data before each loop, we are able to obtain a larger proportion of inliers uniformly distributed in most distorted portions. We summary our adaptive method as Algorithm 1 and show an example of selection results in Fig. 2(b).

3. GLOBAL HOMOGRAPHY

Given a set of inliers $\{p_i, p_i'\}_{i=1}^n$ between the target image I and the reference image I', the global homography $H_g \in \mathbb{R}^{3\times3}$ indicates the projective transform $p' = H_g p$,

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}. \tag{2}$$

Using Direct Linear Transformation (DLT) [21], we can rewrite (2) by cross product as $0_{3\times1} = p' \times H_g p$, linearized as

$$0_{3\times1} = ah = \begin{bmatrix} 0_{1\times3} & -p^T & y'p^T \\ p^T & 0_{1\times3} & -x'p^T \\ -y'p^T & x'p^T & 0_{1\times3} \end{bmatrix} \begin{bmatrix} h_1 \\ \vdots \\ h_9 \end{bmatrix}, \tag{3}$$

where two rows of $a \in \mathbb{R}^{3\times9}$ are linearly independent. We estimate h from all n inliers as

$$\hat{h} = \arg\min_h \sum_{i=1}^{n} ||a_i h||^2 = \arg\min_h ||Ah||^2, \tag{4}$$

with the constraint $||h|| = 1$, where a_i corresponds to the first-two rows of a for the i-th datum $\{p_i, p_i'\}$. The global homography H_g reshaped from $\hat{h}$ can be obtained by calculating the least significant right singular vector of $A \in \mathbb{R}^{2N\times9}$.

As a result of the larger proportion of inliers obtained by using our inlier selection method, the global transform p-reserves the global shape of I extremely well as shown in Fig. 3(a), yet fails to align images in the overlapping region.

4. MULTI-HOMOGRAPHY WARPING

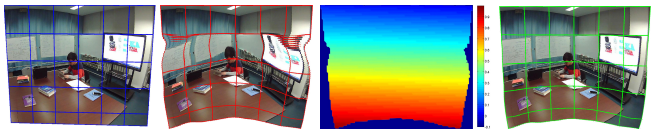
To improve alignment accuracy, we use Moving DLT [17] technique to estimate a local homography

$$h^* = \arg\min_h \sum_{i=1}^{N} ||w_i^* a_i h||^2 = \arg\min_h ||W^* A h||^2 \tag{5}$$

for each position x^* in I, where the weight is calculated as

$$w_i^* = max(exp(-||x^* - p_i||^2/\sigma^2), \gamma), \tag{6}$$

so $W^* = diag([w_1^* w_1^* ... w_N^* w_N^*])$, and $\gamma \in [0, 1]$ is an offset. Moving DLT assumes higher weights to inliers closer to x^*, the homography h^* better respects the local distorted structure around x^*, so its accuracy highly depends on the quality and the quantity of inliers. As inliers obtained by using our method spread out over the overlapping region, the warp h^* varies smoothly and provides the most accurate alignment. While the h^* in the non-overlapping region containing no inliers, becomes a Gaussian-weighted combination of local homographies of the overlapping region, leading to unnatural shape distortion as shown in Fig. 3(b).



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

Fig. 3. (a) The warped I by the global homography. (b) The warped I by the local homographies. (c) Weight map of w_l^*. (d) The result of our method.

In our method, we integrate the local h^* with global H_g as our *multi-homography warping*:

$$H_m^* = w_l^* h^* + (1 - w_l^*) H_g, \tag{7}$$

the weighting coefficient is given by

$$w_l^* = (u^* - u_m)/(u_M - u_m), \tag{8}$$

where u is the new coordinate system (u, v) obtained as a rotation of the original (x, y) of the warped I. The rotation angle is chosen as $\theta = atan2(-h_8, -h_7)$ (h_8, h_7 from H_g), and it has been proved in [19] that the area distortion of H_g only becomes larger along the positive u-axis, from the overlapping region to the non-overlapping region. Therefore we smooth the warp from the local homographies to the global

homography along the positive u-axis. u_m and u_M are smallest and largest value of u of all x^* in I, respectively. For each x^*, the w_l^* ranging from 0 to 1 is computed by Eq. (8). Fig. 3(c-d) present a weight map of w_l^* and the result of our multi-homography warping. To compensate the adjustments of local homographies in the overlapping region, we warp the reference image I' by using $R_m^* = H_m^*(h^*)^{-1}$.

