IMAGE MATCHING BY NORMALIZED CROSS-CORRELATION

Feng Zhao, Qingming Huang, Wen Gao

Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China
Graduate School of the Chinese Academy of Sciences, Beijing, China
{fzhao, qmhuang, wgao}@jdl.ac.cn

ABSTRACT

Correlation is widely used as an effective similarity measure in matching tasks. However, traditional correlation based matching methods are limited to the short baseline case. In this paper we propose a new correlation based method for matching two images with large camera motion. Our method is based on the rotation and scale invariant normalized cross-correlation. Both the size and the orientation of the correlation windows are determined according to the characteristic scale and the dominant direction of the interest points. Experimental results on real images demonstrate that the new method is effective for matching image pairs with significant rotation and scale changes as well as other common imaging conditions.

1. INTRODUCTION

Image matching plays an important role in many applications. A lot of matching algorithms have been proposed in the literature [1,2]. Matching two uncalibrated images with large camera motion such as significant rotation and scale changes still remains a difficult problem. One effective strategy is using feature matching approach, which extracts salient features such as corners in the two images and then establishes reliable feature correspondences [3,4].

Normalized cross-correlation has found application in a broad range of computer vision tasks such as stereo vision, motion tracking, image mosaicing, etc. Normalized cross-correlation is the simplest but effective method as a similarity measure, which is invariant to linear brightness and contrast variations. Its easy hardware implementation makes it useful for real-time applications.

There have been some image matching methods based on normalized cross-correlation [5,6,7]. However, these methods cannot perform well when there are significant rotation and scale changes between the two images. This is due to the limitation that normalized cross-correlation is sensitive to rotation and scale changes. Therefore, traditional correlation based matching methods are not robust against rotation and scale changes. There are also generalized versions of cross-correlation, which calculate the cross-correlation for each assumed geometric transformation of the correlation windows [8,9]. Although they are able to handle more complicated imaging conditions, the computational load grows very fast in the mean time.

This paper presents a new image matching method based on normalized cross-correlation, which can efficiently handle image pairs with significant rotation and scale changes. First, interest points are detected in the two images separately. Each interest point is assigned one characteristic scale and one dominant direction. Then the new method uses rotation and scale invariant normalized cross-correlation as the similarity measures between two interest points to establish the interest point matches. In order to be invariant to rotation and scale changes, both the size and the orientation of the correlation windows are determined according to the characteristic scale and dominant direction of the interest points. Finally, the epipolar geometry constraint is imposed to reject the false matches. Experimental results demonstrate that the new method performs well on real images with different imaging conditions such as large angle rotation and significant scale changes.

The remainder of the paper is organized as follows. Section 2 describes extracting interest points with characteristic scale and dominant direction. Section 3 introduces the matching algorithm based on rotation and scale invariant normalized cross-correlation and presents in detail the calculation of similarity measures between two interest points. Section 4 describes rejecting the false matches by imposing epipolar geometry constraint. Section 5 presents some experimental results on real images and Section 6 concludes the paper.

2. INTEREST POINTS DETECTION

The interest point detector used in our method is Harris-laplacian detector [10]. The results of Harris-laplacian detector have high repeatability under different imaging conditions such as translation, rotation, scale changes and moderate viewpoint changes [11]. The detector is based on

1-4244-0469-X/06/$20.00 ©2006 IEEE

scale normalized auto-correlation matrix, which is built as follows:

$$M(X,\sigma_I,\sigma_D)=\sigma_D^2 g(\sigma_I)\otimes\begin{bmatrix} I_x^2(X,\sigma_D) & I_xI_y(X,\sigma_D) \\ I_xI_y(X,\sigma_D) & I_y^2(X,\sigma_D)\end{bmatrix}, \quad (1)$$

where $g(\sigma_I)$ is a Gaussian window function at the scale σ_I. $I_x(X, \sigma_D)$ and $I_y(X, \sigma_D)$ are the x and y directional derivatives at the point $X = (x, y)$ of the image I, which are computed with Gaussian kernels of scale σ_D. The corner response function of a point X is defined as:

$$C_{Harris} = \det(M) - k\,\mathrm{trace}^2(M), \quad (2)$$

where k is a constant. A point is selected as a corner if its corner response function is the local maximum in the 8-neighborhood of it and a threshold C_t is given to ensure the salience of the corner [12].

