AN AUTOMATIC IMAGE REGISTRATION EVALUATION MODEL ON DENSE FEATURE POINTS BY PINHOLE CAMERA SIMULATION

Wen-Liang Du, Student Menber, IEEE and Xiao-Lin Tian

Faculty of Information Technology, Macau University of Science and Technology, Macao SAR, China

ABSTRACT

Evaluating image registration methods in images with dense feature points is challenging, because it's hard to discriminate real inliers in thousands of resulting correspondences processed by image registration methods. Therefore, this paper presents a dense feature points simulation which could provide ground truth of correspondences for evaluating image registration methods automatically. Moreover, the dense feature points are created by simulating pinhole camera model, parallax of reference image and sensed image, radial distortion of camera lens and random outliers. The performance of five state-of-art image registration methods is evaluated and discussed on the dense correspondences simulated by proposed model. The evaluation results show that the proposed model indeed offers a practical way for evaluating image registration methods on dense feature points.

Index Terms— Automatic image registration evaluation, dense feature points, pinhole camera simulation.

1. INTRODUCTION

Image registration is the process of aligning two images—reference and sensed images geometrically. It is widely used in several image processing applications and pattern recognition areas including remote sensing, medical imaging, computer vision etc [1]. Besides, the accuracy of image registration profoundly affects the accuracy of applications which use it (such as fusion, 3D representation, mosaic etc) [1]. Therefore, image registration is a critical progress in the image analysis tasks where the final information is gained from combination of various data sources.

During the last decades, a large number of approaches have been proposed for image registration [2–7]. Among them, feature-based methods are most robust and could handle complex image distortions [1]. In particular, dense feature-based image registration is widely used in 3D representation [8] and stereo matching [9] for providing sub-pixel accuracy.

However, to evaluate image registration methods in images with dense feature points is still a challenging due to the following two critical problems:

1. The existing high-resolution datasets with ground truth are still scarce [10].
2. For quantitatively evaluating image registration methods on the datasets with no ground truth, the remaining outliers after registration have to be discriminated manually [2, 4–7]. But it is hard to manually distinguish them in thousands of or even more correspondence.

To tackle the above problems, we present a stimulation of dense feature points to provide ground truth of correspondence for evaluating image registration methods quantitatively and automatically.

The main advantages of this proposed model are concluded as follows:

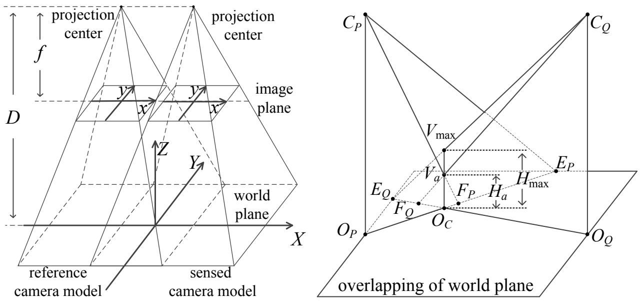
- The proposed model can provide as many datasets with ground truth as we need for designing and evaluating image registration methods.
- The parameters in the proposed model can be customized for simulating various situations.
- The procedure of providing ground truth is fully automatic.

The rest of this paper is organised as follows: the details of the simulation are introduced in section 2. Five state-of-art [3–5] feature-based image registration methods are evaluated on the proposed simulation and the results are discussed in section 3. Finally, the paper is concluded in section 4 and the future work is given in section 5.

2. SIMULATION

The feature points of reference and sensed images in the proposed simulation are generated through pinhole camera model shown in Fig. 1(a). To simplify the placement of objects in the camera model, we assume that the camera is placed overhead and shoot vertically to the objects placed on a flat. In this paper, the percentage of overlapping between reference and sensed camera models is simply set as 50% and the projection centers of two camera models share the same y-coordinate and z-coordinate. Moreover, the depth of filed of

This work is supported by the Science and Technology Development Fund of Macao (No. 99/2016/A3)



(a) Reference and sensed pinhole camera models

(b) An example of calculating $H_{\max}$ and $R_{\min}$ for a certain cone.

Fig. 1. Pinhole camera model.

two camera models is assumed as from focal length to camera distance, which means the features of objects in the world plane are all captured in focus.

Then, the proposed simulation is started with feature points creation of a flat in world plane and the feature points in reference and sensed images are denoted as p and q respectively. Next, the coordinates of p and q are changed by parallax and image distortion simulation. Finally, the outliers are selected randomly between p and q.

