Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2012, Article ID 382369, 14 pages
doi:10.1155/2012/382369

Research Article

Multiscale Point Correspondence Using Feature Distribution and Frequency Domain Alignment

Zeng-Shun Zhao,[1,2] Xiang Feng,[2] Sheng-Hua Teng,[2] Yi-Bin Li,[1] and Chang-Shui Zhang[3]

[1] *School of Control Science and Engineering, Shandong University, Jinan 250061, China*
[2] *College of Information and Electrical Engineering, Shandong University of Science and Technology, Qingdao 266590, China*
[3] *State Key Lab of Intelligent Technologies and Systems, Tsinghua National Laboratory for Information Science and Technology (TNList), Department of Automation, Tsinghua University, Beijing 100084, China*

Correspondence should be addressed to Zeng-Shun Zhao, zhaozengshun@gmail.com

Received 23 July 2012; Revised 19 November 2012; Accepted 20 November 2012

Academic Editor: Asier Ibeas

In this paper, a hybrid scheme is proposed to find the reliable point-correspondences between two images, which combines the distribution of invariant spatial feature description and frequency domain alignment based on two-stage coarse to fine refinement strategy. Firstly, the source and the target images are both down-sampled by the image pyramid algorithm in a hierarchical multi-scale way. The Fourier-Mellin transform is applied to obtain the transformation parameters at the coarse level between the image pairs; then, the parameters can serve as the initial coarse guess, to guide the following feature matching step at the original scale, where the correspondences are restricted in a search window determined by the deformation between the reference image and the current image; Finally, a novel matching strategy is developed to reject the false matches by validating geometrical relationships between candidate matching points. By doing so, the alignment parameters are refined, which is more accurate and more flexible than a robust fitting technique. This in return can provide a more accurate result for feature correspondence. Experiments on real and synthetic image-pairs show that our approach provides satisfactory feature matching performance.

1. Introduction

Given two or more images of a scene, the ability to match reliable corresponding points between these images is a fundamental and very important problem in computer vision field. In fact, many computer vision applications rely on the success of finding corresponding points [1–5], for example, stereo vision, image registration, motion analysis, object

recognition, and 3D reconstruction. Point correspondences are usually established by matching the local descriptors of a small region around the interest points [6–9]. However, usually a large proportion of them are false matches because of perceptual alias, occlusion, change of illumination and view-points, and so forth. One strong feature may appear weak in the two images, which makes feature matching nearly impossible. In extreme situation, the correspondence is physically meaningless, even though they have similar local appearance. Although the point correspondence problem techniques have been much developed in last decades, it still remains a challenge in various situations.

Nowadays, a considerable amount of previous research has been conducted on the works on efficient feature descriptors, which have used the spatial domain representation of various image features [6–9], such as line segments, corners, implicit and parametric curves and surfaces. Actually, any geometrical feature can be represented as a point set to find meaningful correspondence with another point set. A common approach to obtain such feature that possess above properties is known as "key-point" or "interest point" extraction [6, 7], involving identify points that can be reliably extracted from different images of the same scene. Some of them are well known, for example, Harris, SIFT [7], and SURF (Speeded Up Robust Features) [9, 10]. In the spatial domain representation, many works registered the images by selecting a number of windows in high-variance areas of one image, locating the corresponding windows in the other image and using the window geometric centers as control points to determine the registration parameters [11, 12].

However, those spatial domain approaches conduct exhaustive search of local appearance templates, making it very time consuming and difficult, especially in presence of occlusion junctions, large change of viewpoint, multiple similar structure, and handling of appearing and disappearing features. Even when the most effective invariant descriptors are applied, the performance of feature correspondence in spatial representation is not very satisfactory. These drawbacks are the common problems where overall images are used as the search space for exhaustively finding putative correspondence without guidance. Another more difficult problem is that some difference between the images due to object movements, lighting changes, using different types of sensors or with different sensor parameters, cannot be modeled by a spatial transform alone. They make the registration more difficult since accurate registration can be no longer achieved between two images, even after spatial transformation.

Due to the limitations of spatial domain methods, some researchers take advantage of frequency representation information to assist the image registration process or motion analysis [4, 5, 13, 14]. Since the image pairs can be related by the camera motion which consists of relative translation, rotation, scale, and other geometric transformation, so motion estimation techniques could be introduced into our algorithm. Frequency domain processing has several advantages over spatial domain methods. The motion estimation is based on the phase changes of the Fourier Transformation, so it is robust to global illumination changes. The partial occlusion does not affect the deformation analysis, as the initial geometric transformation estimation is to be obtained in the frequency domain instead of spatial information. However, transformation computation in frequency domain processing alone is not adequate for all image registration tasks, so spatial information is used for more accurate correspondence.