The corners are detected over a set of scales $\sigma_n = s^n\sigma_0$, where σ_0 is the initial scale factor and s is the scale factor between sequential scale levels. In our implementation, 17 scale levels are used to build the scale space representation. The factors σ_0 and s are set to 1.0 and 1.2, respectively. The auto-correlation matrix is calculated with $\sigma_I = \sigma_n$ and $\sigma_D = 0.7\sigma_n$. The threshold C_t is set to 2000 and k is set to 0.04. All the corner points detected in different scale levels form the initial set of interest points.

2.1. Characteristic scale selection

The characteristic scale of an interest point is selected by finding the local extremum of the Laplacian scale selection operator over scales. σ_n is the characteristic scale of an interest point X if

$$F(X,\sigma_n) > F(X,\sigma_{n-1}) \wedge F(X,\sigma_n) > F(X,\sigma_{n+1}), \\ F(X,\sigma_n) > S_t, \quad (3)$$

where $F(X, \sigma_n) = \sigma_n^2\ |I_{xx}(X, \sigma_n) + I_{yy}(X, \sigma_n)|$ and S_t is a threshold, which is set to 20 in our implementation. For two interest points that correspond to the same scene point, the ratio of their characteristic scales is equal to the scale factors between the two images. The characteristic scale of the interest point will be used to determine the size of the correlation window in our method. Note that if an interest point in the initial set has more than one characteristic scale, it will be treated as multiple interest points that each of them has one characteristic scale. And interest points with no characteristic scale will be eliminated from the initial set of interest points.

2.2. Dominant direction assignment

In order to achieve invariance to rotation, each interest point is assigned one dominant direction. The histogram based approach for dominant direction assignment [13] is adopted. An orientation histogram with 36 bins covering the range of 360 degrees is used to accumulate the local gradient orientations within a region around an interest point. The gradient orientation of each sample in the region is weighted by its gradient magnitude and by a Gaussian window.

After building the orientation histogram, we perform a smoothing operation on the histogram by iterative local averaging of every 3 consecutive bins in a cyclical fashion. The orientation corresponding to the largest bin in the smoothed histogram is selected to be the dominant direction of the interest point.

3. MATCHING BY NORMALIZED CROSS-CORRELATION

Matching interest points in two uncalibrated images is a fundamental problem in computer vision. Normalized cross-correlation is widely used in many applications that require matching parts of the images. Traditional matching methods based on normalized cross-correlation can only handle the situation where there are only translation or small rotation and scale changes between the two images. We introduce a new image matching method based on rotation and scale invariant normalized cross-correlation, which can handle more complicated imaging conditions such as large angle rotation and significant scale changes.

In our new method, rotation and scale invariant normalized cross-correlation is used as similarity measure to estimate the difference between the interest points. In contrast to traditional normalized cross-correlation, both the size and the orientation of the correlation windows are determined according to the characteristic scale and dominant direction of the interest points. The matching algorithm is presented in detail as follows:

Let $m_1=I_1(x_i, y_i)$ be an interest point with characteristic scale σ_1 and dominant orientation θ_1 in one image and $m_2=I_2(x_j, y_j)$ be an interest point with characteristic scale σ_2 and dominant orientation θ_2 in the other image. Without loss of generality, we can assume $\sigma_1 \geq \sigma_2$. W_1 and W_2 are two correlation windows of size $(2w+1)\times(2w+1)$ centered on each interest point with $w=\lambda\sigma_2$, where λ is a constant. Let $\theta=\theta_2-\theta_1$ be the rotation angle and $r=\sigma_1/\sigma_2$ be the scale change factor. W_1 is rotated by $|\theta|$ around m_1 (the direction of rotation is counterclockwise if $\theta \geq 0$, and otherwise is clockwise). W_1 and W_2 can then be represented as two arrays of pixel intensities A and B:

