A Novel Image Matching Method in Camera-calibrated System

Junwei Yu, Lubin Weng, Yuan Tian, Yanqing Wang, Xianqing Tai
Institute of Automation, Chinese Academy of Sciences, Beijing, China
{junwei.yu, lubin.weng, yuan.tian, yanqing.wang, xianqing.tai}@ia.ac.cn

***Abstract*—A novel approach for matching points from two views in a camera-calibrated system is presented in this paper. The interest points are selected using epipolar gradient features. Then three local region invariant descriptors, together with similarity measures are proposed. These descriptors are constructed from binary-threshold gray histogram, sample statistics of the edge points' epipolar gradients, and average intensities of points on the epipolar line, respectively, all image patches based. Similarity measures relative to the descriptors work in cascade. Experimental results demonstrate that our matching scheme is tolerant to image deformations due to changes of viewpoint and effects of perspective, and can find more corresponding points.**

***Keywords*—image matching, epipolar gradient features, computer vision**

I. INTRODUCTION

In the context of application such as image retrieval[1], object recognition[2], scene reconstruction[3], local invariant features have been commonly used, because they are resistant to nearby clutter, insensitive to partial occlusion, and can be computed efficiently.

When the camera calibration is not available, the most common method to sparse matching in the literature is based on the use of Harris feature points, with some correlation-based measure [4, 5]. Many other interest point detectors such as SUSAN operator [6], FAST operator [7], etc. have also been proposed in the past ten years. Studies on the performance evaluation of these interest point detector can be referenced in [8, 9]. Recently, more complex neighborhood feature representations have been developed, and basic types include SIFT [10] and its variant PCA-SIFT [11], GLOH [12], Shape Context [13] and a number of others [14, 15, 16]. All of these are local scale-invariant and affine-invariant feature detectors. Mikolajczyk and Schmid presented a comparative study of these several different descriptors [9]. Their experiments showed that SIFT method obtained the best matching results. The SIFT feature is computed by sampling the magnitudes and orientations of gradients of neighborhood image region and building smoothed orientation histograms. This representation provides robustness against localization errors and small geometric distortions [16].

Although most feature point detectors mentioned above are relatively stable, it is found performances may degrade significantly in some cases. For example, when the scenes do not contain sufficiently suitable textural content, the detected interest points might not be well distributed or stable enough for further utilization. Another dilemma is that, when the scene is not planar, affine invariant would be out of action as the viewpoint changes. In this situation, perspective effects caused by viewpoint changes must be taken into account. Etienne Vincent and Robert Laganière addressed these problems in a calibrated system with constraint of trinocular geometry [17]. In their approach, interest points were detected relying on the concept of epipolar gradient, and compared using edge transfer, resulting in a measure of consistency for point triplets and the edges on which they lay.

For more common two-view calibrated systems, an alternative feature matching scheme is proposed in this paper to cope with the problems stated above but with only two images. We aim to develop a robust feature matching method for epipolar-geometry constrained stereo images system, such as real-time aerial images matching problems. In this application, camera intrinsic parameters are obtained through precise calibration, and extrinsic parameters are provided by high precision inertial navigation system equipped on the aerial vehicles. The new matching scheme consists of three stages. In the first stage, prominent and stable epipolar gradient feature points are selected as interest points. Then in the second stage, three local invariant descriptors of the interest point's neighborhood are generated in terms of information of binary-threshold gray histogram, sample statistics of the edge points' epipolar gradients, and average intensities of points on epipolar line, respectively. In the final stage, candidate matches are measured by calculating the similarities of the three descriptors, in cascade.

The paper is organized as follows. In next section, we review the epipolar gradient feature point selection introduced in [17]. In section 3, the invariant region descriptors are presented and section 4 presents the similarity measures for potential matches decision-making. The experimental results for image matching are given in section 5.

II. EPIPOLAR GRADIENT FEATURE SELECTION

The intensity gradient along the direction of the epipolar line is termed epipolar gradient. It can be obtained by projecting image gradient $\nabla \mathrm{I}(\mathrm{x})$ onto $l = (l_1, l_2, l_3)$, giving the explicit formula,

$$\nabla_{ep}(\mathrm{x}) = \nabla \mathrm{I}(\mathrm{x}) \cdot \left(\frac{-l_3}{l_1}, \frac{l_3}{l_2} \right) \Bigg/ \left\| \left(\frac{-l_3}{l_1}, \frac{l_3}{l_2} \right) \right\| \qquad (1)$$

Where l can be calculated as:

$$l = F^T [K']_\times Fx \tag{2}$$

In formula (2), F is the fundamental matrix, and K', an arbitrary line used, not going through the second image's epipole.

