CHANGE DETECTION IN HETEROGENEOUS REMOTE SENSING IMAGES BASED ON AN IMAGING MODALITY-INVARIANT MDS REPRESENTATION

Redha Touati^{†‡}, Max Mignotte[†], Mohamed Dahmane[‡]

[†]Image Processing Laboratory, DIRO, University of Montréal, Canada [‡]R&D vision Department, Centre de Recherche Informatique de Montréal (CRIM) Email: touatire@iro.umontreal.ca

ABSTRACT

In this paper, we propose a new multimodal change detection in remote sensing. The proposed method is based on a projection of the two multisensor satellite images to a common feature space, in which the two heterogeneous images share the same statistical properties and on which any classical monomodal change detection methods can be applied. This transformation of the before and after images is mainly based on a Multidimensional Scaling(MDS) representation which can be also viewed as a *de-texturing* approach of the two multisource images. Experimental results involving different types of imaging techniques confirm the reliability of the proposed approach.

1. INTRODUCTION

Multimodal Change Detection (CD) consists in identifying any land cover changes/uses that may have occurred between two satellite images acquired on the same geographical area, at different times, by two different kinds of imaging techniques. It is a recent and challenging task in the area of remote sensing, also called *multi-sensor data fusion*, that actually generalizes the classical monomodal CD issue [1–3] already used for solving the environmental monitoring, geological resources surveys and disaster detection/localization and quantification to name a few. The combination of images acquired by different sensor types (e.g. active and passive) or the finding of reliable imaging modality-invariant features, coming from different data sources is a difficult task. However, this difficulty is widely compensated by the numerous practical and technical advantages of such multimodal analysis procedure. Indeed, with the development of satellite and remote sensing imaging technology, a huge amount of heterogeneous data are acquired every day and stored in data archives for later use. By this fact, it can happen that, for example, an optical image of an area, provided by an archive, have to be necessarily combined with a new SAR image (of the same area) for technical reasons, lack of time, availability or atmospheric conditions in an emergency situation (SAR sensors can operate regardless of weather or thermal conditions, even at night, *i.e.* with less

restrictive conditions compared to optical imaging). It is also worth mentioning that, since a multimodal CD analysis processes heterogeneous data with different statistics, this new technique may be more robust to natural variations in environmental variables such as soil moisture or phenological states or shading effects which should not be detected as major land cover changes.

Until now, among the few research works that have been devoted to heterogeneous CD problem, we can identify four main categories. Namely; parametric models, non-parametric or learning machine based methods, algorithms based on operators using spatial and temporal similarity measures with invariance according to the imaging modality or finally, procedures mainly based on a transformation or projection of the two multimodal images to a common feature space, in which the two heterogeneous images share the same statistical properties and on which classical monomodal CD methods can then be applied.

In parametric models, a mixture or a set of parametric multidimensional distributions are generally used to model the joint statistics or the dependencies between the two imaging modalities [4, 5]. Sometimes, these models take also into account the noise characteristics and the relationships between the sensor responses to the objects and their physical properties [6, 7]. The main problems related with these parametric models are that they have been especially designed with specific distribution laws related to a type of multimodal sensors and are not easily generalizable for another pair of different sensors. In addition, these methods require a Maximum Likelihood (ML) parameter estimation step of the considered distribution, which can be complex and computationally expensive. Sometimes, these models are also semi-supervised and rely on a training set to fit the parametric model.

Among nonparametric methods, an energy minimization model has been specifically designed and solved in the leastsquares sense in [8] for satisfying an overdetermined set of constraints, expressed for each pair of pixels existing in the before-and-after satellite images acquired through different modalities. Deep learning methods through conditional adversarial networks [9] or convolutional coupling networks [10] have also been proposed and turn out to be valuable for the multimodal CD problem. In fact, these nonparametric methods have the ability to adapt to a wide variety of different imaging modalities (with possibly different noise types and levels) but are also generally less accurate than a parametric model dealing with a specific type of multimodality repre-

¹Acknowledgements: we would like to acknowledge the Computer Research Institute of Montreal (CRIM) and the Ministry of Economic Science and Innovation (MESI) of the Government of Québec to have supported this work.

sented by a particular distribution whose shape is clearly theoretically determined.

In the third family of methods relying on similarity measures with invariance according to the imaging modality, Alberga et al. [11] propose to use a technique closed to the co-registration and based on the use of a combination of different invariant similarity measures (such as correlation ratio, mutual information, etc.). Also, authors in [12] presented a CD method to quantify the damages caused by an earthquake to each individual building from a pre-event optical and post-event SAR images. In this work, simulation is used to predict the expected SAR signature of each building from the optical image which is then compared to the actual SAR scene to quantify the damages caused to each building. In [13], an imaging modality-invariant operator that detects the common specific high-frequency pattern of each structural region existing in the two heterogeneous satellite images is proposed. Finally, in the last category in which the bitemporal image data is projected to a common feature space for comparison convenience, [14] proposes also a representation, especially designed to highlight the changes. Another representation which turn out to be invariant to imaging modality is given by a classical segmentation. In this way, Liu *et al.* in [15] propose a general multidimensional evidential reasoning approach for estimating the segmentation map of the two satellite images which are then easily and subsequently compared.