5. EXPERIMENTS

We compared our multi-homography inlier selection with methods proposed by Byröd *et al.* [12] and Ju *et al.* [13]. The distorted image dataset is also from [13] with pre-defined distortion factors ranging from -0.5 to 0.5. The feature points were detected using SURF [22]. On each of 11 image pairs, the average inlier proportion of each method over 10 repetitions of 2000 random sampling iterations was computed. For our method, we ran 10 random and 100 conditional sampling iterations in each loop and the total number of iterations is also limited to 2000 in each repetition.

The results are depicted in Fig. 4. The method [13] has similar inlier proportions unrelated to distortion factors but at a lower rate, and the performance of the method Stitching3pt [12] significantly decreases when $\lambda < 0$, since their method can not estimate accurate distortion factors in this case. It is clear that our multi-homography inlier selection consistently outperforms the others, as we generate local homography hypotheses under conditional sampling instead of estimating distortion factors using random sampling.

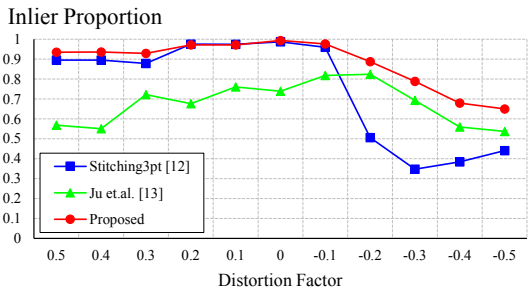
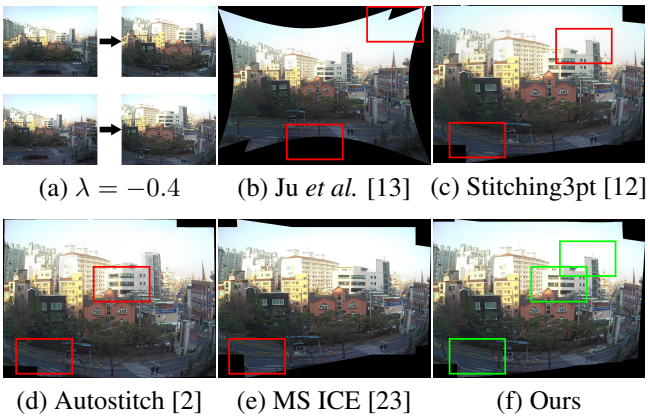


Fig. 4. Comparison of the inlier proportions.

To evaluate the alignment accuracy and shape preservation of our multi-homography warping, we compare our method with methods [12, 13], softwares Autostitch [2], and Microsoft ICE [23], and all of these techniques employ a single homography for image alignment. Fig. 5(b-f) show the stitched results of a pair of distorted image with $\lambda = -0.4$ (Fig. 5(a)), and we highlight a few notable areas where misalignments and shape distortion occur. Because all images are warped according to the corresponding distortion factor, heavy shape distortion are produced by method [13]. The results of Stitching3pt, Autostitch, and Microsoft ICE are similar, and all of these methods produce misalignments and



(a) $\lambda = -0.4$ (b) Ju *et al.* [13] (c) Stitching3pt [12]

(d) Autostitch [2] (e) MS ICE [23] (f) Ours

Fig. 5. (a) Two distorted images with $\lambda = -0.4$. (b-f) Stitched results of (a) via different methods and ours.

Fig. 6. Panorama of building images.

ghosting effects due to the inaccuracy of the single homography projective warp. However, our method aiming for local alignment accuracy and global shape preservation performs better in these areas.

The panorama example in Fig. 6 demonstrates that our method is easily generalized to multiple images taken by wide angle lenses with unknown lens distortion.

6. CONCLUSION

This paper has presented a multi-homography warping to stitch wide-angle images with unknown lens distortion, which integrates multiple local homographies with a global homography. The experiments have shown that our method improves the performance of inlier selection on wide-angle image pairs, and can achieve both accurate alignment in the overlapping region and shape preservation in the non-overlapping region.