$$A_{uv} = I_1(x_i + ru\cos\theta + rv\sin\theta, y_i + rv\cos\theta - ru\sin\theta), \\ B_{uv} = I_2(x_j + u, y_j + v), \quad (4)$$

where u, $v\in[-w, w]$. A_{uv} is calculated using bilinear interpolation. Image I_1 should be smoothed by a Gaussian window function of scale σ_s before calculating A_{uv} when r is large, for example r>2. The similarity measure between the two interest points is then defined as the normalized cross-correlation between A and B.

$$S_{m_1,m_2} = \frac{\sum_{u=-w}^{w} \sum_{v=-w}^{w} \left[A_{uv} - \overline{A} \right] \cdot \left[B_{uv} - \overline{B} \right]}{(2w+1)(2w+1)\sigma(A)\sigma(B)}, \qquad \mathbf{(5)}$$

where $\overline{A}\,(\overline{B})$ is the average and $\sigma(A)(\sigma(B))$ is the standard deviation of all the elements in $A\,(B)$. We use λ=5 and σ_s=1.5 in our experiments.

By adapting the size and the orientation of the correlation window, the similarity measure is robust against rotation and scale changes. The similarity measure decreases monotonically from 1 to -1 with the increase of difference between two interest points. Suppose there are m interest points in the first image and n interest points in the second image, consider a matrix $G \in M_{m,n}$ that its element G_{ij} stands for the similarity measure between the i-th interest point in the first image and the j-th interest point in the second image. In order to establish one to one interest point correspondence, two interest points will be accepted as a match only if their similarity measure G_{ij} is both the greatest element in its row and the greatest element in its column. With selecting all such elements in G, the initial set of interest point matches between the two images can be established.

The new method yields good results in the experiments on real images under different imaging conditions such as large angle rotation and significant scale changes. Some of the results are demonstrated in Section 5.

4. FALSE MATCHES REJECTION

The initial set of interest point matches usually contains some false matches due to the inaccurate characterization of interest point or the improper matches established in the matching procedure. In the case of matching two uncalibrated images, the epipolar geometry can be used to reject the false matches [6,14,15]. The estimation of epipolar geometry from interest point correspondences is performed by the robust estimator RANSAC [16]. Then the interest point matches that are not consistent with the estimated epipolar geometry are identified as false matches and rejected.

5. EXPERIMENTAL RESULTS

In this section, we will demonstrate some experimental results on real images of various content. These images are under different imaging conditions, such as rotation, scale changes, viewpoint changes, etc. Some of the images are from the public domain resources of INRIA and others are collected by our lab. Fig.1-Fig.4 show the final matching results for four different image pairs with significant camera motions.

Figure 1: Matching result for image pair **East_south** (from INRIA) with rotation and scale changes.

Figure 2: Matching result for image pair **Residence** (from INRIA) with rotation and scale changes.

Figure 3: Matching result for image pair **Graffiti6** (from INRIA) with viewpoint changes.

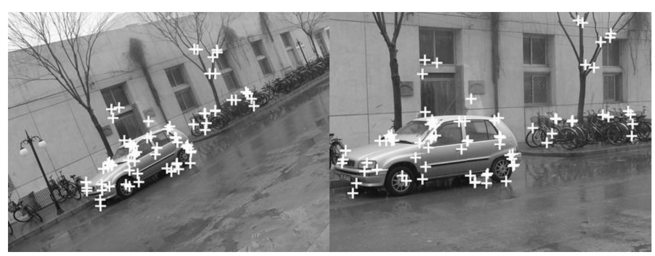

Figure 4: Matching result for image pair **Car** (from our lab) with translation, rotation and scale changes.

