

© © 2016 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Title: Edge-Based Registration Noise Estimation in VHR Multitemporal and Multisensor Images

This paper appears in: IEEE Geoscience and Remote Sensing Letters

Date of Publication: 28 June 2016

Author(s): Y. Han, F. Bovolo, L. Bruzzone,

Volume: 13, Issue: 9

Page(s): 1231 - 1235

DOI: 10.1109/LGRS.2016.2577719

EDGE-BASED REGISTRATION NOISE ESTIMATION IN VHR MULTITEMPORAL AND MULTISENSOR IMAGES

Youkyung Han¹, *Member, IEEE*, Francesca Bovolo¹, *Senior Member, IEEE*, and Lorenzo Bruzzone²,
Fellow, IEEE

¹Fondazione Bruno Kessler, Center for Information and Communication Technology, Trento, Italy

²Dept. of Information Engineering and Computer Science, University of Trento, Trento, Italy

Abstract— Even after co-registration, Very High Resolution (VHR) multitemporal images acquired by different multispectral sensors (e.g., QuickBird, WorldView) show a residual misregistration due to dissimilarities in acquisition conditions and in sensor properties. Residual misregistration can be considered as a source of noise and is referred to as Registration Noise (RN). Since RN is likely to have a negative impact on multitemporal information extraction, detecting and reducing it can increase multitemporal image processing accuracy. In this paper, we propose an approach to identify RN between VHR multitemporal and multisensor images. Under the assumption that dominant RN mainly exists along boundaries of objects, we propose to use edge information in high frequency regions to estimate it. This choice makes RN detection less dependent on radiometric differences and thus more effective in VHR multisensor image processing. In order to validate the effectiveness of the proposed approach, multitemporal multisensor datasets are built including QuickBird and WorldView VHR images. Both qualitative and quantitative assessments demonstrate the effectiveness of the proposed RN identification approach compared to the state-of-the-art one.

Index Terms— Registration noise, very high resolution images, multitemporal multisensor data, co-registration

I. INTRODUCTION

Since Earth observation satellites equipped with Very High Resolution (VHR) sensors (e.g., QuickBird, GeoEye, WorldView, etc.) have been launched, the availability of images with a spatial resolution less than 1m has increased. As satellites periodically acquire images on the Earth surface, multitemporal images acquired from various sensors play a crucial role in a wide range of applications. However, in order to effectively exploit multitemporal images, they should be geometrically aligned each other. Therefore, image co-registration, which is defined as the process of spatially overlaying two or more images of the same scene [1], is a prerequisite. When multitemporal images show a high co-registration level, a better performance in multitemporal information extraction is possible.

Many studies exist on VHR multitemporal images co-registration [2]–[4]. However, even after co-registration, multitemporal images still show a residual misregistration caused by dissimilarities in the acquisition conditions such as stability of the acquisition platform, off-nadir acquisition angles of the sensor, geometrical properties of the considered scene, and so on. The most critical component of misregistration arises from spatially corresponding pixels belonging to different objects at two dates (i.e., the border region of objects or the high frequency area in the images) [5]. Residual misregistration can be considered as a source of noise, referred as Registration Noise (RN) [5]–[7], and it has a negative impact on multitemporal information extraction. For example, it reduces the accuracy of change detection with an increase of false alarms [8], and of multitemporal data fusion or segmentation [9] as well. Accordingly, a proper detection of RN can improve the multitemporal image processing performance. To the authors' knowledge only one study exists on the analysis of RN properties and its identification [5]. The approach is based on Change Vector Analysis (CVA) in polar domain applied to spectral bands, and assumes that multitemporal images show similar radiometric behaviors. Thus radiometric differences can be associated to real changes on the ground or misregistration effects, only. This is reasonable when multitemporal single sensor VHR images are considered. However, the assumption is seldom satisfied in multisensor multitemporal datasets where radiometric properties are intrinsic due to the acquisition sensors.

