ICIC Express Letters ICIC International ©2016 ISSN 1881-803X
Volume **10**, Number **11**, November **2016** pp. **2527–2532**

ROBUST QUASI-DENSE WIDE BASELINE MATCH PROPAGATION

Yiran Wang, Wei Wang*, Hai Zhu and Xiaowei Zhang

School of Network Engineering
Zhoukou Normal University
No. 6, Wenchang Road, Chuanhui District, Zhoukou 466001, P. R. China
*Corresponding author: wangwei@zknu.cn

Received April 2016; accepted July 2016

Abstract. *The paper proposes a robust three-view matching method to produce quasi-dense 3D points in multi-view stereo. In order to overcome the limitations of traditional wide baseline match propagation methods, the proposed method simultaneously models illumination variations and perspective distortions in images, and performs an effective affine parameter error detection and rectification mechanism in the process of match propagation. Experimental results show the proposed method can effectively enhance the stability and robustness of match propagation, and achieve better performance in both accuracy and quantity of the quasi-dense 3D points.*

Keywords: Match propagation, Multi-view stereo, 3D reconstruction, Affine model, Illumination variation

1. **Introduction.** Outdoor scene reconstruction from multiple images is one of hot pursued topics in computer vision community in recent years. The quasi-dense match propagation methods [1-4] are commonly used since the reconstructed 3D points are usually sufficient for describing scene structures or can be directly taken as initialization to enhance the reliability of dense matching methods. However, most existing quasi-dense methods [1] designed for small baseline image pairs are unsuitable for wide baseline ones, and some wide baseline matching methods [2-4] are not robust for significant illumination variations and perspective distortions in images obtained from outdoor scenes.

To address these issues, in this study, we improve the previous methods [2,3] by simultaneously modeling illumination variations and perspective distortions between the patches around the current seed in both images, and immediately optimize the corresponding affine parameters in the process of match propagation. Moreover, we also enhance the applicability of these methods by extracting more seeds by other feature detection methods (e.g., SIFT, Harris) other than Hessian-Affine detector.

The remainder of our paper is organized as follows. Section 2 describes the problem under consideration. Section 3 introduces the proposed method. Experimental results are demonstrated in Section 4. Section 5 concludes the paper.

2. **Problem Description.** The major advantage of the method [2] lies in that it can adaptively estimate the affine parameters of a new match during the propagation by computing the corresponding second order intensity moments. This often allows the seed to propagate into more regions with different local transformation.

However, in practice, for regions with significant illumination variations or perspective distortions in images, the affine parameters estimated by the second order intensity moment may be unreliable, and thus should be immediately optimized in order to enhance the reliability of the subsequent propagation and reduce the occurrence of outliers caused by the error accumulation of unreliable affine parameters.

To illustrate the method, we conduct experiments on two wide baseline images of the Herz-Jesu dataset [5]. Figure 1(a) shows that the initial eighth seed is propagated 1000

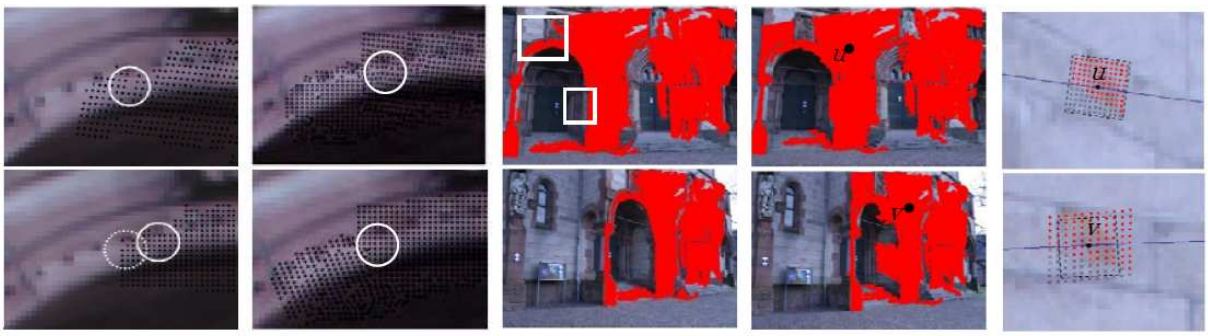

FIGURE 1. Affine transformation estimation: (a) outliers caused by erroneous affine parameters in method [2]; (b) more reliable matches obtained by our method; (c) the results obtained by the method [2]; (d) the results obtained by our method; (e) the illustration of the affine transformation estimation

times by the method [2]. Clearly, because of unreliable affine parameters updated in regions with significant illumination variations, the points in the solid circle of the current image match erroneously with the points in the solid circle of the matching image. The real points that match with them should be in the dotted circle of the matching image. In addition, as shown in Figure 1(c), the method fails in certain regions (e.g., rectangular regions) with large perspective distortion.