In epipolar geometry, as we all know, epipolar line l in the first image corresponds to the epipolar line l' in the second image. For the continuity of object and images, points on l that are immediately next to x should correspond to points that lie on l' and are immediately next to x'. Consequently, epipolar gradient of point x should be similar to epipolar gradient of point x'. In other words, a point with a high absolute epipolar gradient in one image should have a similar high absolute epipolar gradient in the other as well. This is why epipolar gradient features are good candidates for matching.

In our scheme, those points that satisfy $\|\nabla_{ep}(x)\| \geq a_{ep}$ are selected as interest points, and lower ones are discarded as they are not appropriate for matching. Fig. 1 shows some extracted points, and we can see these points are mostly on the significant edges of the images.

Figure 1. Selected interest points in an aerial image

III. INVARIANT REGION FEATURE DESCRIPTOR

In previous section, interest points are selected using epipolar gradient threshold, and then they should be presented distinctively for further comparing. Unlike those affine-invariant feature descriptors [10-16], we propose three different descriptors which are all developed to reduce the impact of perspective effects caused by changing camera viewpoint. Among these, two descriptors are based on neighboring field-semicircles region, and the third is relative to the neighboring pixels on the epipolar line.

Neighborhood is defined as the region within the interest-point-centered circle, and the radius is determined empirically to be R. As shown in Fig. 2, every epipolar line divides the image into two parts, and each part corresponds to a destined part of the two ones in the second image also divided by the corresponding epipolar line. The parts' correspondence may be determined in terms of the neighboring epipolar lines. Accordingly, the neighborhood of the interest point is also divided into two field-semicircles and each field-semicircle is corresponded loosely to another one in the other image as well. Therefore, descriptors could be generated according to the image patch in the two field-semicircles respectively, and then we can match points field-semicircle by field-semicircle. In this paper, we denote the two field-semicircles by superscript 0 and 1, respectively.

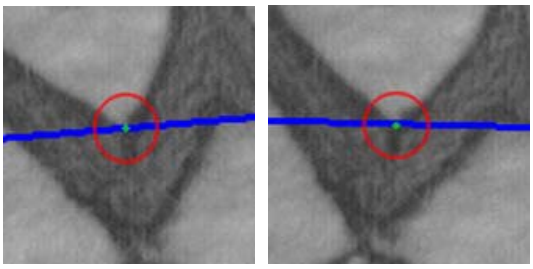


Figure 2. Epipolar lines and neighborhoods (in red circles) illustrations of corresponding points in an image pair

The first descriptor is the binary encoded representation of the two gray histograms built upon two field-semicircles. To deal with the impact of image deformation, the histograms are Gaussian weighed, with weight determined by the distance between the current pixel and the interest point. After the histograms collection, they are divided by a predefined suitable threshold α_{gh} as follows:

$$s_j^i = \begin{cases} 1 & s_j^i \geq \alpha_{gh} \\ 0 & s_j^i < \alpha_{gh} \end{cases} \tag{3}$$

s_j^i is the bin value of the histograms, with superscript i denoting the field-semicircle' s index, and j for the bin number. This representation is robust because the binary forms of the histograms always keep invariant.

The second descriptor is the sample statistics representation of the epipolar gradients of all feature points ($\|\nabla_{ep}(x)\| \geq a_{ep1}$, where a_{ep1} is a new threshold smaller than a_{ep}) falling in the two field-semicircles neighborhood of the interest point. The statistical properties we use here are the Gaussian weighted mean $\mu_{\mathrm{eg}}^{\mathrm{i}}(\mathrm{x})$ and standard deviation $\sigma_{\mathrm{eg}}^{\mathrm{i}}(\mathrm{x})$, with i denoting the field-semicircle's index. As mentioned above, feature points usually locate on significant image edges, and neighboring points often appear to be coherent, concentrated along lines and some curves. Furthermore, the percentage of these points is often low in the image patch; hence there are not a sufficient number of independent measurements for histogram building. Consequently, mean and standard deviation of the epipolar gradients are suitable, and selected for feature description. The descriptor is compound, formed as $(\mu_{eg}(\mathrm{x}), \sigma_{eg}(\mathrm{x}))$, with $\mu_{eg}(\mathrm{x}) = (\mu_{\mathrm{eg}}^{0}(\mathrm{x}), \mu_{\mathrm{eg}}^{1}(\mathrm{x}))^{\mathrm{T}}$, and $\sigma_{eg}(\mathrm{x}) = (\sigma_{\mathrm{eg}}^{0}(\mathrm{x}), \sigma_{\mathrm{eg}}^{1}(\mathrm{x}))^{\mathrm{T}}$, being two column vectors respectively.

