Efficient Globally-Optimal Registration of Remote Sensing Imagery via Quasi-Random Scale Space Structural Correlation Energy Functional

Wen Zhang, Alexander Wong, *Member, IEEE,* Akshaya Mishra, *Member, IEEE,* Paul Fieguth, *Member, IEEE,* and David A. Clausi, *Senior Member, IEEE*

***Abstract*— A novel energy functional for automatic registration of remote sensing imagery based on quasi-random scale space structural correlation is presented. The structural correlation energy functional takes advantage of the fact that for many types of remote sensing imagery, there exists common structures at different scales even if the acquired images have very different intensity characteristics. The proposed energy functional also takes advantage of the noise robustness and feature localization properties of quasi-random scale space theory. An efficient globally exhaustive optimization strategy in the frequency domain is developed for registering remote sensing imagery based on the proposed energy functional. Promising test results on interband, intraband, and intermodal remote sensing image sets show that the proposed method has the advantages of being robust to differing sensing conditions and large misalignments.**

***Index Terms*— registration, energy functional, quasi-random scale space, structural correlation.**

I. Introduction

Remote sensing image registration can involve images taken at different times and/or captured using different sensors and/or using different bands. Registration is of significant value in remote sensing applications such as building extraction, environmental modeling, and change detection. Given a pair of remote sensing images f and g, the underlying goal is to determine the transformation T that brings f into alignment with g such that the energy functional C is maximized

$$T^* = \arg\max_T \Big[C\big(T[f], g\big) \Big] . \tag{1}$$

To handle the different intensity characteristics of image pairs acquired under different conditions, a class of algorithms [1], [2] have been developed based on mutual information (MI) [3], [4]. However, due to the presence of many local optima on the convergence plane [5], converging to the global optima using iterative optimization methods such as conjugate gradient [6] and Nelder-Mead simplex [7] can be difficult to achieve with a MI-based energy functional. In response to these difficulties, energy functionals based on intensity remapping such as correlation ratio [5] have been proposed but are still sensitive to large initial misalignments.

Globally exhaustive search methods, on the other hand, are not sensitive to local optima or to large initial misalignments, and a class of efficient frequency domain algorithms have been proposed [8], [9], [10], [11], [12]. However, since most of these methods are based on the conventional cross-correlation as the energy functional, they are not well-suited to registering remote sensing images acquired using different sensors and/or different bands because such images may have very different intensity characteristics.

To overcome this limitation of globally exhaustive search methods, we propose a novel energy functional that fuses structural correlation with quasi-random scale space theory [13], which allows for robust registration of remote sensing imagery acquired under different sensing conditions. The proposed energy functional attempts to overcome the difficulties faced when using cross-correlation by taking advantage of the presence of common structures at different scales between images that may have very different intensity characteristics. Note that the proposed energy functional is designed to take advantage of common structural characteristics between images and may not be suitable for situations where little to no common structures are captured between the acquired images.

Based on this energy functional, a globally exhaustive optimization framework in the frequency domain is developed. Although the optimization framework discussed in this paper assumes rigid transformations (rotations and translations only), the proposed energy functional can be applied in optimization frameworks dealing with non-rigid transformations. Furthermore, achieving an accurate rigid registration between the images under evaluation is an important first step in aiding the subsequent non-rigid registration process [14], [15] to converge to the correct solution.

To the best of the authors' knowledge, this approach to registration for remote sensing imagery has not been proposed before, and presents a significant advancement in the capabilities of efficient globally exhaustive search algorithms, since other energy functionals such as correlation ratio and mutual information cannot be performed exhaustively for all possible rotations and translations in an efficient manner.

II. Quasi-Random Scale Space Structural Correlation Energy Functional

Remote sensing images taken under different sensing conditions such as different sensors and different bands often have very different intensity characteristics. An example of this is shown in Fig. 1, where an optical/LIDAR image pair is shown.