2.1. Feature Points Initialization

We firstly set the length-width ratio of the image plane as 36 to 24 which is the same length-width ratio of a "full-frame" image plane [11]. Then, N feature points captured by a flat in world plane, of reference image for example, are generated with the equal distance to their eight adjacent feature points, which is like pixels aligned in image plane.

2.2. Parallax simulation

The corresponding extrinsic parameters of p and q should be calculated as pinhole camera geometry model firstly. Then, their elevation are adjust as the height/depth of some random right circular cones produced on the overlapping world plane. Therefore, the parallax is simulated by the coordinates updating of p and q though the elevation adjustment. The amount, height and radius of random cones are discussed as follows.

Assume that, there are C cones created in the overlapping world plane, and the horizontal coordinates of their vertices follow uniform distribution. In order to ensure there is no feature points sheltered by others while elevation adjustment, the height/depth and radius of the cones are randomly selected lower than a maximum height ($H_{\max}$) and larger than a minimum radius ($R_{\min}$) respectively, where $H_{\max}$ and $R_{\min}$ are calculated as Algorithm 1 and illustrated in Fig. 1(b) (F_P and F_Q are the intersection points of line C_PV_a and line C_QV_a to the overlapping word plane). In addition, the orientations of the cones are also determined randomly, which means a certain cone could be a peak or a valley.

Algorithm 1 Determine the $H_{\max}$ and $R_{\min}$ for a Certain Cone

```
Input: $C_P$ and $C_Q$: projection centers of reference and sensed camera models respectively; $O_P$ and $O_Q$: world planes' centers of reference and sensed camera models respectively; $O_C$: a base's center of a certain cone; $D$: The camera distance of reference and sensed camera models; $E_P$ and $E_Q$: the end points of line $O_PO_C$ and line $O_QO_C$ to edge of overlapping world plane respectively; $V_{\max}$: a point which shares the horizontal coordinates with $O_C$, and the distance from it to $O_C$ is $H_{\max}$; $V_a$: the vertex of a certain cone, which randomly locates between $V_{\max}$ and $O_C$; $H_a$: the height of the line segment $V_aO_C$.
Output: $H_{\max}$ and $R_{\min}$
if $|E_PO_C| \leq |E_QO_C|$ then
    $H_{\max} = \frac{|E_PO_C|}{|O_CO_P|} \times D$;
else
    $H_{\max} = \frac{|E_QO_C|}{|O_CO_Q|} \times D$;
end if
Randomly select a point between $V_{\max}$ and $O_C$ as $V_a$;
$R_P = \left(|O_CO_P| \times \frac{H_a}{D}\right) / \left(1 - \frac{H_a}{D}\right)$;
$R_Q = \left(|O_CO_Q| \times \frac{H_a}{D}\right) / \left(1 - \frac{H_a}{D}\right)$;
$R_{\min} = \max\left(R_P, R_Q\right)$;
```

Furthermore, we denote (x_i, y_i) and $\left(x_i^{(c)}, y_i^{(c)}\right)$ as the ith original and ajusted coordinates of p respectively, and the center of p is assumed as $(0, 0)$. Then, $\left(x_i^{(c)}, y_i^{(c)}\right)$ can be calculated as pinhole camera geometry:

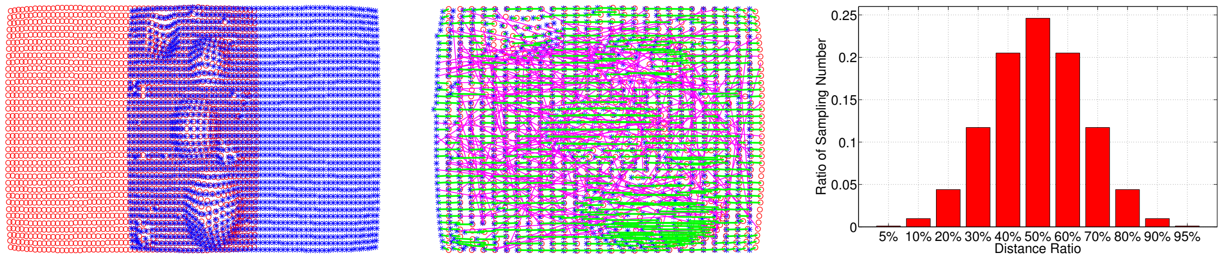
$$\begin{bmatrix} x_i^{(c)} \\ y_i^{(c)} \end{bmatrix} = \begin{bmatrix} x_i \left(\frac{D}{D-Z_i}\right) \\ y_i \left(\frac{D}{D-Z_i}\right) \end{bmatrix}, \tag{1}$$

where D is the camera distance and Z_i is the z-coordinate of the ith corresponding extrinsic parameter.