In this paper, we aim at designing an automatic method for identifying RN in VHR multitemporal and multisensor images. The proposed method is based on the assumption that RN appears as thin linear features along object borders. We thus work on edge magnitude images (instead of spectral bands) generated by applying the Difference of Gaussian (DoG) filter. The high edge magnitude pixels on both images are first extracted, after which the pixels not involved in border regions of objects are estimated as RN pixels by using the difference of edge magnitude between the two images. In order to assess the proposed approach, QuickBird and WorldView VHR multitemporal multisensor data are employed. RN extraction results by the proposed approach are compared with the ones obtained according to [5].

This letter is organized into five sections. Section II explains background of the CVA-based RN extraction approach and drawbacks when it is applied to VHR multitemporal images acquired by different sensors. Section III describes the proposed RN extraction approach. Section IV illustrates the dataset and reports experimental results. Section VI presents the conclusion of this letter.

II. CVA-BASED REGISTRATION NOISE IDENTIFICATION

In [5], RN pixels are identified in multitemporal images acquired by the same VHR sensor by means of CVA in polar domain applied to spectral bands. Let X_1 and X_2 be two VHR multispectral images acquired over the same area at different times t_1 and t_2 , respectively. The multispectral difference image X_Δ consists of Spectral Change Vectors (SCVs) defined as

$$X_\Delta = X_2 - X_1 \quad (1)$$

In a 2D feature space, the change information can be described by the magnitude ρ and the direction ϑ of SCVs:

$$\rho = \sqrt{(X_{\Delta,1})^2 + (X_{\Delta,2})^2}, \rho \in [0, \rho_{max}] \quad (2)$$

$$\vartheta = \tan^{-1} \left(\frac{X_{\Delta,1}}{X_{\Delta,2}} \right), \vartheta \in [0, 2\pi) \quad (3)$$

where $X_{\Delta,b}$ represents the b th spectral band of X_{Δ} ($b = \{1,2\}$) and ρ_{max} is the highest magnitude value of SCVs in the image.

In the polar domain, unchanged SCVs have a low magnitude and cluster around the origin, whereas changed SCVs have a high magnitude and cluster far from the origin [10]. Therefore, the magnitude variable can be used to distinguish between change and no-change SCVs according to

$$X_{\Delta}(x, y) \in \begin{cases} \text{change,} & \text{if } \rho(x, y) \geq T \\ \text{no change,} & \text{otherwise} \end{cases} \quad (4)$$

where T is a threshold value.

In this feature space, dominant RN pixels behave like changed pixels, they are likely to show magnitude above T and to be detected as changes [5]. However, [5] demonstrated that changed samples have a stable behavior along the direction variable when scale decreases, whereas the dominant RN noise ones do not. Thus, a multiscale analysis of the direction variable allows one to discriminate between RN and changed pixels. Let us assume that (X_1^N, X_2^N) is a low-resolution version of the original (X_1, X_2) image pair. The conditional density of RN distribution along the direction variable $\hat{p}^{RN}(\vartheta|\rho \geq T)$ can be estimated as:

$$\hat{p}^{RN}(\vartheta|\rho \geq T) = C[P(\rho \geq T)\hat{p}(\vartheta|\rho \geq T) - P^N(\rho \geq T)\hat{p}^N(\vartheta|\rho \geq T)] \quad (5)$$

where $P(\rho \geq T)$ and $P^N(\rho \geq T)$ are the probabilities of SCVs having magnitude values higher than T at full resolution and at resolution level N , respectively; $\hat{p}(\vartheta|\rho \geq T)$ and $\hat{p}^N(\vartheta|\rho \geq T)$ are the marginal conditional densities of the direction variable of the SCVs at full resolution and at resolution N , respectively; and C is a constant value that satisfies the condition $\int_0^{2\pi} \hat{p}^{RN}(\vartheta|\rho \geq T)d\vartheta = 1$. Higher value of $\hat{p}^{RN}(\vartheta|\rho \geq T)$ (i.e., high difference between distributions at low and high resolution) corresponds to a higher probability of having pixels contaminated by RN. Detailed explanation of CVA-based RN identification can be found in [5].