Wen Zhang, Alexander Wong, Akshaya Mishra, Paul Fieguth, and David A. Clausi are with the Department of Systems Design Engineering, U. of Waterloo, Waterloo, Canada N2L 3G1. {wxzhang,a28wong,akmishra,pfieguth,dclausi}@uwaterloo.ca

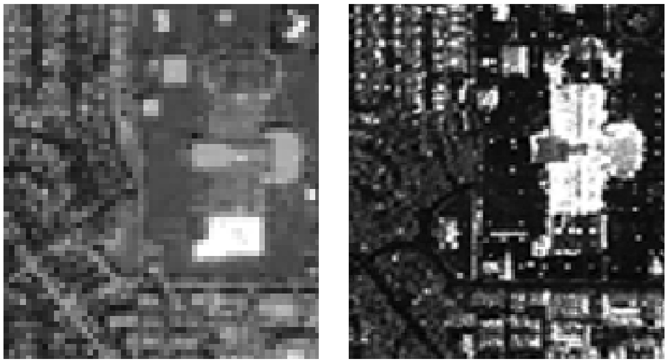


Fig. 1. In this example, the optical image (left) has very different intensity characteristics than the LIDAR image (right) of the same area. As such, a conventional cross-correlation approach would fail to register such an image. However, the two images share a large number of common structures at different scales, thus motivating the use of energy functionals that take structural characteristics into account.

The underlying intensity characteristics of the optical image differs significantly from that of the LIDAR image, making the use of conventional cross-correlation ill-suited. However, the image pair shares a large number of common structures at different scales. Hence, we are motivated to design an energy functional that captures the *structural* information as the basis of registering the images so that we can take common structural characteristics into account.

In most remote sensing imagery, the important structural characteristics exist at a variety of different scales [16]. Thus, an effective approach to capturing structural information would be to decompose the image into the scale space domain, the most popular form of which is the Gaussian scale-space [17]. However, the Gaussian scale space is limited by its sensitivity to noise and poor structural localization [18], therefore to overcome these limitations, Mishra et al. proposed a quasi-random scale space theory [13], which is obtained as follows.

In scale space theory, an image $I(s)$ is represented as a family of M scale space representations $L_0(s), L_1(s), \ldots, L_M(s)$ representing image detail at different scales. As the scale increases, more and more of the details contained within the image becomes removed. Therefore, at lowest scale $i = 0$, the scale space representation $L_0(s)$ is just the original image $I(s)$ and as such contains all of the fine details, while at the highest scale $i = M$, the scale space representation $L_M(s)$ consists of mainly the coarse, large scale details of the original scene. This separation of detail amongst the family of scale space representations allows us to identify structural details that exist at the finer, smaller scales as well as at the coarser, larger scales within the image, which is very important for emphasizing common structures that exists between the images under alignment. Unlike conventional multi-resolution approaches where the resolution changes at each level, the resolution is maintained at each level of the scale space decomposition. Based on testing, a total of $M = 5$ scales was used as it was found to provide the best results.

Let us now study the concept of scale space decomposition from a mathematically perspective. Let $\mathcal{L}$ be a discrete lattice with a set of sites S. Let the original scene $I(s)$, the scale space representation $L_i(s)$, and the residual fine scale structure $C_i(s)$ be random fields on $s \in S$, where the index $i = 0, \ldots, M$ denotes the scale. Since more and more fine scale structures are removed at each increasing scale, the scale space decomposition is expressed in the recursion relation

$$L_{i-1}(s) = L_i(s) + C_i(s) \text{ for } i = 1, \ldots, M. \tag{2}$$

Starting with $L_0 = I(s)$, we formulate the computation of each coarser scale L_i as an inverse problem, where $L_{i-1}(s)$, $L_i(s)$, and $C_i(s)$ are the measurement, state, and noise respectively. The Bayesian least-squares estimate for this problem is given by

$$\hat{L}_i(s) = E[L_i(s)|L_{i-1}(s)] \tag{3}$$

$$= \int_0^\infty p\big(L_i(s)\big)p\big(L_i(s)|L_{i-1}(s)\big)\, dL_i(s)\,. \tag{4}$$