The third descriptor is not based on two field-semicircles, but based on two radius line segments on the epipolar line. The Gaussian weighted average intensities on both side of the interest point along the epipolar line are computed to be another invariant representation of the patch, for these values should be preserved in different views of the same point taken simultaneously. Which side of the line corresponds to which in the other image is determined by the comparing of the intensity values. We define the third descriptor as $\mu_{in}(\mathrm{x})=(\mu_{\mathrm{in}}^{0}(\mathrm{x}),\mu_{\mathrm{in}}^{1}(\mathrm{x}))^{\mathrm{T}}$, where $\mu_{\mathrm{in}}^{0}(\mathrm{x})$ is the bigger one of the two computed averages on both sides, and $\mu_{\mathrm{in}}^{1}(\mathrm{x})$ is the smaller one.

IV. MATCHING STRATEGY

In this section, point-to-point correspondence guided by the epipolar geometry would be built using similarity measures defined in the following between the features descriptors proposed in section 3. Then a disparity consistency constraint introduced in [18] is applied to eliminate outliers.

The new similarity measures consist of two feature distance computing, work in cascade. Each one works independently, except that candidate matches denied by first one would no longer be input to the second one for further comparison. The first similarity measure is related to the first feature descriptor mentioned in previous section, and the second measure is a composite similarity measure of the second and third descriptors.

The similarity measurement between the first descriptors of points **x** and **y** is defined in terms of a bitwise exclusive OR operator，

$$s(\mathrm{x},\mathrm{y})=S(\mathrm{x})\wedge S(\mathrm{y}) \tag{4}$$

where $S(\mathrm{x})$ is the collection form of $(s^{0}_{1},\ldots,s^{0}_{n},s^{1}_{1},\ldots,s^{1}_{n})$ in compactly bitwise representation, and n is the total number of bins of the binary-threshold gray histogram. The point y is kept as the candidate match of x, only when $s(\mathrm{x},\mathrm{y})$ is equal to zero.

The qualified candidate point y is then further tested by the similarity measure based on the second and third local image region descriptors. We first define a 2×3 matrix A, as follows:

$$A=\begin{pmatrix}\dfrac{\mu_{eg}(x)-\mu_{eg}(y)}{\sigma_{\Delta\mu eg}} & \dfrac{\sigma_{eg}(x)-\sigma_{eg}(y)}{\sigma_{\Delta\sigma eg}} & \dfrac{\mu_{in}(x)-\mu_{in}(y)}{\sigma_{\Delta\mu in}}\end{pmatrix} \tag{5}$$

where $\sigma_{\Delta\mu eg}$, $\sigma_{\Delta\sigma eg}$, $\sigma_{\Delta\mu in}$, the standard deviations of the corresponding item differences of descriptors, are used to normalize the differences to a similar range. Then the feature distance between interest point x and y is calculated as

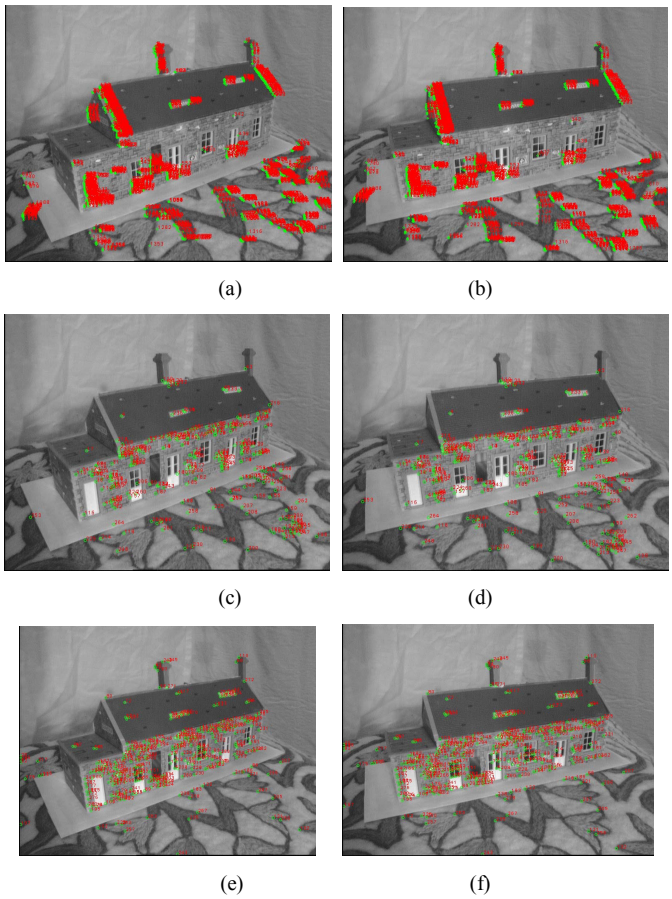
$$d(\mathrm{x},\mathrm{y})=\|A\|_{F} \tag{6}$$