The CVA-based RN detection approach shows some open issues when it is applied to multisensor multitemporal image pairs. The first open issue is a difficulty in separating between change and no-change

pixels using magnitude variable. This is critical because RN pixels are separated from the changed SCVs under the assumption that the latter ones are effectively identified first. However, because of sensor-dependent differences, multitemporal multisensor image pairs may show stronger radiometric dissimilarities compared to single sensor multitemporal pairs even if no changes occurred. Accordingly, unchanged SCVs show higher magnitude in average and tend to be clustered in biased position far from the origin. Thus, their behavior in the polar domain becomes similar to that of both changed and RN pixels. In this situation, a proper detection of unchanged samples would require a high value of T , which implies an increase of missed alarms and a decrease of accuracy in detecting changed pixels and thus of RN. On the other hand, a lower T value would imply that a larger number of unchanged pixels is identified as being changed. Accordingly, RN pixels would be over-estimated. In both cases RN detection step in multisensor multitemporal image pairs is less accurate than in single sensor ones.

The second open issue is whether common spectral bands between multitemporal multisensor images exist or not. The approach in [5] assumes that at least two common spectral bands exist in the multitemporal image pair. However, when multisensor images are considered, the assumption may not be satisfied. Thus, CVA on spectral bands is less effective in RN detection. Moreover, even if similar acquisition bands exist, the selection of bands leading to good performance may be complicated due to the variability of SCVs distribution in the different spectral bands.

In order to mitigate the abovementioned issues arising from radiometric and spectral dissimilarities that affect multitemporal multisensor images, we propose an approach to RN identification for multisensor VHR images based on the use of geometric features.

III. PROPOSED EDGE-BASED REGISTRATION NOISE IDENTIFICATION

As mentioned, RN mainly occurs along boundaries among objects. It means that RN has a high probability to be detected in high frequency content regions. Therefore, we propose to use high frequency features as edge magnitude image to extract RN. By working on edge magnitude images independently generated from

each multitemporal image, we can minimize the impact of the radiometric dissimilarities between multisensor images. Let us assume that X_1^h and X_2^h are the edge magnitude images at time t_1 and t_2 . Here we exploit the DoG filter to construct the edge magnitude image because of its reliability on scale and rotation differences [11]. The edge magnitude image generated by DoG filter is defined as

$$X_i^h = \frac{\sum_{b=1}^B (G_{k\sigma}(X_{i,b}) - G_{\sigma}(X_{i,b}))}{B} \quad (6)$$

where i ($i = \{1,2\}$) is the acquisition time t_i , and G_{σ} is the Gaussian filter with standard deviation σ . k is a constant multiplicative factor. The values of σ and k control the thickness and smoothness of edges in the edge magnitude image. Small values of σ and k allow to preserve detail edges, but fail in removing small linear noisy segments. On the other hand, large values effectively reduce noise, but also remove detail edge information associated with RN. Therefore, selecting suitable values of σ and k is important to detect RN in a reliable way. B is the number of similar bands b ($b = \{1, \dots, B\}$) between multitemporal multisensor images. The magnitude image can be generated even if there is one common band (i.e., $B = 1$) only.

In order to extract the RN on the edge magnitude images, two conditions should be satisfied. The first condition is that RN occurs in the surrounding of dominant edges in both images when the dominant edges are misaligned. Pixels located along and nearby the dominant edges have larger edge magnitude values, and the values decrease while moving toward homogeneous regions. Thus, candidate RN pixels are the ones having high edge magnitude in both images, i.e.,

$$RN_1(x, y) \in \begin{cases} 1, & \text{if } \min(|X_1^h(x, y)|, \alpha|X_2^h(x, y)|) \geq T_1 \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

where T_1 is a threshold value. The higher T_1 , the more strict the selection criteria. (x, y) is a generic spatial position of image, and α is a normalization parameter that controls the balance of the edge magnitude between two multisensor images. Since we suppose that X_1 and X_2 have radiometric dissimilarities, the edge magnitude images are also different. The impact on relative edge magnitude can be normalized and thus

compared each other by dividing the magnitude values by their standard deviations [3]. Accordingly, the normalization parameter can be derived as $\alpha = \sigma(X_1^h)/\sigma(X_2^h)$, where $\sigma(X_1^h)$ and $\sigma(X_2^h)$ are standard deviations of each edge magnitude image X_1^h and X_2^h , respectively.