Since computing the posterior distribution $p\big(L_i(s)|L_{i-1}(s)\big)$ is analytically intractable, a quasi-random estimation approach is employed. In this approach, n quasi-random samples from a Sobol sequence [19] are drawn with respect to site s at scale i. A Gaussian mixture model is fitted to $p(L_{i-1}(s))$ and those samples which fall within one standard deviation of the nearest local maximum of $p\big(L_{i-1}(s)\big)$ are selected as a realizable sample of $p\big(L_i(s)|L_{i-1}(s)\big)$. From the set Ω of selected samples, the posterior distribution is estimated as

$$\hat{p}(L_i(s)|L_{i-1}(s)) = \frac{1}{G\sqrt{2\pi}\sigma_{L_i}} \sum_{k\in\Omega} f_1(k)f_2(k)f_3(k) \exp\left(-\frac{1}{2}\left(\frac{L_i(s) - L_{i-1}(s_k)}{\sigma_{L_{i-1}}}\right)^2\right), \tag{5}$$

where G is a normalization factor, and where $f_1(k)$, $f_2(k)$, $f_3(k)$ are objective functions of sample relevance assessed by intensity, gradient, and spatial offset respectively [13]. This completes the computation of the scale-space estimates via the method of Mishra et al. [13].

Next, given the scale space estimates $L_i(s)$, we suppress the modality-specific intensity information and capture the structural characteristics by computing the discrete derivative magnitude of $|\nabla_i|^2(s)$ at each scale,

$$|\nabla_i|^2(s) = \left(\frac{\partial L_i}{\partial x}\right)^2 + \left(\frac{\partial L_i}{\partial y}\right)^2. \tag{6}$$

Taking advantage of the fact that salient structural features have strong responses across multiple scales [20], the quasi-random scale space structural representation of the image is given by

$$Q(s) = \left[\sum_{i=1}^{M} \alpha^i |\nabla_i|^2(s)\right]^{\frac{1}{2}}, \tag{7}$$

where the response at each scale is weighted by α^i to emphasize coarser scales as a way to suppress noise. Based on testing, $\alpha = 2$ was found to provide the best results.

After computing the quasi-random scale space structural representations, Q_f and Q_g for the image pair f and g, the quasi-random scale space structural correlation (QRSC) energy functional C can be defined as

$$C(T[f], g) = \sum_{\tau} Q_f(T[\tau])Q_g(\tau). \tag{8}$$

A. Efficient Globally Exhaustive Optimization in the Frequency Domain

In this section, we now develop an efficient globally exhaustive optimization framework in the frequency domain based on the QRSC energy functional C introduced in Eq. 8. Suppose we wish to find the rigid transformation that, when applied to Q_f, maximizes the energy functional in Eq. 8, according to Eq. 1. The cross-correlation energy functional can be efficiently maximized using the Fast Fourier Transform (FFT) [8], [9], hence making this approach a good direction to follow for the proposed optimization framework. Given two images f and g, their quasi-random scale space structural correlation energy functional at all integer shifts Δx is found by

$$C_{f,g}(\underline{x}) = \sum_{\tau} Q_f(\tau - \Delta x) Q_g(\tau) \tag{9}$$

$$= \mathcal{F}^{-1}\left(\overline{\mathcal{F}\{Q_f\}}\mathcal{F}\{Q_g\}\right), \tag{10}$$

where $\mathcal{F}$ is the FFT operation and $\overline{\mathcal{F}\{Q\}}$ denotes the complex conjugate. Hence, for the space of possible rigid transformations, we can determine the maximum quasi-random scale space structural correlation as a function of rotation θ, and translation $\Delta x = [x, y]$ by decoupling the the rotational and translational components.

First, let F and G be the Fourier coefficients of Q_f and Q_g respectively. Since a rotation in the spatial domain is equivalent to the same rotation in the frequency domain, we would like to find the rotation θ^* that, when applied to F, gives the greatest correlation with G.