This suggests a simple but effective approach that we denote a coarse-to-fine hierarchical approach. In fact, coarse-to-fine hierarchical ways have been applied by various researchers [3, 15, 16]. However, until now, there have been few approaches to solve the

strongly interconnected problems that can take advantage of both frequency and spatial domain information.

In this paper, we propose a novel way to hierarchically integrate the estimation in the frequency domain and in the spatial domain. These problems are alleviated by firstly resorting to a rough estimate of the transformation parameters between the image pairs at the coarsest level using the frequency information, then this reasonable approximation guide the matching process in spatial domain at the original level. In such a way, we fuse spatial and frequency domain information in a new efficient manner. The integration of frequency and spatial domain can not only avoid the drawbacks of spatial domain methods, but also make use of spatial information for precise feature localization. It should be noted that our idea in certain steps seems similar to some image registration approaches, that employ a set of correspondence features to determine the transformation between the image pairs. Indeed, to the best of our knowledge, this is the first time that by introducing Fourier-Mellin Transform at the coarse-scale, the captured deformation parameters are then applied to assist the feature matching procedure in spatial domain.

The paper is organized as follows. Section 2 introduces our hybrid image registration scheme. Section 3 presents experimental results that demonstrate the advantage of this combination of matching schemes. Section 4 summarizes the paper.

2. Problem Presentation and Our Approach

It is assumed that the image pairs containing the same scene are taken at different times, from different imaging devices, or from different perspectives, due to changes in camera position and pose.

We present a novel algorithm that takes advantage of both spatial and frequency domain information in a hierarchical multiscale decomposition way, as is described in Figure 1. The main idea is to take advantage of the estimation obtained in the frequency domain at the coarse scale, to guide the accurate spatial feature matching. Meanwhile, the multiscale decomposition also reduces much time for the point correspondences. The Fourier-Mellin Transform (in frequency domain) is applied to determine the coarse transformation parameters that map the current image to the source image, which is beneficial to establish the quick correspondence of a set of features. This strategy alleviates the difficult and time-consuming identification of corresponding features in the image pairs and is not dependent upon exact exhaustive searching of point correspondences. Our approach saves much computational efforts, since it need not to search through overall image space for each key point, but in a small window guided by the transformation.

2.1. Coarse Transformation Estimation by Frequency Domain Alignment

Since the image pairs can be related by the camera motion which consists of relative translation, rotation, scale, and other geometric transformation, so motion estimation techniques could be introduced into our algorithm. The affine motion model [11] is adopted in this paper as it provides good tradeoff between generality and ease of estimation. Actually, any Rotation-Scale-Translation (RST) transformation may be expressed as a combination of a single translation, rotation, and scale factor, all operating in the plane of the image. The wrap

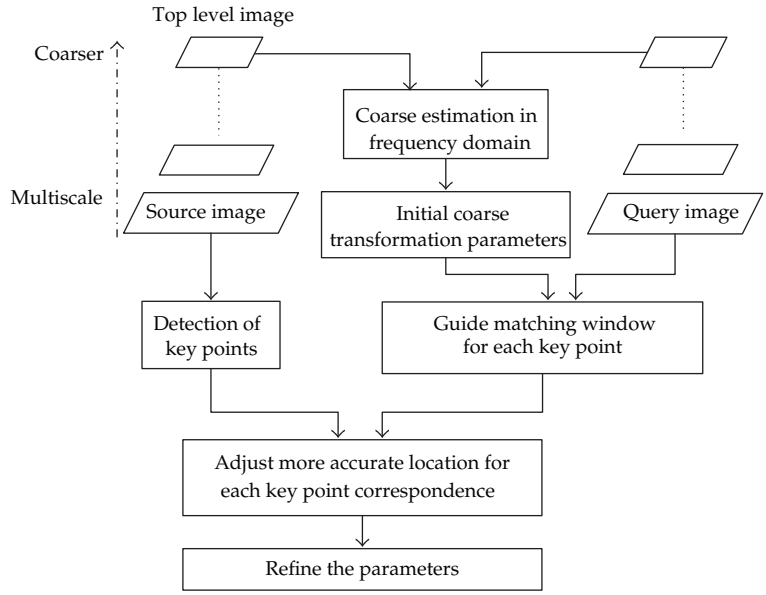


Figure 1: Framework of the presented method.

model between the reference image $f_1(X)$ and current image $f_2(X)$ is mathematically expressed as,

$$f_2(X) = f_1(sRX + t),$$
$$X = \begin{bmatrix} x \\ y \end{bmatrix}, \qquad R = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}, \qquad t = \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix}, \tag{2.1}$$

where X are coordinates of corresponding pixels in current image. sR is linear part and t is translational part of the affine motion parameters, s, θ, Δx, and Δy are the scaling, rotation, and shift along the x- and y-axis. It means that each point $r(x_1; y_1)$ in the reference image maps to a corresponding point $p(x_2; y_2)$ in the current image, according to the matrix equation

$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} s \cdot \cos\theta & -s \cdot \sin\theta & \Delta x \\ s \cdot \sin\theta & s \cdot \cos\theta & \Delta y \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}. \tag{2.2}$$

Global motion estimation methods can be broadly classified into two categories: spatial domain [2, 15, 16] and frequency domain. Frequency domain [4, 5, 13, 14] processing has several advantages over spatial domain methods. The motion estimation is based on the phase changes of the Fourier Transformation, so it is robust to global illumination changes. Its computational cost is significantly lower, making it more useful for practical applications. The partial occlusion does not affect the motion analysis, as the initial geometric transformation estimation is to be obtained in the frequency domain instead of spatial information.