2.3. Image Distortion Simulation

In real cameras, image would be encountered non-linear aberrations due to complex lens systems [12], and the most visible aberration is radial distortion which makes the actual image points to be displaced radially in the image plane [13]. Consequently, in this paper, the radial distortion is introduced for simulating non-linear distortion. Note that, the image distortion simulation is processed after parallax simulation (See Fig. 2(a)). Hence, the distorted coordinates of p, $\left(x_i^{(d)}, y_i^{(d)}\right)$, is implemented as follows [13–16]:

$$\begin{bmatrix} x_i^{(d)} \\ y_i^{(d)} \end{bmatrix} = \begin{bmatrix} x_i^{(c)} \left(1 + k_1 r_i^2 + k_2 r_i^4 + ...\right) \\ y_i^{(c)} \left(1 + k_1 r_i^2 + k_2 r_i^4 + ...\right) \end{bmatrix}, \tag{2}$$



(a) Simulated feature points by 30 random cones and radial distortion

(b) Simulated feature points in overlapping of reference and sensed images with 20% random outliers

(c) Distribution of distances between outliers in reference and sensed images ($n = 10, p = 0.5$)

Fig. 2. Examples of simulated feature points of reference and sensed images with 50% overlapping and 20% random outliers, and the distribution of distances between corresponding outliers. (red cycles and blue stars denote feature points in reference and sensed images respectively, green and magenta lines denote correct and incorrect correspondences respectively)

where $k_1, k_2, ...$ are the coefficients of radial distortion, and $r_i = \sqrt{\left(x_i^{(d)}\right)^2 + \left(y_i^{(d)}\right)^2}$.

2.4. Outliers Selection

We denote p' and q' as the feature points in the overlapping image plane. In particular, they are supposed as one-to-one correspondence (inliers) initially, and the distances between the corresponding feature points are supposed as zero. In this paper, the outliers are assumed as additive white Gaussian noise of the inliers. Therefore, the outliers in p' are selected as uniform distribution, while the outliers in q' are selected as the distances to the corresponding outliers in p'. Generally, the distances could be generated follows binomial distribution (See Fig. 2(c)). Fig. 2(b) shows the overlapping feature points of Fig. 2(a) with 20% outliers.

3. EVALUATION

3.1. Execution Environments and Parameter Discussion

All the experiments are performed on a personal computer with 3.3-GHz Intel i5 CPU, 20-GB memory and MATLAB Code. The five state-of-art feature-based image registration methods are chosen as LLTA, LLTR, LLTV [3], WGTM [4] and SOCBV [5]. Our simualtion codes are available at **https://github.com/WenliangDu/CorrespondencesSimulation**.

The radial distortion is stimulated as the relative distortion curve of Zeiss Distagon 2.8/21 [12]. Hence, the focal length (f) used in this paper is 21 millimeter. The camera distance (D) is set as 2 meters. Furthermore, the numbers of cones (C) in the overlapping of world plane is set to 30, and their heights and radius are set as half of $H_{\max}$ and double of $R_{\min}$ respectively.

3.2. Evaluation Results

The performance of the five state-of-art image registration methods is evaluated on four generic criteria [4, 17]:

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP}, \tag{3}$$

$$Precision = \frac{TP}{TP + FP}, \tag{4}$$

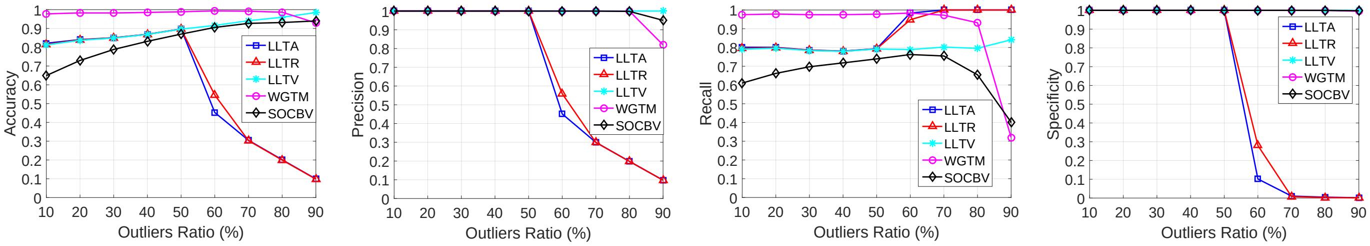
$$Recall = \frac{TP}{TP + FN}, \tag{5}$$