Note that, the method [3] extends the two-view matching in [2] to the case of three views in order to improve the accuracy of matching. However, it does not fundamentally solve the problems in [3], and fails to have satisfactory results in certain cases.

3. Our Method.

Our method is meant to overcome the above limitations in the methods [2,3] to robustly generate as many reliable matches or 3D points as possible.

3.1. Local affine model.

Let I_r and I_N respectively denote the reference image and its left/right neighboring image, respectively. For the current match points $u \in I_r$ and $v \in I_N$, we model the geometric transformation by an affine transformation and the illumination effect by a linear transformation between the two corresponding neighborhoods $N(u)$ and $N(v)$, respectively. Then, the affine transformation between $N(u)$ and $N(v)$ can be obtained by minimizing the following residual:

$$\varepsilon = \sum_{m \in N(u)} \left[\left(\mu I_N \left(Am + d \right) + \delta \right) - I_r(m) \right]^2 \tag{1}$$

where $A_{2\times2}$ is a 2D affine matrix and $d_{2\times1}$ is the translation vector; μ and δ respectively depend on the reflection angle of light source and the camera gain, respectively.

In general, (1) can be solved by traditional gradient descent method based on the first-order Taylor expansion at $A = I$ (identity matrix), $\mu = 1$ and $\delta = 0$. The translation component, i.e., d, can be eliminated by centering the coordinate systems to the points under consideration.

However, for the wide baseline images, such a method frequently becomes trapped into local minimum due to non-linear rotation and scale change components implied in A. To address these issues, we choose the neighborhoods $N(u)$ and $N(v)$ with one side parallel to the corresponding epipolar lines to eliminate changes in rotation. For the scale change, we choose the reference image that the affine transformation to the matching image is magnifying, i.e., $|\det A| \geq 1$. With these settings, we found that five iterations are sufficient when solving (1). At each iteration, we compute the ZNCC value between $N(u)$ and $N(v)$. If the ZNCC value is larger than a predefined threshold τ_1 (set 0.9 in this paper) and the number of iterations is less than a predefined threshold τ_2 (set 5 in this paper), then the iteration is terminated and the affine transformation is considered optimal. Otherwise, the affine transformation is considered unreliable, and such unreliability often occurs in regions with large depth changes (e.g., occlusion regions). In this study, matches with unreliable affine transformation are no longer considered seed matches to avoid error accumulation.

As an example, as shown in Figure 1(e), the dotted rectangular regions are the chosen neighborhoods $N(u)$ and $N(v)$ with one side parallel to the corresponding epipolar line (straight line) of the current seed point (big black dot). After five iterations, the points (black dots and crosses) in $N(u)$ better match those corresponding points (black dots and crosses) in $N(v)$ and the ZNCC value is more reliable than the traditional one (e.g., the ZNCC value based on the fixed size of horizontal neighborhoods). A black dot denotes the point where the distances of the corresponding epipolar line to itself are less than 0.05 pixels, and a black cross denotes the point where the distance of corresponding epipolar line to itself is larger than 0.05 pixels but less than 0.1 pixels. Overall, the resulting matches are reliable, and the corresponding affine transformation can be determined approximately.

Furthermore, the size of the neighborhood also needs to be considered considering efficiency and robustness. For an image with a 1024×768 resolution, our method performs well in both accuracy and efficiency using the size of 11×11, and the size can be increased slightly as the image resolution increases. Of course, oversized neighborhoods may incur a heavy computational load and make the affine model unreliable.

3.2. **Effective three-view match propagation.** We improve the method [2,3] in the following four aspects.

(1) Initial seeds are extracted by several feature detection methods (e.g., SIFT, Harris) other than Hessian-Affine detector, and the corresponding initial affine transformation is computed using the above method.

(2) To avoid error accumulation, the affine transformation is optimized by the above method when the ZNCC score of the current seed match is less than τ_1. This helps to guide the propagation proceed toward a better direction and obtain more new matches.

(3) When the queue of seeds is empty, for those regions that are unreached by the previous propagation, a new point matching process based on the ZNCC feature is activated with a much reduced searching range because of the previous matching results. Then, match propagation is initiated in turn from these newly established matches. To ensure the robustness of the ZNCC measures in wide baseline images, choosing the neighborhood with one side parallel to the corresponding epipolar line is necessary.

(4) To improve efficiency, not all new matches detected from the neighborhood of the current seed match are taken as new seed matches to propagate. For a new match, we compute the ratio (denoted by Φ) of all other new matches to all points in the neighborhood of a new match and simply choose the new match with a Φ value that is less than a predefined threshold γ (set 0.4 in this paper) to insert into the queue of seed matches.