To do this, we transform F and G into polar coordinates, expressed as $F_{pol}(\theta, r)$ and $G_{pol}(\theta, r)$. To find the optimal shift in the θ axis, we maximize the correlation between the magnitudes $|F_{pol}|$ and $|G_{pol}|$ according to

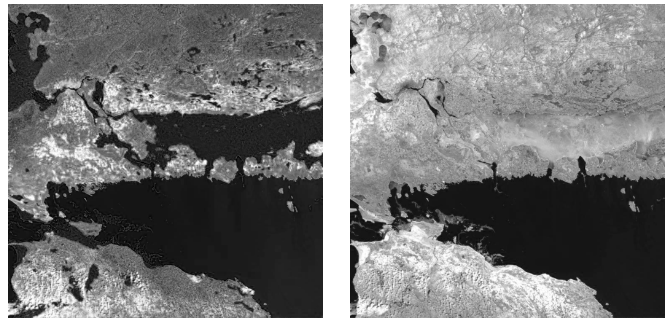
$$\{\theta^*, r^*\} = \arg\max_{\theta, r}\{C_{|F_{pol}|,|G_{pol}|}(r, \theta)\}, \tag{11}$$

where $C_{|F_{pol}|,|G_{pol}|}(r, \theta)$ can be computed for all possible values of θ and r in a simultaneous manner according to Eq. 10. Note that while both θ^* and r^* are found, only θ^* is important for our purposes as the goal is to determine the optimal rotation that brings the images into alignment. Furthermore, due to the conjugate symmetry of the FFTs of real-valued images Q_f and Q_g, both θ^* and $\theta^* + 180°$ are possible maxima.

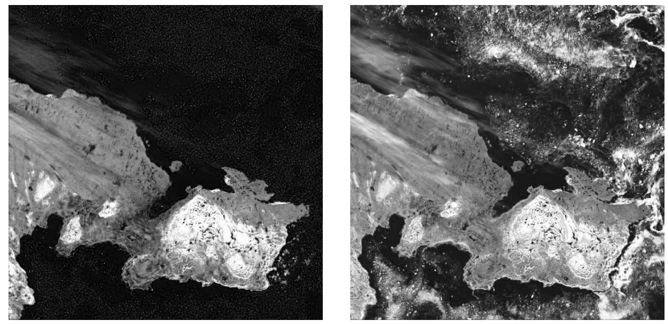
Next, we rotate Q_f by θ^* to obtain Q'_f. Now, we compute the optimal translation that brings Q'_f into alignment with Q_g by maximizing the cross correlation

$$\{x^*, y^*\} = \arg\max_{x^*, y^*}\{C_{Q'_f, Q_g}(x, y)\}. \tag{12}$$

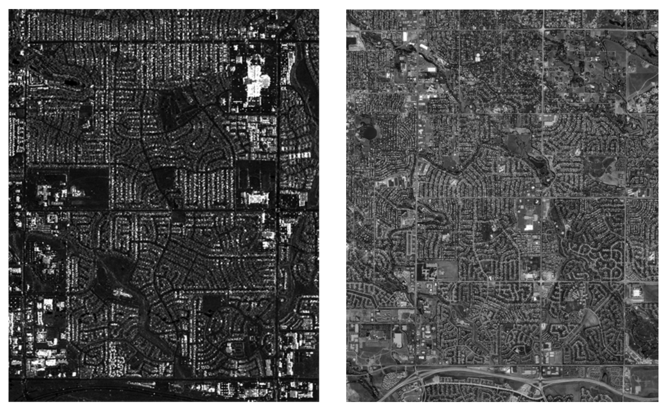
Again, the quasi-random scale space structural correlation energy functional $C_{Q'_f, Q_g}(x, y)$ can be computed globally for all possible translations in a simultaneous manner using Eq. 10. To deal with the fact that both θ^* and $\theta^* + 180°$ are possible maxima, this process is repeated for the case of $\theta^* + 180°$, and the rotation-translation combination that gives the highest maximum value is chosen.



RS1: Interband test image pair: Band 3 and Band 5



RS2: Intraband test image pair: 2002/7/26 and 2002/7/17



RS3: Intermodal test image pair: LIDAR and optical

Fig. 2. Remote sensing test image pairs

III. Results

To evaluate the performance of the proposed QRSC approach, automatic registration was performed on both interband and intraband remote sensing image sets from the United States Geological Survey (USGS), as well as optical-LIDAR image sets from Intermap Technologies Inc. The first test pair, RS1, consists of two images taken using different sensors on different bands. The second test pair, RS2, consists of two images taken by the same sensor on the same band on different dates. The third test pair, RS3, consists of an aerial optical image and a LIDAR image, and as such are acquired using different sensing technologies:

1) **RS1**: Interband, Landsat-7 ETM+ **Band 3** (GSD: 240m) and Landsat 5 TM **Band 5** (GSD: 240m), different dates.
2) **RS2**: Intraband, Landsat-7 ETM+ Band 3 (GSD: 240m), taken on **2002/7/26** and **2002/7/17**.