where $\|\bullet\|_{F}$ is the Frobenius matrix norm. This measure will have a low value for corresponding points. Thus if d (**x**, **y**) is the minimum of all the candidate matches of point **x**, and below a threshold a_{d}, the point **y** is accepted as the correspondence. When there are more than one candidate points get the very similar minimum value of d (**x**, **y**), a trick, the order relation of these points on the epipolar line, is used to adjust the points to the appropriate correspondences.

Thus all point-to-point correspondences form a set of initial matches. We refine the matches with disparity consistency constraint for mismatches would be still expected. This constraint is based on the fact that many matches are identified throughout the images, and mistakes are relatively few, therefore they can be eliminated by simply enforcing that matches near to each other have similar disparities. Details of this technology can be referenced in [18].

V. EXPERIMENTAL RESULTS

Our algorithm is verified by stereo pair images matching experiments, and results are compared to two classical methods in the literature, SIFT and Harris/Cross-correlation approach, both appending epipolar geometry constraint to eliminate some outliers. The new matching scheme is applied to a number of image pairs, in this paper, two of them are listed below. To detect more interest points, the threshold a_{ep} used in these experiments, are smaller than they are used in Fig. 1.



(a) (b) (c) (d) (e) (f)

Figure 3. Matches with proposed method (a) and (b), with SIFT (c) and (d), with Harris/Cross-correlation approach (e) and (f), using the model house image pair

Fig. 3 shows the matching result for a model house image pair using the three difference method. The corresponding points are indicated by the same number. Our new matching scheme obtains 1358 corresponding points, with only a few mismatches. The mismatches are relative to the similarities of certain points along the epipolar line. It is observed that points matched using our approach are much more than those using other two methods and some correspondences in deformation areas are robustly identified as expected.

Another experimental results comparison is showed in Fig. 4 using an aerial image pair. The number of detected correspondences is 1606, which is much greater than the other two methods. The impact of the perspective effect is reduced, as some distorted image patches are still matched to one anther. This shows our scheme is effective to aerial images matching issues.