The entire process of the improved three-view match propagation is described in Algorithm 1.

Through the above modifications, we found that the improved method is more robust for wide baseline images with large illumination variation and perspective changes, and can generate more reliable quasi-dense matches across three views. As shown in Figure 1(b), the improved method can obtain more reliable matches in the same propagation times in the sample shown in Figure 1(a). In Figure 1(d), certain points in regions (e.g., rectangular regions) in which the method [2] fails are also propagated by the improved method. Our method also performs faster than the method [2] because our method avoids

Algorithm 1: Effective three-view match propagation

Input: three calibrated images.
Output: quasi-dense 3D points.
1: The queue of initial seed matches is constructed.
2: Basic three-view match propagation is performed to obtain new matches.
3: Seeds are selected from these new matches. Each seed should satisfy the following two conditions: (1) with reliable affine parameters; (2) with low Φ value (see the details of the text).
4: Step 2 commences.
5: If the queue of seeds is empty, then new seeds are detected in regions that are unreached by the previous propagation, and Step 2 commences.
6: If each point in the reference image is traversed, then the process is stopped, and the 3D points computed from the propagated matches are output.

the repeated propagation of certain seed matches. More experiments will be shown in a later section.

4. **Experiment.** To evaluate the performance of our method, we have conducted experiments on the following several data sets: (1) multi-view stereo data sets [5] including the Herz-Jesu images (3072×2048) and the Castle images (3072×2048); (2) our own data sets: the Tsinghua Physics Buliding (TPB) images (1224×1848) and the Life Science Building (LSB) images (1092×728).

All the experiments are conducted on desktop PC with Intel Core 2 Duo 2.33 GHz CPU and 16G RAM, and the implementation of each algorithm in all experiments is in parallel C++.

We compare the proposed three-view match propagation method with the method [3] and the PMVS method [4]. In order to make the methods comparable, all methods are set to reconstruct each pixel in the reference image. In addition, in order to evaluate the accuracy of these methods, a reconstructed 3D point is considered as a reliable one according to the depth consistency criterion proposed in [6].

As shown in Figure 2, our method performs better than the method [3] and the PMVS method. Especially, in some regions with larger perspective distortions (e.g., the TPB image and the roofs of the Castle image), the PMVS method has a poorer performance due to the unreliability of the measure used in the propagation. Relatively, although the method [3] takes the affine transformations of local regions into account and thus obtains better results than the PMVS method, it fails to reconstruct more 3D points due to the unreliability of the affine parameters updated in the propagation. In contrast, our method overcomes the limitations in the method [3] and obtains dense results by rectifying the affine parameters in the propagation. In addition, it seems that our method can also obtain good results in some poorly textured regions (e.g., the doors of the Herz-jesu image and the walls of the TPB image). This is because that, in these regions, a new point matching process is activated with a much reduced searching range and then match propagation is initiated in turn from these newly established matches.

Table 1. All 3D points, reliable 3D points and time (min) of different methods

Data sets	Method [3]			PMVS			Ours		
	All	Rel.	Time	All	Rel.	Time	All	Rel.	Time
Herz-Jesu	266765	235567	19.5	230090	218445	17.5	299877	279800	14.3
Castle	198564	160988	18.1	109811	88496	9.8	284451	237884	11.1
TPB	172010	141113	11.8	50030	43693	7.1	220617	201325	8.2
LSB	103871	998530	12.7	63124	53781	8.4	159636	147404	7.8