3) **RS3**: Intermodal, aerial passive optical (GSD: 1m) and LIDAR (GSD: 1m)

The images in RS1 and RS2 measure 740×740 pixels while the optical and LIDAR images in RS3 measure 902×1131 and 449×567 respectively, and are shown in Fig. 2. Each image was distorted with 30 random rigid transformations, with translations up to 50 pixels for RS1 and RS2, translations up to 200 pixels for RS3, and rotations between $0°$ and $360°$ for all cases, giving us a total of 90 randomized tests. Since the images are initially aligned, the gold standard transformations are known for all tests.

For comparison, we also performed registration with GO-EDGE [9], a state-of-the-art multimodal globally-optimal FFT-based approach that was shown to provide superior performance when compared to correlation ratio based methods for large misalignments [9], and with normalized mutual information [4] maximized with the Nelder-Mead simplex method [7] as described in [2]. For QRSC, a total of $M = 5$ scales was used to represent the structural characteristics of the images. The registration accuracy is determined by the fiducial registration error (FRE), defined as the root mean square distance between the fiducial points.

Since the images were originally aligned and the gold standard transformations are known for all tests, a set of 60 fiducial points were randomly placed by the computer to allow for a fair evaluation amongst different techniques, since human placement of fiducial points could be prone to error in this particular case and tend to be biased towards structured landmarks such as road intersections and buildings and as such may give an advantage to techniques that make use of structure information. The choice of 60 fiducial points was considered sufficient as they are based on unbiased computer placement and are well spread out throughout the images.

Table I summarizes the results for the remote sensing test pairs. In all three cases, QRSC has the greatest success rate, resulting in the lowest average registration errors. Sample results using QRSC are illustrated in Fig. 3 with the structural representation of the left images overlaid on the transformed images. Given the large number of randomized tests conducted, a discussion of the dependence on the amount of misalignment is important. The effect of different levels of misalignment (defined here in terms of pixel displacement from the true alignment) on the average FRE across all test pairs was studied, as shown in Fig. 4. The FRE obtained using the QRSC approach remained relatively stable as the amount of misalignment increased, while the FRE obtained using the other approaches increased significantly as the amount of misalignment increased. This weak dependence of FRE on the amount of misalignment is due to the fact that the proposed approach employs an efficient globally exhaustive optimization framework in the frequency domain based on the QRSC energy functional, which is highly robust to the presence of local optima that can have a tremendous effect on registration performance for situations characterized by large misalignments between the images.