In this paper, we recover the coarse rotation, translation, and scale parameters of the transformation at the top level by using the Fourier-Mellin Transform (FMT), which

is essentially a phase correlation method based on the Fourier and Log-polar transform. The idea behind FMT method is to makes use of the Fourier Shift Theorem and the Fourier Rotation Theorem to provide invariance to rotation, translation, and scale. Then it is performed by phase correlation of the cross-power spectra.

Equation (2.1) can be expressed in Fourier domain as

$$F_2(u,v) = \frac{1}{s^2} e^{-j2\pi(x_0 u + y_0 v)} \cdot F_1\left(\frac{u\cos\theta + v\sin\theta}{s}, \frac{-u\sin\theta + v\cos\theta}{s}\right), \tag{2.3}$$

with $F_2(u,v)$ the Fourier transform of $f_2(X)$ and so is $F_1(u,v)$. Let $\|\cdot\|$ represent the magnitude notation, then $\|F_2(u,v)\|$ and $\|F_1(u,v)\|$ are related by

$$\|F_2(u,v)\| = \frac{1}{s^2}\left\|F_1\left(\frac{u\cos\theta + v\sin\theta}{s}, \frac{-u\sin\theta + v\cos\theta}{s}\right)\right\|. \tag{2.4}$$

We can see that, the Fourier Transform (FT) itself is translation invariant. So the rotation and scaling parameters can be determined independent of the translation parameter.

Since dynamic range of the output of FFT is very high, making interpolation in the frequency domain difficult; this range is compressed by resampling the Fourier magnitude spectra on log-polar grid. When the Fourier magnitude spectra are converted from Cartesian coordinate system to a log-polar representation (ρ, γ) as follows,

$$\mu = \log(\rho), \qquad \nu = \log(s), \qquad \gamma = \gamma. \tag{2.5}$$

Then it converts to polar-logarithmic coordinates so that rotation θ and scale s effects appear as translational shifts along orthogonal γ and $\log\rho$ axes. We can obtain

$$\|F_2(\rho,\gamma)\| = \left\|F_1\left(\frac{\rho}{s}, \gamma - \theta\right)\right\|. \tag{2.6}$$

In other words, it can be written in the following way,

$$\|F_2(\log\rho,\gamma)\| = \|F_1(\log\rho - \log s, \gamma - \theta)\|. \tag{2.7}$$

It can be seen that, the Fourier-Mellin transform (FMT) gives a transform that its resulting spectrum is invariant in rotation, translation, and scale.

We can summarize the Fourier-Mellin Transform (FMT) as follows. Firstly, by working in this translation invariant (Fourier-Mellin) domain, linear component A of affine transformation (rotation angle and scale factor) can be determined by phase-correlation method [8, 9, 13], independent of translational component B. Once linear component has been determined, it can be compensated for and translation. Then, the translation parameters of x-axis and y-axis can be also calculated by the same method. The procedure that obtains the coarse estimation by Fourier-Mellin Transform at the top level of the multiscale image pyramid is also depicted in Figure 2.

Since Fourier magnitude spectrum is applied as translation invariant domain, FFT of whole original image is needed, making it computational expensive. This problem is

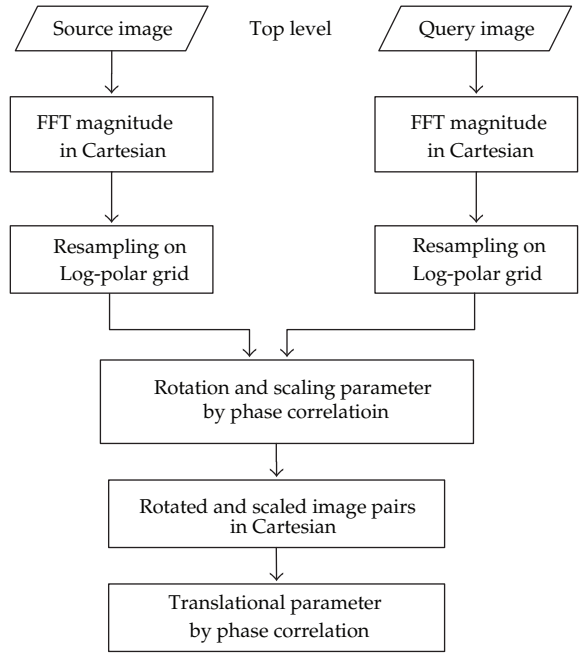


Figure 2: Coarse Estimation By Fourier-Mellin Transform at the top level of the multiscale image pyramid.