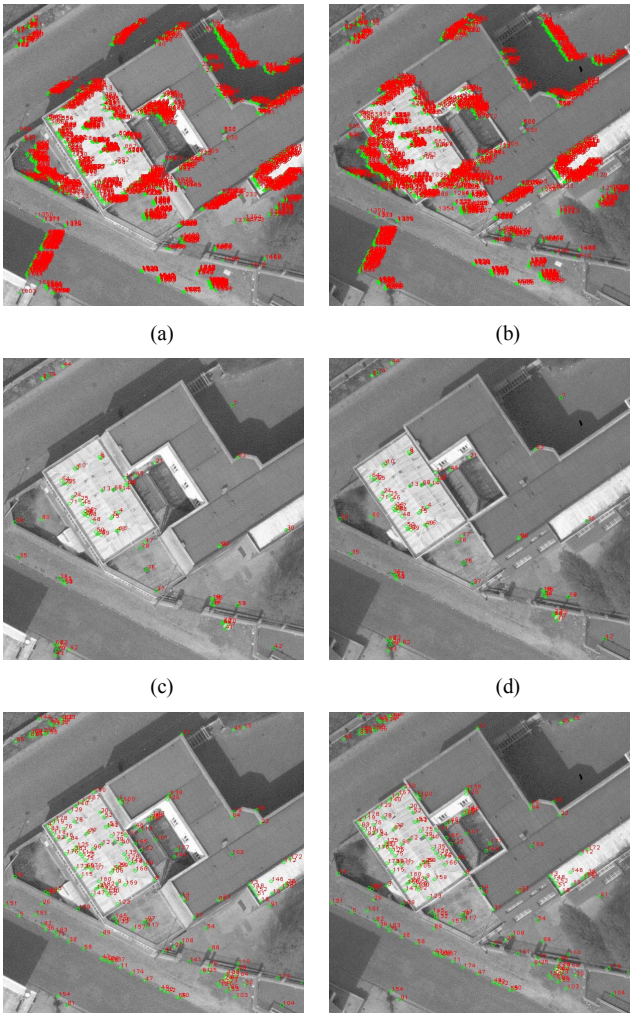


Figure 4. Matches with proposed method (a) and (b), with SIFT (c) and (d), with Harris/Cross-correlation approach (e) and (f), using the aerial image pair

VI. CONCLUSION AND FUTURE WORK

In this paper, three local image region descriptors, together with two similarity measures in cascade are proposed, which are appropriate for matching epipolar geometry constrained image pairs. Our matching scheme is compact, distinctive, and robust to deformation caused by changes of viewpoint, and is capable of providing much more matched corresponding couples. The experimental results verify the algorithm.

In future work, additional statistical properties will be explored for more robust local invariant image representation. And the global matching scheme along the epipolar line using the dynamic programming method is also a topic for further work.

REFERENCES

[1] K. Mikolajczyk, and C. Schmid, "A performance evaluation of local descriptors", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10), 2005, pp. 1615-1630.

[2] R. Fergus, P. Perona, and A. Zisserman, "Object class recognition by unsupervised scale-invariant learning", *In Proceedings of Computer Vision and Pattern Recognition*, June 2003.

[3] R.Hartley, and A.Zisserman, *Multiple view geometry in computer vision*, Cambridge University Press, Cambridge, UK, 2000.

[4] G. Roth, A. Whitehead, "Using projective vision to find camera positions in an image sequence", *Proc. of Vision Interface*, 2000, pp. 225-232.

[5] P. Torr, and A. Zisserman, "Robust computation and parameterization of multiple view relations", *In Proceedings of International Conference on Computer Vision*, 1998, pp. 727-732.

[6] S.M. Smith, and J.M. Brady, "SUSAN-a new approach to low-level image processing", *International Journal of Computer Vision*, 23 (1) ,1997, pp. 45-78.

[7] E. Rosten, and T. Drummond, "Fusing points and lines for high performance tracking", *In Proceedings of International Conference on Computer Vision*, 2005, pp. 1508-1511.

[8] C. Schmid, R. Mohr, and C. Bauckhage, "Evaluation of interest point detectors", *International Journal of Computer Vision*, 37 (2), 2000, pp. 151-172.

[9] F. Mokhtarian, and F. Mohanna, "Performance evaluation of corner detectors using consistency and accuracy measures", *Computer Vision and Image Understand*, 102 , 2006, pp. 81-94.

[10] D. G. Lowe, "Distinctive image features from scale-invariant keypoints", *International Journal of Computer Vision*, 60(2), 2004, pp.91-110.

[11] Y. Ke, and R. Sukthankar, "PCA-SIFT: a more distinctive representation for local image descriptors", *In Proceedings of Computer Vision and Pattern Recognition*, 2004, pp. 511–517.

[12] K. Mikolajczyk, and C. Schmid, "A performance evaluation of local descriptors", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10), 2005, pp. 615–1630.

[13] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(4), 2002, pp. 509–522.

[14] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust widebaseline stereo from maximally stable extremal regions", *Image and Vision Computing*, 22(10), 2004, pp. 761–767.

[15] R. Sára, and M. Matousek, "FAIR: towards a new feature for affinely-invariant recognition", *In International Conference Pattern Recognition*, 2006, pp. 412-416.

[16] Lei Qin, Wei Zeng, Wen Gao, and Weiqiang Wang, "Local invariant descriptor for image matching", *ICASSP2005*, Philadelphia, PA, USA , March, 2005.

[17] E. Vincent, and R. Laganière, "Models from image triplets using epipolar gradient features", Image Vision Computing, 25, 2007, pp. 1699-1708.

[18] E. Vincent, and R. Laganière, "Matching feature points in stereo pairs: a comparative study of some matching strategies", Machine Graphics and Vision, Oct, 2001, pp. 237-259.