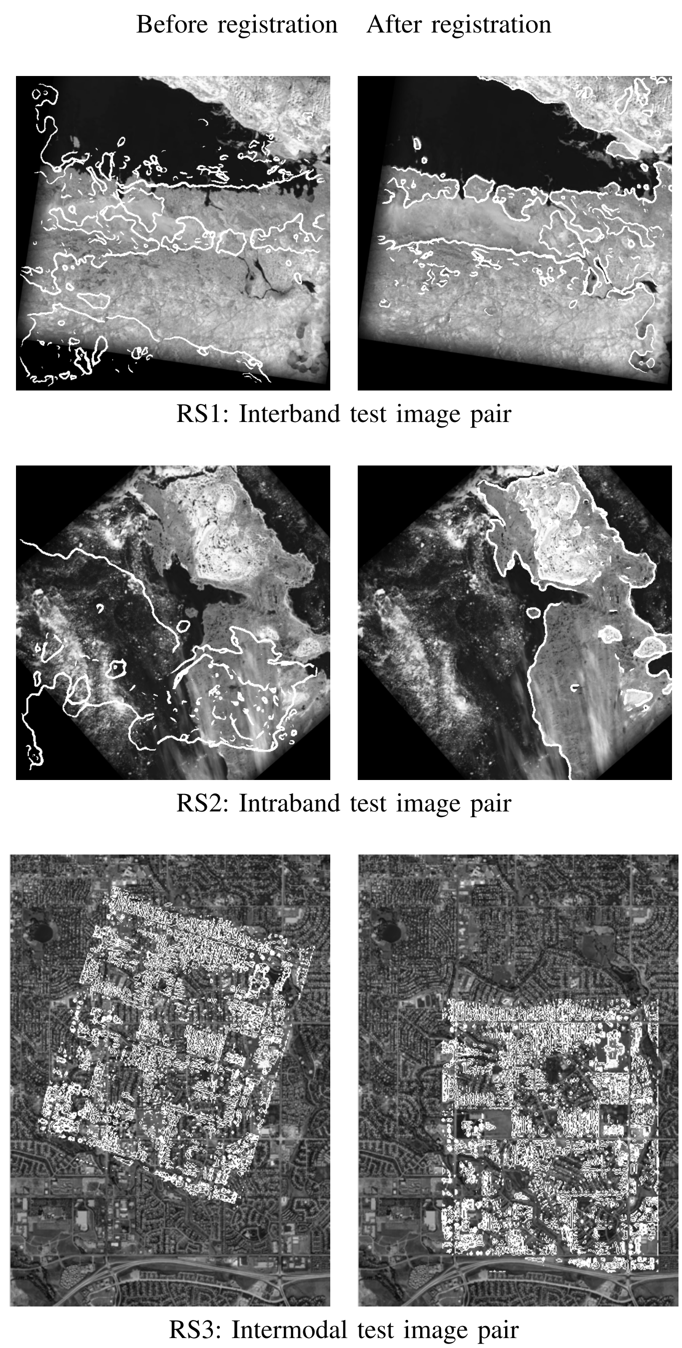


Fig. 3. Example of QRSC registration result for remote sensing image pairs. The overlay shows the structural representation of the Band 3 image in the (*top row*), the 2002/7/26 image in the (*middle row*), and the LIDAR image in the (*bottom row*). Only a zoomed-in section of the results for RS3 is shown for illustrative purposes to improve clarity.

TABLE I
FIDUCIAL REGISTRATION ERRORS (FRE) FOR REMOTE SENSING TEST PAIRS. A TOTAL 30 RANDOMIZED TESTS WERE PERFORMED FOR EACH TEST CASE.

Test	**FRE in pixels [mean / stddev]**		
	QRSC	GO-EDGE	NMI
RS1	6.26/1.22	442/321	495/277
RS2	1.35/0.87	405/272	464/179
RS3	4.98/1.72	342/114	459/125

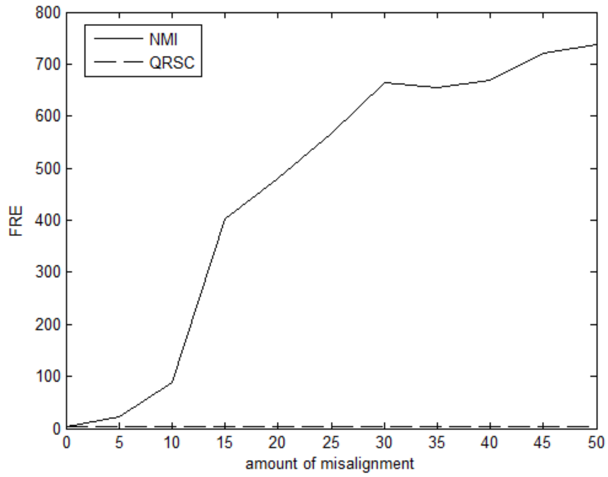


Fig. 4. The effect of different amounts of misalignment on the average FRE across all test pairs. The FRE obtained using the QRSC approach remained relatively stable as the amount of misalignment increased.

IV. Conclusion

In this paper, we presented a novel quasi-random scale space structural correlation energy functional for registering remote sensing images acquired under different sensing conditions. An efficient globally exhaustive optimization framework in the frequency domain was also introduced based on the proposed energy functional. It offers advantages in being suitable for registering remote sensing images acquired at different times, using different sensors, and/or at different bands, particularly under large misalignments. Based on the results, this approach shows great promise in allowing the use of efficient globally exhaustive optimization methods for robust registration of remote sensing imagery. Future work involves investigating the design and incorporation of a non-rigid registration component to the existing framework based on this energy functional.