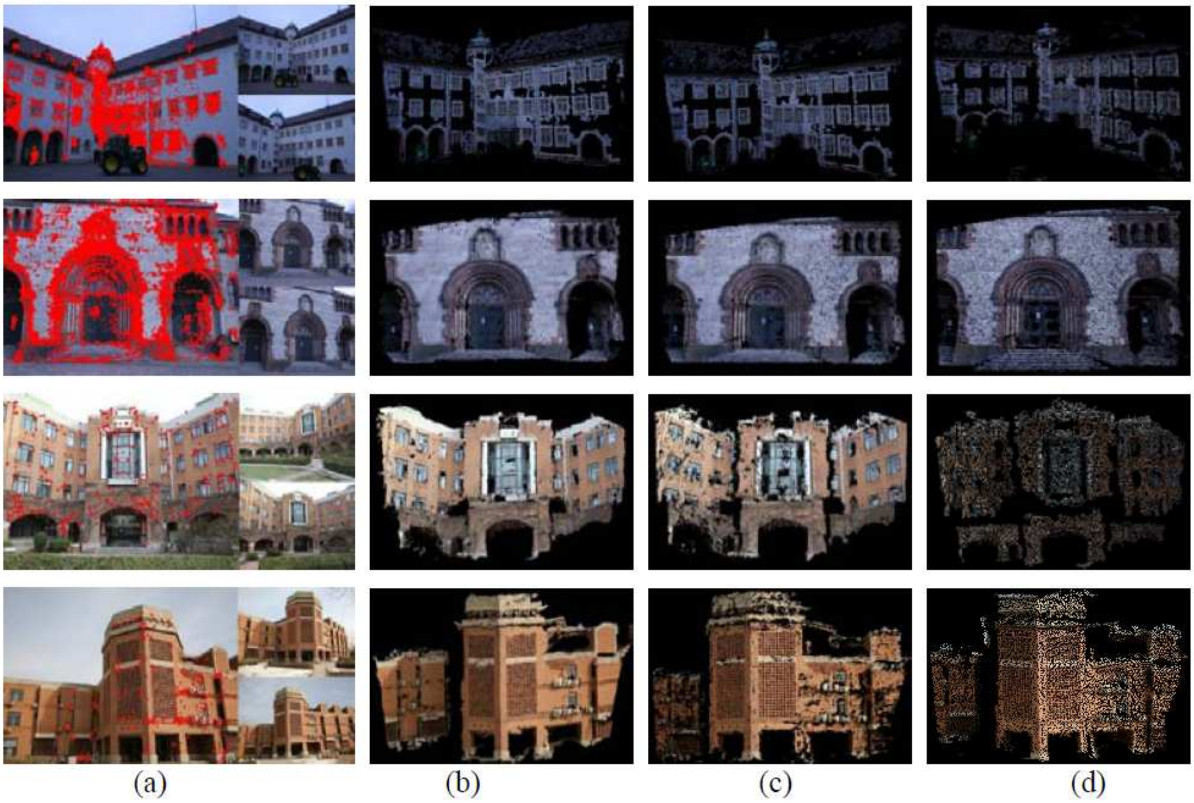

FIGURE 2. Quasi-dense 3D points obtained by different methods: (a) Three images (the left image is the reference image, and shows initial seeds); (b) our method; (c) the method [3]; (d) the PMVS method

Table 1 shows some qualitative results and efficiencies of these methods. Obviously, our method performs well in both the number of reliable reconstructed 3D points and computational efficiency. The method [3] seems to be the slowest among the three methods, and our method overcomes the drawback of the repetitive propagation of seed matches in the method [3], and thus performs faster. For the Castle image, our method is slightly slower than the PMVS method. The reason lies in that our method detects some potential matches in regions where the seeds cannot reach. Overall, our method has better performance.

5. **Conclusion.** In the paper, we have proposed a robust wide baseline matching method which can be used to generate quasi-dense reliable 3D points in multi-view stereo. The proposed method improves the performance of traditional match propagation methods by modeling illumination variations and perspective distortions in images, and performing an effective affine parameter error detection and rectification mechanism in the propagation process. This allows the seed to effectively and efficiently propagate into more regions with different local transformation, and thus can generate better results. In the future, we would explore the possibility of further improving the efficiency of the proposed method by performing a region matching in advance based on color and texture similarity to guide the point match propagation.

Acknowledgment. This work is supported by the National Natural Science Foundation of China (Nos. 61103143, U1404620, U1404622), the Key Scientific and Technological Project of Henan Province (No. 162102310589), the Scientific Research Starting Foundation for Advanced Talents of Zhoukou Normal University (No. zknuc2015103), the Basic and Frontier Project of Science and Technology Department of Henan Province (No. 142300410334), the Higher School Key Research Project of Henan Province (Nos. 17A

520018, 17A520019, 15A520125) and the Application Project of Zhoukou Normal University (No. zknub215204).

REFERENCES

[1] M. Lhuillier and L. Quan, A quasi-dense approach to surface reconstruction from uncalibrated images, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.27, pp.418-433, 2005.

[2] J. Kannala and S. S. Brandt, Quasi-dense wide baseline matching using match propagation, *Proc. of Computer Vision and Pattern Recognition*, pp.1-8, 2007.

[3] P. Koskenkorva, J. Kannala and S. S. Brandt, Quasi-densewide baseline matching for three views, *Proc. of International Conference on Pattern Recognition*, pp.806-809, 2010.

[4] Y. Furukawa and J. Ponce, Accurate, dense, and robust multi-view stereopsis, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.32, pp.1362-1376, 2010.

[5] C. Strecha, W. Von Hansen, L. Van Gool and P. Fua, On benchmarking camera calibration and multi-view stereo for high resolution imagery, *Proc. of Computer Vision and Pattern Recognition*, pp.1-8, 2008.

[6] W. Wang, W. Gao and Z. Hu, Dense 3D scene reconstruction based on semantic constraint and graph cuts, *Science China information sciences*, vol.44, no.6, pp.774-792, 2014.