Recalling that object boundaries have strong edge magnitude in both images, candidate RN pixels estimated by (7) can be: i) precisely co-registered border regions; or ii) RN pixels. To discriminate between the two options, similarly to [5], the difference image is considered, but computed on the edge magnitude feature. The edge magnitude difference is small for precisely aligned edges and becomes larger and larger as misregistration increases. This is because high edge magnitude regions (i.e., dominant edges) in one image are compared to lower edge magnitude regions (i.e., regions in the neighborhood of dominant edges) in the other one. Accordingly, the second condition RN_2 becomes

$$RN_2(x, y) \in \begin{cases} 1, & \text{if } \|X_1^h(x, y) - \alpha X_2^h(x, y)\| \geq T_2 \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

where T_2 is a threshold value that controls the sensitivity to the amount of misregistration. $\alpha = \sigma(X_1^h)/\sigma(X_2^h)$ is the normalization factor.

The final RN map is identified as simultaneously satisfying conditions (7) and (8) as

$$RN(x, y) = RN_1(x, y)RN_2(x, y) \quad (9)$$

IV. EXPERIMENTS AND RESULTS

A. Dataset description

To evaluate the effectiveness of the proposed RN identification approach, we employed multitemporal multisensor images acquired over the city of Trento (Italy). The QuickBird and WorldView satellite multispectral sensors were considered. The QuickBird image was acquired in July 2006 with 14.1° off-nadir angle. It consists of a panchromatic band with 0.6m spatial resolution and four spectral bands [blue (450-520nm), green (520-600nm), red (630-690nm), and NIR (760-900nm)] with 2.4m spatial resolution. The WorldView image was acquired in August 2010 with 12.0° off-nadir angle. It consists of a panchromatic

band with 0.5m spatial resolution and eight spectral bands [coastal (400-450nm), blue (450-510nm), green (510-580nm), yellow (585-625nm), red (630-690nm), red edge (705-745nm), NIR 1 (770-895nm), and NIR 2 (860-1040nm)] with 2.0m spatial resolution.

The QuickBird (X_1) and WorldView (X_2) images were pansharpened by the Gram-Schmidt method [12], and 2 subsets of 980×980 pixels were selected for the experiments. The datasets show both urban and field areas and were preliminary co-registered in different ways. Automatic co-registration was applied to the first one. Tie-points were extracted based on Scale Invariant Feature Transform (SIFT) approach [3], and an affine transformation was estimated by applying the RANdom SAmples Consensus (RANSAC) on the tie-points for warping the WorldView image (X_2) to QuickBird one (X_1) [13]. For the second dataset, we manually selected tie-points and applied the polynomial transformation for warping. For both of them residual misregistration was around 1 RMSE pixel on the tie-points, but larger misalignment can be observed locally. Fig. 1 shows a true color composition of the two multitemporal multisensor datasets.

B. Experimental Results

To assess the effectiveness of proposed RN identification, we applied the proposed approach and compared the result with the one proposed in [5] (standard approach). The values of parameters have been selected based on trials. More specifically, we exploited the red and NIR bands of the VHR images to identify RN for the fair comparison with the standard approach. k and σ were set to 2 and 1.6, respectively. Threshold values T_1 , T_2 and T were automatically selected by applying a Bayesian decision rule [14]. The selected values of T_1 , T_2 and T are 10.2, 20.5 and 282.1 for the first dataset, and 10.9, 23.0 and 208.1 for the second one, respectively.

Qualitative assessment was conducted by comparing RN maps extracted according to the proposed approach and the standard one [5] (Fig. 2). As one can see, the use of spectral bands leads to the identification of RN pixels in some regions that did not change. This is because they have linear behaviors and large radiometric differences due to the different acquisition sensors. This is the case of roads and some plant rows in the fields.

Examples of such situations can be observed in Fig. 3.a and 4.a. As expected radiometric differences in multisensor images result in poor performance on the discrimination between changed and no-changed SCVs. The increment of false alarms thus increases the amount of detected RN pixels. Those effects are significantly mitigated by the proposed approach (Fig. 3.b and 4.b).