alleviated by multiscale decomposition [3, 15, 16]. The coarse-scale images contain only the main shapes and general features and less noise, so resulting transformation procedure is much faster than in original fine scale. For better efficiency, integer low-pass filter is used with very few nonzero bits in the coefficients. Thus, supplement the frequency method with a coarse-to-fine multiresolution approach and feature-based registration can overcome most limitations of the previous scheme.

2.2. Guided Constrained Search in Spatial Domain

Thus, by Fourier-Mellin Transform in the frequency domain of the coarse-scale image, the initial transformation parameters in (2.1) are obtained. Once the transformation is computed, the coarse locations of corresponding points are simply handled by applying the transformation to each interest point extracted in the reference image:

$$f_2\left(x_i'\right) = f_1(sRx_i + B), \tag{2.8}$$

where in the query image, x_i' is the ideal corresponding location relating the interest point x_i in the reference image. Due to measurement errors and other uncertainties in camera position and orientation, matching points may not occur exactly on the estimated mapping locations in the image plane; in this case, a search in a small neighborhood is necessary. So a search window centered about the ideal mapping location x_i' is used to significantly limit the search space for finding conjugate point-pairs. Under the guidance of transformation of each interest point, efficient candidate correspondence can be estimated within a small region, without

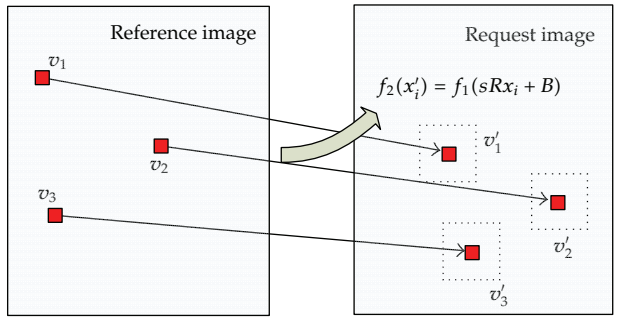


Figure 3: The transformation obtained by frequency information guide the correspondence search within small windows, whose centers are ideally mapped from those interest points detected in the reference image.

meaningless exhaustive search in the entire image. This strategy is illustrated in the following figure.

2.3. Correspondence Refinement by RANSAC and Geometrical Relationships

Once the transformation parameters are obtained in the frequency domain, under the guidance for each interest feature extracted in the source image, a set of features correspondence can be established by search within a small area around the ideal projected center. Even so, it is not assured that all of the matches are necessarily the exact correspondences. Sometimes, even a small error can have a large influence in the recovered parameters. Actually, in the case of occlusion or removal, a most similar feature point within that window will be proved to be false candidate match.

To identify and eliminate outliers, we apply the robust estimation algorithm RANSAC [17] to find the transformation that is consistent with the largest number of inliers. Inliers are defined as those putative matches $\{x_i, x'_i\}$ such that

$$\left\| f_2(x'_i) - f_1(x_i) \right\| < \lambda, \tag{2.9}$$

where λ is a threshold. RANSAC returns the transformation with the largest consensus and the list of matches in the consensus set. If the set of inliers changes with the improved transformation, we continue to reestimate the transform until the consensus set converges. Those false matches which are not consistent with the dominant transformation are rejected as outliers, ignored at the rest steps. To further improve the estimate, we use the consensus set to reestimate the transform with all inliers. At each iteration of RANSAC, the location of each inlying corresponding feature is rectified as follows:

$$x'_{i,k} = \frac{\left(x'_{i,k} + x'_{i,k-1}\right)}{2}, \tag{2.10}$$

where the $x'_{i,k-1}$ and $x'_{i,k}$ are the location of the ith corresponding point mapped by dominant transformation at the $(k-1)$th and kth iteration. This update can undoubtedly accelerate

the convergence process. By this step, simultaneous precise feature correspondence and refined wrap between the image pairs are achieved in a hybrid way.

Then, we make use of the distribution of collections of nearby interest points to increase the correspondence belief for each other . But how to select such a group of points and what metrics can be utilized to enhance the performance is a challenge. Following the works in [18], we make use of structural relationship of interest points to avoid the false matches caused by local similar regions. The stable geometrical relationships between a set of interest points can make such a group of points distinct from the similar ones, even in the case that they looks similar within the single local neighborhood.