$$Specificity = \frac{TN}{TN + FP}, \tag{6}$$

where TP and FP denote the number of resulting inliers and outliers respectively, while FN and TN denote the number of missing inliers and outliers. Therefore, the $Accuracy$ describes the degree of right matches to all matches, the $Precision$ gives the degree of remaining inliers to all remaining corresponding points, the $Recall$ tells the degree of remaining inliers to all inliers, the $Specificity$ states the degree of discriminated outliers to all outliers.

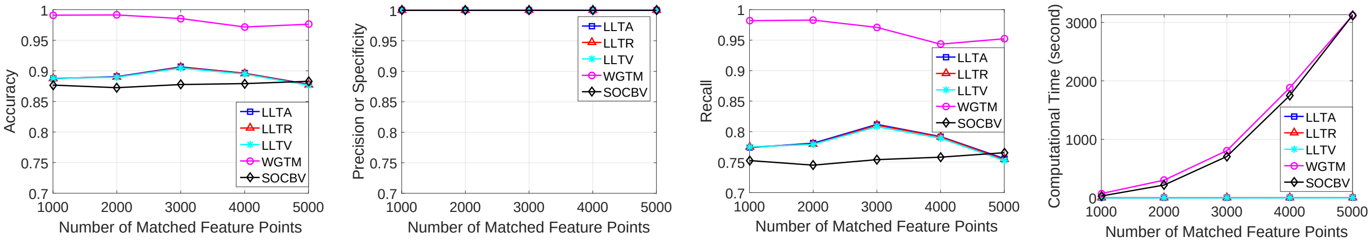
Note that, the five methods are evaluated for each experiments 50 times and the outliers, cones are added randomly in each experiment.

Fig. 3 shows the average *Accuracy*, *Precision*, *Recall* and *Specificity* of the five methods on 2000 pairs of simulated feature points in overlapping image plane, where the ratio of outliers is varied from 10% to 90% with an increment of 10%. There is an apparent tendency that the *Accuracy*, *Precision* and *Specificity* of LLTA and LLTR decrease dramatically when outliers ratio is over 50%. That is because LLTA and LLTR not only preserve all the inliers but also retain most outliers when the percentage of outliers is more than 50%. By contrast, LLTV provides reasonable *Accuracy*, *Precision*, *Recall* and *Precision* even the outliers ratio is 90%. WGTM gives highest *Accuracy*, *Precision*, *Recall* and *Precision*



(a) Accuracy (b) Precision (c) Recall (d) Specificity

Fig. 3. Performance comparison of LLTA, LLTR, LLTV, WGTM and SOCBV methods on 2000 pairs of simulated feature points and varies outliers ratio ($10\% \sim 90\%$).



(a) Accuracy (b) Precision and specificity (c) Recall (d) Computational time

Fig. 4. Performance comparison of LLTA, LLTR, LLTV, WGTM and SOCBV methods on varies pairs of simulated feature points ($1000 \sim 5000$) and 50% outliers.

among the five methods until the outliers ratio goes to 90%. SOCBV also discriminates all the outliers when the outliers ratio is lower than 90%, but the number of inliers that it preserved is relative lower than other methods preserved (except WGTM at 90% outliers ratio, see Fig. 3(c)).

Fig. 4 presents the average *Accuracy*, *Precision*, *Recall*, *Specificity* and *computational time* of the five methods on 1000 to 5000 pairs of simulated feature points in 50% outliers. In particular, since all the five methods remove every outliers in these cases, their *Precision* and *Specificity* are always one (see Fig. 4(b)). Moreover, WGTM preserves more inliers than others, hence it gives highest *Accuracy* and *Recall*. However, WGTM and SOCBV cost much more computational time than other three methods with the increasing number of feature points (see Fig.4(d)).