For further assessment, fine co-registration was applied according to [15]. Fine co-registration performance depends on the accuracy of RN. Thus we can say that the higher the fine co-registration performance, the better the RN noise detection. For visual assessment, chessboard images generated by the CVA-based and the proposed approaches are shown in Fig. 3 and Fig. 4. From border regions of blocks in the chessboard image, we can see whether the co-registration is carried out properly or not. As one can see from the alignment quality in critical areas (white circles in Fig. 3 and Fig. 4), the proposed approach (Fig. 3.d and Fig. 4.d) led to a better co-registration of the multitemporal images with respect to the standard one (Fig. 3.c and Fig. 4.c).

Quantitative assessment was conducted by analyzing fine co-registration performance on: i) local regions showing large radiometric dissimilarity, and ii) whole dataset, independently. Three indices were selected: Mutual Information (MI), RMSE and Standard Deviation (SD) of the residual errors. MI evaluates the similarity between multitemporal multisensor images. The higher the MI value, the better the performance. To compute the RMSE and the SD, 40 checkpoints were selected. 12 checkpoints (see red crosses in Fig 3 and Fig 4) were selected in regions with radiometric dissimilarities, and 28 checkpoints were evenly sampled over the whole dataset. The performance is presented in Table I and Table II for the first and second datasets, respectively. The analysis of fine co-registration performance over 28 evenly sampled checkpoints (i.e., global columns in Table I and Table II) points out that both the proposed and the standard RN approaches introduce significant improvements in the co-registration accuracy with respect to methods that does not account for RN information (i.e., SIFT-based alignment for the first dataset, and manual alignment for the second one). However, the proposed RN detection approach guarantees a more significant improvement. Let

us now analyze performance at a local level in critical checkpoints affected by radiometric differences. Here, the proposed approach generates an RMSE around 1 pixel, whereas the standard RN detection approach fails and RMSE increases up to 3.49 pixels in first dataset and 3.32 pixels in second one. The standard RN approach could not improve the fine co-registration performance locally. Whereas, the proposed one is more effective and allows for a significant reduction of RMSE from 2.47 to 1.03 in first dataset and from 2.53 to 1.01 in second dataset. This is further confirmed by the lower SD values.

TABLE I

NUMERICAL ASSESSMENT OF CO-REGISTRATION ACCURACY

(FIRST DATASET).

Registration method	Local			Global		
	MI	RMSE	SD	MI	RMSE	SD
SIFT approach	0.562	2.47	1.31	0.514	2.29	1.16
Based on standard RN detection	0.538	3.49	1.84	0.521	1.86	0.98
Based on proposed RN detection	0.577	1.03	0.43	0.538	1.05	0.50

TABLE II

NUMERICAL ASSESSMENT OF CO-REGISTRATION ACCURACY

(SECOND DATASET).

Registration method	Local			Global		
	MI	RMSE	SD	MI	RMSE	SD
Manual alignment	0.614	2.53	1.03	0.462	2.93	1.23
Based on standard RN detection	0.622	3.32	1.82	0.485	1.90	0.90
Based on proposed RN detection	0.688	1.01	0.40	0.529	0.92	0.32