An instance of the distribution of nearby corresponding point sets is designed as follows. For every initial matched feature point pair $\{z_i \leftrightarrow z_i'\}$, three nearest points $\{v_1, v_2, v_3\}$ around z_i in the source image and three nearest points $\{v_1', v_2', v_3'\}$ around z_i' in the target image are selected in the reference image, as well as in the captured image, as illustrated in Figure 4. It is assumed $d_1 < d_2 < d_3$ and $d_1' < d_2' < d_3'$ in Euclidean distance. Then any two of these three points and z_i can construct an angle. Next, we start with the point v_1 and compute the angle from it to the second point v_2 and z_i which is marked as α:

$$\begin{aligned} \tan\alpha &= \frac{(k_{v_2 z_i} - k_{v_1 z_i})}{(1 + k_{v_2 z_i} k_{v_1 z_i})}, \\ d_1 &= \sqrt{(v_1 \cdot x - z_i \cdot x)^2 + (v_1 \cdot y - z_i \cdot y)^2}. \end{aligned} \tag{2.11}$$

The angle from v_2 to v_3 and z_i is marked as β. These variables can be computed in accordance with forums to the above. Also, we compute α', β', d_1', d_2', and d_3' by the corresponding points in the other image. For a correct matching point pair, the ratio between α and β is close to ratio between α' and β'. Further, the ratio between d_1 and d_2 or d_2 and d_3 is close to that in the other image, as expressed in the following equations:

$$\begin{aligned} \left|\frac{\alpha}{\beta} - \frac{\alpha'}{\beta'}\right| &< c_1, \\ \left|\frac{d_1}{d_2} - \frac{d_1'}{d_2'}\right| &< c_2, \\ \left|\frac{d_1}{d_3} - \frac{d_1'}{d_3'}\right| &< c_3, \end{aligned} \tag{2.12}$$

where c_1, c_2, and c_3 are the predefined thresholds to justify whether the neighboring points around the potential corresponding point are also matched well.

3. Experiments

In this section, we conduct the point correspondence experiments using both the real images and synthetic deformed image-pairs. We demonstrate the accuracy and the robustness of the algorithm presented in the second section.

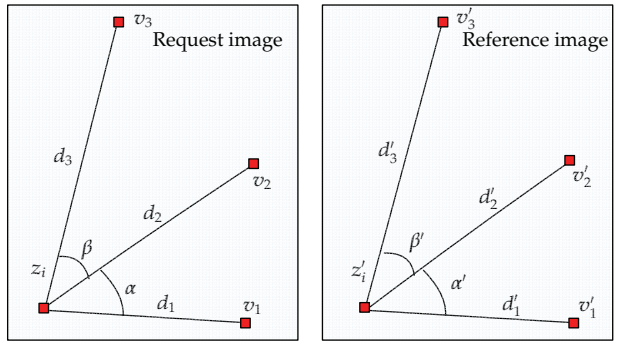
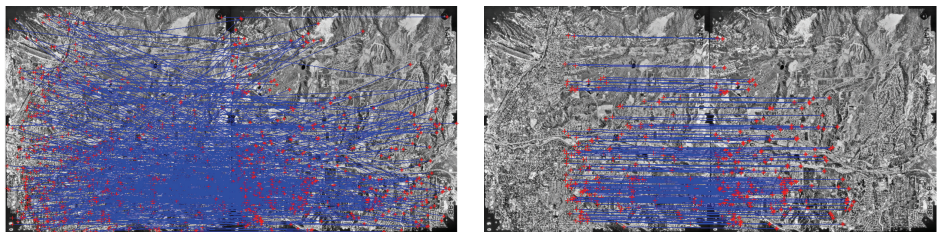


Figure 4: The illustration of an instance our strategy. For each initial matched feature point pair, the geometrical relationships between the three nearest points around z_i are applied to describe the distribution.



(a) (b)

Figure 5: Some examples of feature matches in the image pairs named as col90p1 and col91p1, which are supplied by Leica Geo-systems Geospatial Imaging. A few of them are false matches due to the similarity of local appearance information.

3.1. Real Remote-Sense Image Matching

Image matching plays a critical important role in remote sensing applications. Due to the large volumes of remote-sensing data available, automated feature correspondence is highly desirable. We will consider images that differ by a approximate planar motion, which is suitable for remote sensing image registration. To measure the performance of the proposed method, we apply our algorithm to two sets of images. The first set is the image pairs named as col90p1 and col91p1 (showed in Figure 5), which are supplied by Leica Geo-systems Geospatial Imaging. The second set is the image pairs of Ji-Ning coal area captured by the satellite SPOT5.

As noted above, any geometrical feature can be represented as a point set to find meaningful correspondence with another point set. However, it is important for feature-based methods to adopt discriminative and robust feature descriptors that are invariant to the differences between the two image pairs. Lowe [7] presented the SIFT method to extract distinctive invariant features from images. These features are invariant to image scale and rotation and provide robust matching across a substantial range of affine distortion, addition of noise, and changes in illumination. Gauglitz [10] had shown that of several currently used key point descriptors, SIFT descriptors are the most effective. In this experiment, SIFT detector and descriptor was adopted for its effective invariant attribute.