Therefore, according to the evaluation results of the five methods on our proposed dense feature points model, WGTM provides highest *Accuracy*, *Precision*, *Recall* and *Specificity* among other four methods when the percentage of outliers lower than 80%. But it costs near one hour for discriminating outliers in 5000 pairs of simulated feature points. However, LLTV gives lower than WGTM but still reasonable *Accuracy*, *Precision*, *Recall*, *Specificity* and spend only around 1 second for removing outliers in 5000 pairs of simulated feature points.

4. CONCLUSION

In this paper, we proposed an automatic image registration evaluation model by simulating dense feature points which is generated though modelling pinhole camera model, parallax of images, radial distortion of camera lens and random outliers. According to the evaluation results of five state-of-art methods testing on the proposed model, WGTM gives most precise registration when the ratio of outliers is lower than 80%, while LLTV is more satisfactory than WGTM for image registration with dense feature points. Therefore, the proposed model indeed be able to provide adequate and customized ground truth of correspondences for evaluating image registration methods.

5. FUTURE WORK

The proposed model could be more realistic by simulating perspective distortion, the more complicated objects and viewpoints.

6. ACKNOWLEDGMENT

The authors would like to thank Mohammad Izadi and Fanyang Meng for their generous help with WGTM and SOCBV algorithms.

7. REFERENCES

[1] B. Zitova and J. Flusser, "Image registration methods: a survey," *Image and Vision Computing*, vol. 21, no. 11, pp. 977–1000, 2003.

[2] W. Aguilar, Y. Frauel, F. Escolano, M. E. Martinez-Perez, A. Espinosa-Romero, and M. A. Lozano, "A robust graph transformation matching for non-rigid registration," *Image and Vision Computing*, vol. 27, no. 7, pp. 897–910, 2009.

[3] J. Ma, H. Zhou, J. Zhao, Y. Gao, J. Jiang, and J.W. Tian, "Robust feature matching for remote sensing image registration via locally linear transforming," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 12, pp. 6469–6481, 2015.

[4] M. Izadi and P. Saeedi, "Robust weighted graph transformation matching for rigid and nonrigid image registration," *IEEE Transactions on Image Processing*, vol. 21, no. 10, pp. 4369–4382, 2012.

[5] F. Meng, X. Li, and J. Pei, "A feature point matching based on spatial order constraints bilateral-neighbor vote," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 4160–4171, 2015.

[6] M. L. Uss, B. Vozel, V. V. Lukin, and K. Chehdi, "Multimodal remote sensing image registration with accuracy estimation at local and global scales," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 11, pp. 6587–6605, Nov 2016.

[7] H. Zhou, J. Ma, C. Yang, S. Sun, R. Liu, and J. Zhao, "Nonrigid feature matching for remote sensing images via probabilistic inference with global and local regularizations," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 3, pp. 374–378, March 2016.

[8] Peter Henry, Michael Krainin, Evan Herbst, Xiaofeng Ren, and Dieter Fox, "Rgb-d mapping: Using kinect-style depth cameras for dense 3d modeling of indoor environments," *The International Journal of Robotics Research*, vol. 31, no. 5, pp. 647–663, 2012.

[9] D. Scharstein, R. Szeliski, and R. Zabih, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," in *Proceedings IEEE Workshop on Stereo and Multi-Baseline Vision (SMBV 2001)*, 2001, pp. 131–140.

[10] Daniel Scharstein, Heiko Hirschmüller, York Kitajima, Greg Krathwohl, Nera Nešić, Xi Wang, and Porter Westling, *High-Resolution Stereo Datasets with Subpixel-Accurate Ground Truth*, pp. 31–42, Springer International Publishing, Cham, 2014.

[11] "Full-frame digital slr," `https://en.wikipedia.org/wiki/Full-frame_digital_SLR`.

[12] "Distortion," `http://toothwalker.org/optics/distortion.html`.

[13] Slama, *Manual of Photogrammetry*, American Society of Photogrammetry, 1980.

[14] J. Heikkila and O. Silven, "A four-step camera calibration procedure with implicit image correction," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Jun 1997, pp. 1106–1112.

[15] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330–1334, Nov 2000.

[16] J. Heikkila, "Geometric camera calibration using circular control points," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1066–1077, Oct 2000.

[17] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, pp. 1615–1630, Oct 2005.