Acknowledgment

This research was sponsored by Geomatics for Informed Decisions (GEOIDE) and the Natural Sciences and Engineering Research Council (NSERC) of Canada through individual Discovery Grants, as well as the United States Geological Survey and Intermap Technologies Inc. for the remote sensing imagery.

References

[1] J. Kern and M. Pattichis, “Robust multispectral image registration using mutual-information models,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 5, pp. 1494–1505, 2007.

[2] P. V. H. Chen, M. Arora, “Mutual information-based image registration for remote sensing data,” *International Journal of Remote Sensing*, vol. 24, no. 18, pp. 3701–3706, 2003.

[3] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, “Multi-modality image registration by maximization of mutual information,” *Proceedings of the Workshop on Mathematical Methods in Biomedical Image Analysis*, 1996.

[4] C. Studholme, D. L. G. Hill, and D. J. Hawkes, “An overlap invariant entropy measure of 3D medical image alignment,” *Pattern Recognition*, vol. 32, no. 1, pp. 71 – 86, 1999.

[5] A. Roche, G. Malandain, X. Pennec, and N. Ayache, “The correlation ratio as a new similarity measure for multimodal image registration,” *Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention*, vol. 1496, pp. 1115–1124, 1998.

[6] M. Hestenes and E. Stiefel, “Methods of conjugate gradients for solving linear systems,” *J. Res. Nat. Bureau Stand.*, vol. 49, pp. 409–436, 1952.

[7] J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright, “Convergence properties of the Nelder–Mead simplex method in low dimensions,” *SIAM Journal on Optimization*, vol. 9, no. 1, pp. 112–147, 1998.

[8] B. Reddy and B. Chatterji, “An FFT-based technique for translation, rotation, and scale-invariant image registration,” *IEEE Transactions on Image Processing*, 1996.

[9] J. Orchard, “Globally optimal multimodal rigid registration: An analytic solution using edge information,” *IEEE International Conference on Image Processing, 2007. ICIP 2007.*, 2007.

[10] H. Foroosh, J. Zerubia, and M. Berthod, “Extension of phase correlation to subpixel registration,” *IEEE Transactions on Image Processing*, 2002.

[11] M. Guizar-Sicairos, S. T. Thurman, and J. R. Fienup, “Efficient subpixel image registration algorithms,” *Optics Letters*, vol. 33, no. 2, pp. 156–158, 2008.

[12] R. Li, J. Sun, F. Yin, and F. Guo, “An automatic registration method for different temporal remote sensing images based on improved fourier-mellin algorithm,” *Proceedings of the SPIE*, vol. 7651, 2009.

[13] A. Mishra, A. Wong, D. A. Clausi, and P. W. Fieguth, “Quasi-random nonlinear scale space,” *Pattern Recognition Letters*, vol. 31, no. 13, pp. 1850–1859, 2010.

[14] J. Maintz, E. Meijering, and M. Viergever, “General multimodal elastic registration based on mutual information,” *Proceedings of the SPIE Image Processing*, vol. 3338, pp. 144–154, 1998.

[15] B. Likar and F. Pernus, “A hierarchical approach to elastic registration based on mutual information,” *Image and Vision Computing*, vol. 19, pp. 33–44, 2001.

[16] T. Lindeberg, *Scale-Space Theory in Computer Vision*. Dordrecht: Kluwer Academic, 1994.

[17] A. Witkin, “Scale-space filtering,” *Proceedings of the 7th International Joint Conference on Artificial Intelligence*, pp. 1019–1022, 1983.

[18] P. Perona and J. Malik, “Scale-space and edge detection using anisotropic diffusion,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1990.

[19] I. M. Sobol, “Uniformly distributed sequences with an additional uniform property,” *U.S.S.R. Comput. Maths. Math. Phys.*, vol. 16, pp. 236–242, 1976.

[20] D. J. Field and N. Brady, “Visual sensitivity, blur and the sources of variability in the amplitude spectra of natural scenes,” *Vision Research*, vol. 37, no. 23, pp. 3367 – 3383, 1997.