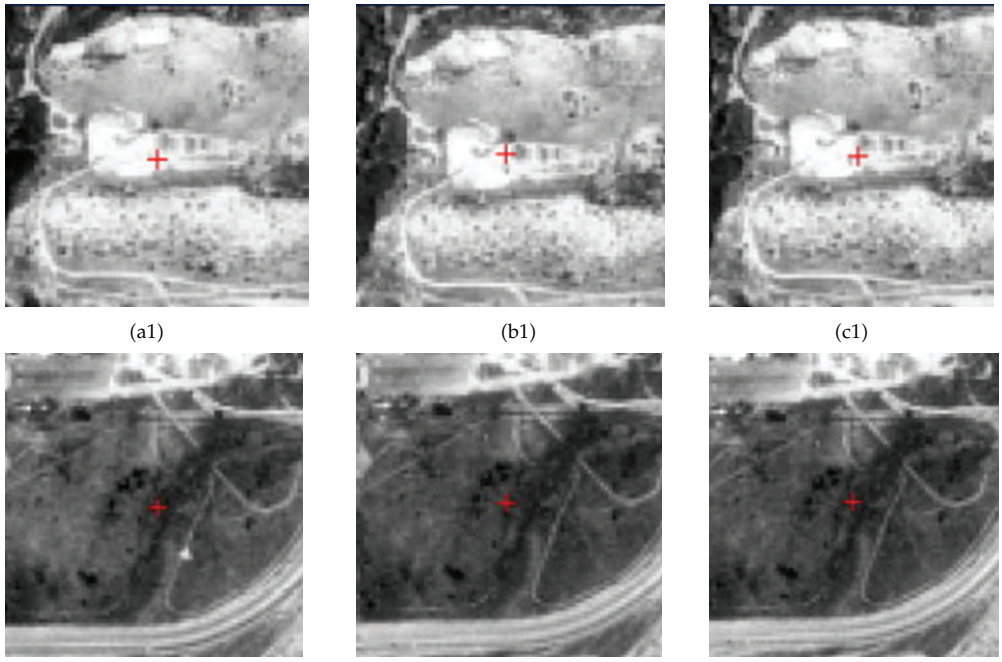
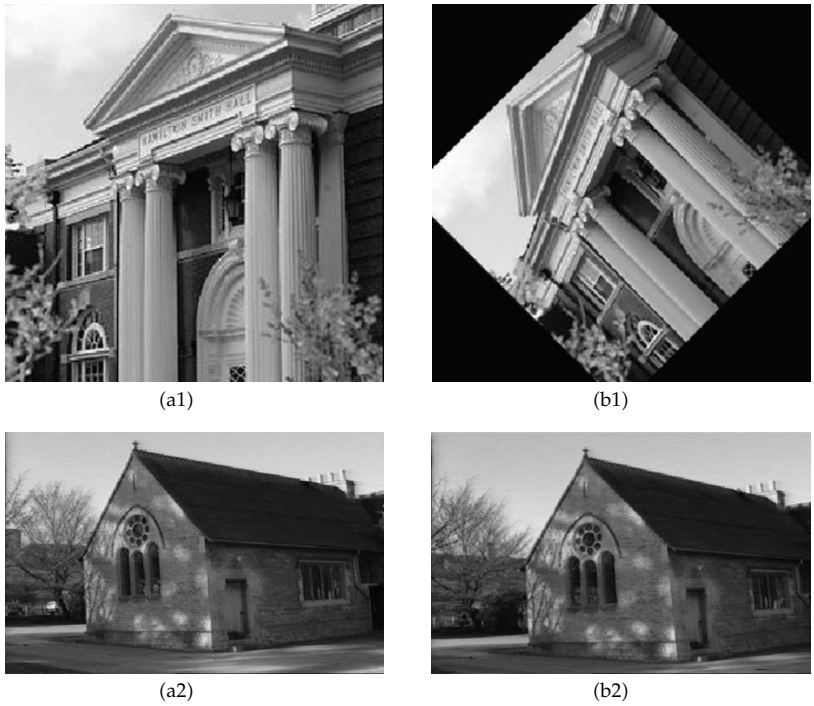


Figure 6: Two examples of performance comparison of standard SIFT and the proposed method. The first column: original interest point in the reference image. The second column: corresponding interest point using standard SIFT detection in the current image. The third column: the rectified locations yield subpixel accuracy using our method.