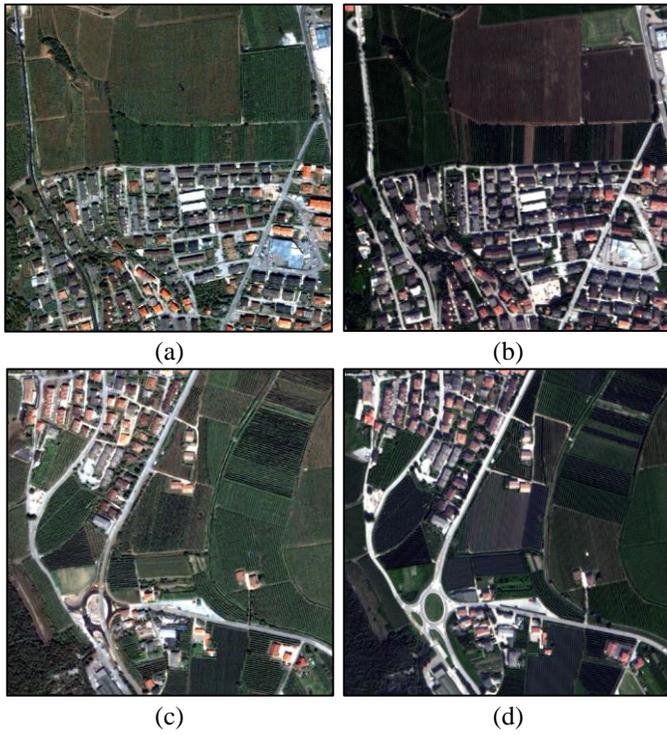


Fig. 1. True color composition of first dataset: (a) Quickbird image (X_1) and (b) WorldView image (X_2); and second dataset: (c) QuickBird image (X_1) and (d) WorldView image (X_2).



Fig. 2. RN noise maps obtained with the standard (left) and the proposed (right) approach on: (a), (b) the first dataset; and (c), (d) the second dataset.

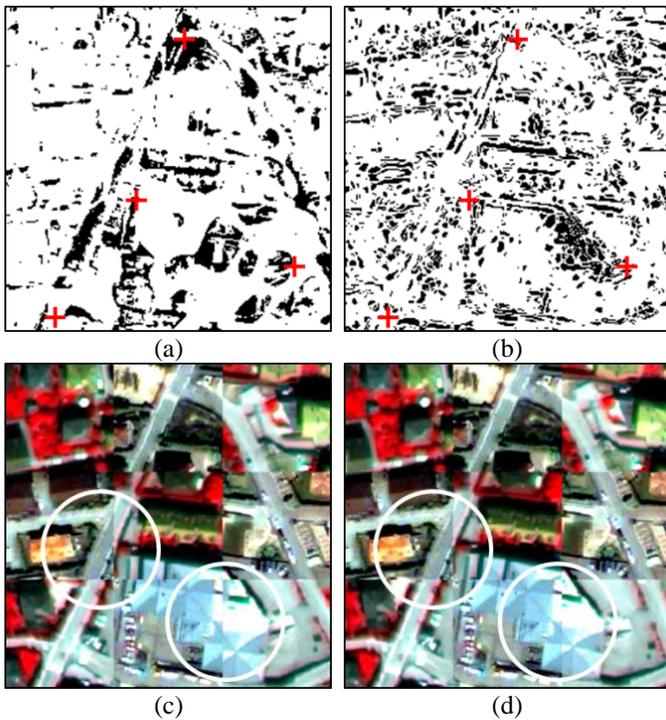


Fig. 3. Detail of RN maps and chessboard images generated with the standard approach (a), (c); and the proposed one (b) and (d). (first dataset).

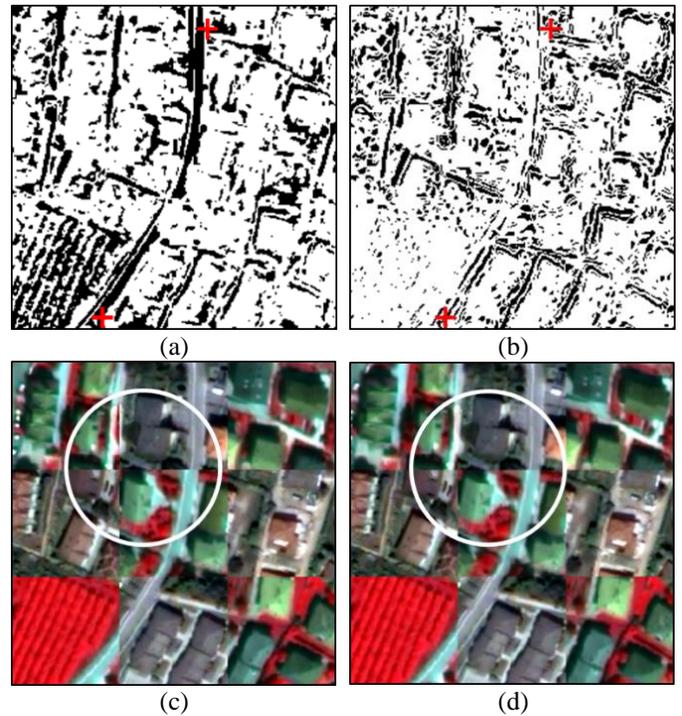


Fig. 4. Detail of RN maps and chessboard images generated with the standard approach (a), (c); and the proposed one (b) and (d). (second dataset).