Fig. 1 and Fig. 2 show the matching results for image pair **East_south** and **Residence** with large rotation and scale changes. There also exist self-similarity structures in the two images of image pair **Residence**. Fig. 3 shows the matching results for image pair **Graffiti6** with viewpoint changes. And Fig. 4 shows the matching results for image pair **Car** with translation, rotation and scale changes. The numbers of correct matches and the average distances from epipolar lines are illustrated in Table 1.

Table 1: Final matching results for Fig.1-Fig.4

Image Pair	Correct Matches	Average Distance
East_south	49	0.235
Residence	83	0.246
Graffiti6	76	0.341
Car	79	0.294

The experimental results on real images of various content show that our new method is effective for matching image pairs with significant rotation and scale changes, which cannot be effectively handled by traditional correlation based matching methods. Moreover, the new method can handle the situation where there are moderate viewpoint changes between the two images. We also test our method under other common imaging conditions. Fig. 5 demonstrates more matching results with the new method.

6. CONCLUSIONS

We have presented a new correlation based method for matching two uncalibrated images with significant rotation and scale changes. Our method employs rotation and scale invariant normalized cross-correlation defined as the similarity measure between two interest points. The calculation of normalized cross-correlation adapts the size and the orientation of the correlation window according to the characteristic scale and the dominant direction of the interest points. Compared with traditional matching methods based on correlation, our method is able to handle more complicated imaging conditions such as large rotation and scale changes. Experimental results on real images of various content verify the effectiveness of our method.

Acknowledgements

This work is supported by National Hi-Tech Development Programs of China under grant No. 2003AA142140.

References

[1] Heipke C., "Overview of image matching techniques", *OEEPE Workshop on the Application of Digital Photogrammetric Work-stations*, *OEEPE Official Publications*, no.33, pp.173-189, 1996.

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

[3] C. Schmid and R. Mohr, "Local grayvalue invariants for image retrieval", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.19, no.5, pp.530-534, 1997.

[4] A. Baumberg, "Reliable feature matching across widely separated views", In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 774-781, 2000.

[5] W. Förstner, "A feature-based correspondence algorithm for image matching", *International Archives of Photogrammetry and Remote Sensing*, vol.26, no.3, pp.150-166, 1986.

[6] Z. Zhang, R. Deriche, O. Faugeras, and Q.-T. Luong, "A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry", *Artificial Intelligence Journal*, vol.78, no.1-2, pp.87-119, 1995.

[7] M. Pilu, "A direct method for stereo correspondence based on singular value decomposition", In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.261-266, 1997.

[8] H. Hanaizumi and S. Fujimura, "An automated method for registration of satellite remote sensing images", In *Proceedings of the International Geoscience and Remote Sensing Symposium IGARSS'93*, pp.1348-1350, 1993.

[9] R. Berthilsson, "Affine correlation", In *Proceedings of the 14th International Conference on Pattern Recognition*, pp. 1458-1461, 1998.

[10] K.Mikolajczyk and C.Schmid, "Indexing based on scale invariant interest points", In *Proceedings of the 8th International Conference on Computer Vision*, pp. 525-531, 2001.

[11] K. Mikolajczyk and C. Schmid, "Scale & affine invariant interest point detectors", *International Journal of Computer Vision*, vol.60, no.1, pp.63-86, 2004.

[12] C. Harris and M. Stephens, "A combined corner and edge detector", In *Proceedings of the 4th Alvey Vision Conference*, pp. 147-151, 1988.

[13] D. G. Lowe, "Distinctive image features from scale-invariant keypoints", *International Journal of Computer Vision*, vol.60, no.2, pp. 91-110, 2004.

[14] Eiji Nishimura, Gang Xu and Saburo Tsuji, "Motion segmentation and correspondence using epipolar constraint", In *Proceedings of the 1st Asian Conference on Computer Vision*, pp.199-204, 1993.

[15] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, UK, 2000.

[16] M. A. Fischler and R. 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, 1981.

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

Figure 5: Matching results for different imaging conditions. (a) shows 60 inliers for illumination case. (b) shows 49 inliers for blurring case. (c) shows 94 inliers for JPEG compression case. (d) shows 62 inliers for large angle rotation, scale changes plus random noise case.