First, the standard SIFT matching procedure is implemented to the image pairs supplied by Leica Geo-systems Geospatial Imaging. Figure 5 shows the feature correspondences between the image pairs. It can be seen from Figure 5 that, although most classical SIFT features contain enough discriminative information to match with other corresponding ones, some of them are false matches due to the similarity of local appearance information. If the corresponding point pairs are enlarged to watch the details, just like we do in Figures 6 (a1), (b1), (a2) and (b2). From the images comparison of every left columns and middle columns, it shows that the locations extracted by SIFT algorithm are not accurate sufficiently, at least several pixels apart.

Currently in our algorithm, 3-pixel is used as the projection error threshold for RANSAC and we repeat the RANSAC loop just for 10 times using 3 putative matches to compute the affine wrap parameters. The length and width of the constrained window for each mapped point are both 20 pixels, as showed in the right part of the Figure 3.

In the first experiment, given two sets of interested points detected and described by SIFT feature in the Figure 5(a), the initial set of putative correspondences is established by standard SIFT matching procedure, which contains 1952 inliers and 1327 outliers totally. All of the 3279 putative point correspondences are presented in Figure 5(a), from which we can see almost half of them are mismatches (actually 40.47% are outliers). From Figure 5(b), we can see that the results are satisfactory, where 98.6% of the outliers are correctly detected by checking their consistency with the known transformation parameters. The percentage of outliers is dramatically reduced from 40.47% down to 0.56%.



(a1) (b1)

(a2) (b2)

Figure 7: The Example of Fixed Source Image and the Synthesized Image pairs.

The experiment on SPOT5's Ji-Ning coal satellite images also achieves satisfactory result. The transformation parameters are [1.0007, −0.0226, 48.4592, 0.0323, 0.9969, and −806.2157] in the form of $[s \cdot \cos\theta, -s \cdot \sin\theta, x_0; s \cdot \sin\theta$, and $s \cdot \cos\theta, y_0]$ according to (2.2).

We also performed the experiments on ten pairs of images from several distinct domains, including medical scans, natural scenes, and military surveillance. The proposed method is computationally efficient. This comes from the shift property of the Fourier Transform (FT) and the use of Fast Fourier Transform. Another reason is that the transformation obtained in frequency domain at the coarse layer, can serve as an initial good guess to the matching process, leading to an easy search within a small region.

3.2. On Synthetic Image-Pairs

To further quantitatively evaluate the accuracy of the proposed technique, we perform an easy way to do the evaluation, in which a known transformation (we take affine transformation as the concrete example of the transformation style in this case) is applied to the source image, and the estimated transformation is compared with the known transformation parameters, to see the accuracy. Two examples are shown in Figure 7. We then decompose the image pairs into 5 hierarchical layers. The initial guess of the transformation parameters is obtained by frequency analysis using the first coarse scale images, and then is tuned at the original level by robust spatial features matching techniques.

Under various parameters of transformation, we compute three metrics, including Root Mean Square (RMS) error between the point sets after alignment, the correlation coefficients between the original image and rectified image, and ratio of outlier to putative

Table 1: RMS error and the correlation coefficient between after the transformation.

Trial number	1	2	3	4	5
RMS error	0.021	0.033	0.043	0.058	0.038
Correlation coefficient	0.993	0.992	0.989	0.988	0.993
Outlier rate	0.42%	0.46%	0.58%	0.65%	0.43%

Table 2: The average relative error and computation time performance compared to FFT and HMIR method.

	Average relative error (%)				Time (sec)
	Δx	Δy	θ	s	
FMT method	0.9	1.2	4.3	2.7	18.4
HMIR method	4.3	3.9	3.6	2.2	34.7
Proposed method	0.4	0.6	1.4	1.9	21.8

matched point-pairs. The RMS error represents the difference between the original control points and the new control point locations calculated by the transformation process. Note that optimum value of the RMS error is 0, indicating exact matching between the images before and after the rectification; while poor matches result in large RMS error values, small correlation coefficients and high outlier rate.

In detail, for the fixed source image, we constructed a target image set which contains 100 frame images that were synthesized from random affine transformations with rotation $\theta \in (-45°, 45°)$, scale $s \in (0.6, 1.4)$, and translation $t \in (\pm 0.25 \times \text{width}, \pm 0.25 \times \text{height})$ pixels. The input image pairs present scale, rotation, and shift changes. We list the 3 metric items using the first 5 synthesized image pairs under random affine transformations (without explicit deformation parameter). The 3 metric items, such as correlation coefficients, RMS error, and outlier rate presented in Table 1 quantitatively confirmed the accuracy of the proposed method. For all of the image pairs between fixed image and the various synthesized images, the proposed technique significantly improves the 3 metric items. The results in Table 1 demonstrate that the algorithm successfully recover the deformation between the image pairs. The capability of detecting outlier is very robust and the capability of matching correctly is not weakened by the variation of deformation parameter such as scale, rotation, and translation.