V. CONCLUSION

In this paper, we proposed an approach to identify the RN between VHR multitemporal multisensor images based on the edge magnitude. The use of edge information makes the proposed approach robust to radiometric differences and thus particularly suitable when multitemporal multisensor datasets are considered that show radiometric differences induced by sensor differences. On the multisensor multitemporal datasets, the proposed approach visually and quantitatively improved the results compared to the standard RN identification approach, which is penalized by intrinsic radiometric differences. As future work, we plan to further mitigate the impact of radiometric differences in multitemporal multisensor VHR images by combining the proposed RN identification approach with a change detection approach designed for such kind of data.

VI. ACKNOWLEDGEMENT

The authors are thankful to Digital Globe Foundation for providing WorldView image within the project “MS-TS – Analysis of Multisensor VHR image Time Series”.

VII. REFERENCES

- [1] B. Zitová and J. Flusser, “Image registration methods: A survey,” *Image Vis. Comput.*, vol. 21, no. 11, pp. 977–1000, Aug. 2003.
- [2] C. Huo, C. Pan, L. Huo, and Z. Zhou, “Multilevel SIFT matching for large-size VHR image registration,” *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 2, pp. 171–175, Mar. 2012.
- [3] Y. Han, J. Choi, Y. Byun, and Y. Kim, “Parameter optimization for the extraction of matching points between high-resolution multisensory images in urban areas,” *IEEE Trans. Geosci. Rem. Sens.*, vol. 52, no. 9, pp. 5612–5621, Sep. 2014.
- [4] J. Oh and C. Lee, “Automated bias-compensation of rational polynomial coefficients of high resolution satellite imagery based on topographic maps,” *ISPRS J. Photogramm. Remote Sens.*, vol. 100, pp. 14–22, Feb. 2015.

- [5] F. Bovolo, L. Bruzzone and S. Marchesi, "Analysis and adaptive estimation of the registration noise distribution in multitemporal VHR images," *IEEE Trans. Geosci. Rem. Sens.*, vol. 47, no. 8, pp. 2658–2671, Aug. 2009.
- [6] L. Bruzzone and R. Cossu, "Adaptive approach to reducing registration noise effects," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 11, pp. 2455–2465, Nov. 2003.
- [7] L. Bruzzone and R. Cossu, "Adaptive approach to reducing registration noise effects," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 11, pp. 2455–2465, Nov. 2003.
- [8] J. Townshend, C. Justice, C. Gurney, and J. McManus, "The impact of misregistration on change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 30, no. 5, pp. 1054–1060, Sep. 1992.
- [9] J. Inglada and A. Giros, "On the possibility of automatic multisensor image registration," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 10, pp. 2104–2120, Oct. 2004.
- [10] F. Bovolo and L. Bruzzone, "A theoretical framework for unsupervised change detection based on change vector analysis in the polar domain," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 1, pp. 218–236, Jan. 2007.
- [11] T. Lindeberg, "Scale-space theory: A basic tool for analysing structures at different scales," *J. Appl. Statist.*, vol. 21, no. 2, pp. 224–270, 1994.
- [12] Y. Zhang and G. Hong, "An IHS and wavelet integrated approach to improve pan-sharpening visual quality of natural colour IKONOS and QuickBird images," *Inf. Fusion*, vol. 6, no. 3, pp. 225–234, Sep. 2005.
- [13] M. Fischler and R. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Comm. ACM*, vol. 24, no. 6, pp. 381–395, Jun. 1981.
- [14] C. Carson, S. Belongie, H. Greenspan, and J. Malik, "Blobworld: Image segmentation using expectation-maximization and its application to image querying," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 8, pp. 1026–1038, Aug. 2002.

[15]Y. Han, F. Bovolo, L. Bruzzone, “An approach to fine coregistration between very-high-resolution multispectral images based on registration noise distribution,” *IEEE Trans. Geosci. Rem. Sens.*, vol. 53, no. 12, pp. 6650–6662, Dec. 2015.