7. REFERENCES

[1] Matthew Brown and David G Lowe, "Recognising panoramas.," in *IEEE International Conference on Computer Vision (ICCV)*. IEEE, 2003, vol. 3, p. 1218.

[2] Matthew Brown and David G. Lowe, "Automatic panoramic image stitching using invariant features," *International Journal of Computer Vision (IJCV)*, vol. 74, no. 1, pp. 59–73, 2007.

[3] Yael Pritch, Alex Rav-Acha, and Shmuel Peleg, "Nonchronological video synopsis and indexing," *IEEE Transactions on Pattern Analysis & Machine Intelligence (TPAMI)*, vol. 30, no. 11, pp. 1971–1984, 2008.

[4] Xiangyun Meng, Wei Wang, and Ben Leong, "Skystitch: A cooperative multi-uav-based real-time video surveillance system with stitching," in *ACM International Conference on Multimedia*, 2015, pp. 261–270.

[5] Hong Cheng, Zicheng Liu, Nanning Zheng, and Jie Yang, "Enhancing a driver's situation awareness using a global view map," in *International Conference on Multimedia and Expo (ICME)*. IEEE, 2007, pp. 1019–1022.

[6] Steven Johnson, Irene Rae, Bilge Mutlu, and Leila Takayama, "Can you see me now?: How field of view affects collaboration in robotic telepresence," in *The 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 2015, pp. 2397–2406.

[7] Jiajun Shen, Bin Xu, Mingtao Pei, and Yunde Jia, "A low-cost tele-presence wheelchair system," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2016, pp. 2452–2457.

[8] Jia Yunde, Xu Bin, Shen Jiajun, and Pei Mintao, "Telepresence interaction by touching live video images," *arXiv preprint arXiv:1512.04334v2*, 2016.

[9] J. Weng, P. Cohen, and M. Herniou, "Camera calibration with distortion models and accuracy evaluation," *IEEE Transactions on Pattern Analysis & Machine Intelligence (TPAMI)*, vol. 14, no. 10, pp. 965–980, 1992.

[10] Zhengyou Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis & Machine Intelligence (TPAMI)*, vol. 22, no. 11, pp. 1330–1334, 2000.

[11] Hailin Jin, "A three-point minimal solution for panoramic stitching with lens distortion," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2008, pp. 1–8.

[12] Martin Byröd, Matthew Brown, and Karl Åström, "Minimal solutions for panoramic stitching with radial distortion," in *The 20th British Machine Vision Conference (BMVC)*, 2009.

[13] Myung-Ho Ju and Hang-Bong Kang, "Panoramic image generation with lens distortions," in *International Conference on Image Processing*. IEEE, 2013, pp. 1296–1300.

[14] Andrew W Fitzgibbon, "Simultaneous linear estimation of multiple view geometry and lens distortion," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2001, vol. 1.

[15] Martin A. Fischler and Robert C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the Acm*, vol. 24, no. 6, pp. 381–395, 1980.

[16] Junhong Gao, Seon Joo Kim, and M. S. Brown, "Constructing image panoramas using dual-homography warping," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011, pp. 49–56.

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

[18] Chung-Ching Lin, Sharathchandra U Pankanti, Karthikeyan Natesan Ramamurthy, and Aleksandr Y Aravkin, "Adaptive as-natural-as-possible image stitching," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1155–1163.

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

[20] Tat Jun Chin, Jin Yu, and David Suter, "Accelerated hypothesis generation for multistructure data via preference analysis," *IEEE Transactions on Pattern Analysis & Machine Intelligence (TPAMI)*, vol. 34, no. 4, pp. 625–638, 2012.

[21] Richard Hartley and Andrew Zisserman, *Multiple view geometry in computer vision*, Cambridge university press, 2003.

[22] Herbert Bay and Tinne Tuytelaars, "Surf: Speeded up robust features," in *European conference on computer vision (ECCV)*. Springer, 2006, pp. 404–417.

[23] Microsoft Research, "Microsoft image composite editor," version 2.0.3, 2016.