Comparison to Fast Fourier Transform (FFT) [13, 14] and mutual-information-based registration (HMIR) [16] is provided in Table 2. We repeated 10 times on the five pairs of test images, which are synthesized using random affine transformations, as stated above. The table lists the average relative error and computational time that each algorithm needs. For each of the three algorithms, there are five columns, Δx, Δy, θ, s, and the shift parameters along the x- and y-axis, the rotation, and scaling parameters using the three algorithms. The table shows that the transformation parameters estimated using the proposed method are very close to those actual parameters. Our approach demonstrates robustness with high accuracy. The mutual information based approach tends to be unstable, especially for large rotation angles.

In the results, SIFT points are used for the comparison but not for a final application (the computation time for our proposed method does not take the SIFT point detection into account). Actually, we also test using Harris corners could be greatly faster without the degradation of accuracy. The type of the point detector (using SIFT detector or Harris corner detector) do not have any influence on the performance of proposed method.

4. Conclusions

In this paper, the combination of these different methods in spatial-frequency domain tends to compensate for any deficiencies in the individual methods. The integration of frequency and spatial domain can simultaneously find the correct feature correspondences within small support windows, the mapping between these image pairs of a same scene and have a more accurate location result after the rectification step. It has been shown that our hybrid hierarchical estimation techniques can achieve efficient and robust performance.

Acknowledgments

The authors would like to thank to the associate editor and the anonymous reviewers for their careful work. This research has been supported by the National Natural Science Foundation of China (Grant nos. 60805028, 60903146), Natural Science Foundation of Shandong Province (ZR2010FM027), Zhejiang Provincial Natural Science Foundation of China (no. Y1110661), and China Postdoctoral Science Foundation (2012M521336).

References

[1] M. Z. Brown, D. Burschka, and G. D. Hager, "Advances in computational stereo," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 8, pp. 993–1008, 2003.

[2] Y. Bentoutou, N. Taleb, K. Kpalma, and J. Ronsin, "An automatic image registration for applications in remote sensing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 9, pp. 2127–2137, 2005.

[3] P. Thévenaz, U. E. Ruttimann, and M. Unser, "A pyramid approach to subpixel registration based on intensity," *IEEE Transactions on Image Processing*, vol. 7, no. 1, pp. 27–41, 1998.

[4] H. Foroosh, J. B. Zerubia, and M. Berthod, "Extension of phase correlation to subpixel registration," *IEEE Transactions on Image Processing*, vol. 11, no. 3, pp. 188–200, 2002.

[5] B. S. Reddy and B. N. Chatterji, "An FFT-based technique for translation, rotation, and scale-invariant image registration," *IEEE Transactions on Image Processing*, vol. 5, no. 8, pp. 1266–1271, 1996.

[6] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," in *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, pp. II257–II263, June 2003.

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

[8] R. Fergus, P. Perona, and A. Zisserman, "Object class recognition by unsupervised scale-invariant learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '03)*, pp. II 264–II 271, June 2003.

[9] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, "SURF: speeded up robust features," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.

[10] S. Gauglitz, "Evaluation of interest point detectors and feature descriptors for visual tracking," *International Journal of Computer Vision*, vol. 94, no. 3, pp. 335–360, 2011.

[11] R. N. Bracewell, K. Y. Chang, A. K. Jha, and Y. H. Wang, "Affine theorem for two-dimensional Fourier transform," *Electronics Letters*, vol. 29, no. 3, article 304, 1993.

[12] G. Hager and K. Toyama, "Xvision: combining image warping and geometric constraints for fast visual tracking," in *Proceedings of the 4th European Conference on Computer Vision*, pp. 507–517, 1996.

[13] Y. Keller, A. Averbuch, and M. Israeli, "Pseudopolar-based estimation of large translations, rotations, and scalings in images," *IEEE Transactions on Image Processing*, vol. 14, no. 1, pp. 12–22, 2005.

[14] P. Vandewalle, S. Süsstrunk, and M. Vetterll, "A frequency domain approach to registration of aliased images with application to super-resolution," *Eurasip Journal on Applied Signal Processing*, vol. 2006, p. 233, 2006.

[15] R. Szeliski and J. Coughlan, "Hierarchical spline-based image registration," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 194–201, June 1994.

[16] A. A. Cole-Rhodes, K. L. Johnson, J. Lemoigne et al., "Multiresolution registration of remote sensing

imagery by optimization of mutual information using a stochastic gradient," *IEEE Transactions on Image Processing*, vol. 12, no. 12, pp. 1495–1511, 2003.

[17] 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 Association for Computing Machinery*, vol. 24, no. 6, pp. 381–395, 1981.

[18] Z.-S. Zhao, Q.-J. Tian, J.-Z. Wang, and J.-M. Zhou, "Image match using distribution of colorful SIFT," in *International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR '10)*, pp. 150–153, 2010.