In this work, we propose a new multimodal CD method belonging to the last category and based on a common feature space thus making possible the direct comparison between the two input images. This transformation of the before and after images can be viewed as a *de-texturing* approach [16] of each satellite image.

2. PROPOSED CHANGE DETECTION MODEL

Imaging Modality Invariant Projection: This step aims at finding a common feature space in which the pixels of the two satellite image should ideally have the same statistical properties. This task is not trivial, especially when the SAR imaging modality has to be combined with the optical imaging technique because the textural properties of the two images are radically different; For the SAR image, the inherent multiplicative speckle noise creates for each land cover class, a kind of macro-texture with grainy patterns (related to the back-scattered intensity of the different object surfaces). For the optical image, the noise is additive and degrades either piece-wise uniform areas or micro-textured and structured regions (representing in fact the reflection intensity of objects).

A solution consists in *de-texturing* the two satellite images, *i.e.*, to create a new (grey level) mapping in which two textured areas (around pixels at distant locations) gives, in the transformed image, two pixels whose grey-level intensity difference is proportional to a distance measure between these two textures. Otherwise said, in this new mapping, two non-adjacent or distant pixels with the same local texture (around the pixel) should have the same (grey-level) intensity.

To this end, a de-texturing approach, close to the one proposed in [16], is applied respectively on the first and second input satellite images. To this end, each pixel of an image is characterized by a feature vector gathering the values of the coarsely quantized grey level histogram followed by the values of the coarsely quantized gradient magnitude histograms in the four directions (respectively vertical, horizontal, right diagonal and left diagonal). These two histograms are computed over the set of pixels existing in an overlapping squared fixed-size (N_w) neighborhood centered around the pixel to be characterized. In our application, this local histogram is respectively quantized with q_l and q_g equidistant binnings for the grey level space and for each of the four gradient magnitude histograms. This simple texture feature extraction step thus yields to a $D = q_l + 4 q_g$ -dimensional feature vector for each pixel. This local feature descriptor turn out to be both discriminant to characterize the different grainy patterns of a SAR image or the different textural patterns specific of another imaging modality.

Once this set of feature vector are extracted for each pixel, we reduce the dimensionality of this set of feature vectors to one dimension with a Multidimensional Scaling (MDS) technique [17,18]. This allows us to project each textured image on a one-dimensional representation or concretely as a new grey-level transformed image. The interest of the MDS over other dimensionality reduction methods lies in the fact that this technique has the particularity of being able to estimate (optimally, in the least-squares sense) an embedding from the set of feature vectors in the high dimensional space $(\dim = D)$ such that the distances are faithfully preserved in the low dimensional (dim=1) target space and thus to ensure that two distant pixels (in the transformed image) will necessarily have a grey-level intensity difference proportional to a L_2 (in our application) distance between the two corresponding texture descriptors extracted on the input satellite image. Nevertheless, for computationally reasons, the originally proposed MDS algorithm (called *metric MDS* is not appropriate in our application and more generally for all large scale applications) because this algorithm requires a complexity of $O(N^2)$ (N being the number of pixels). Instead, we have herein used a fast alternative, called FastMap [19] whose main advantage is its linear complexity (at the price of a slightly less good approximation in the least- squares sense).

At this level, it lacks one very important aspect of the common feature space we search to build. Indeed, as already said, the L₂ distance between each textural feature vectors $(D_{\langle s,t \rangle})$, at locations s and t, in the high dimensional space $(\dim = D)$ and the distances between the grey level $(d_{s,t})$ at these same locations in the low dimensional (dim=1) target space is preserved as faithfully as possible and thus the relation $D_{\langle s,t\rangle}^{t_1} \equiv d_{\langle s,t\rangle}^{t_1}$ is true for the pre-event satellite image (at time t_1) and for the post-event image satellite (at time t_2) $D_{\langle s,t\rangle}^{t_2} \equiv d_{\langle s,t\rangle}^{t_2}$ (for any $\langle s,t\rangle$). Nevertheless, for two distant pixels s and t belonging to the class label unchanged area in satellite image t_1 and t_2 , currently, nothing ensures that the grey level at location s in the first (pre-event) projected image and second (post-event) projected image are similar. The MDS technique respects the monotonicity of the grey level order (linear correlation) existing in an image, nevertheless, a nonlinear monotonic scale factor between the two transformed images could however exist. In order to correct this, we resort to a double histogram matching method [20]. More precisely, let us consider the two bi-temporal remote sensing images, y^{t_1} and y^{t_2} acquired before and after a given event and \hat{y}^{t_1} and \hat{y}^{t_2} their MDS projection. \hat{y}^{t_1} is histogram matched to the *after* image \hat{y}^{t_2} to give $\hat{y}^{t_1}_{\bigstar}$ and \hat{y}^{t_2} is then



Fig. 1. First row: SAR/Optical dataset; before and after images; Second row shows the before and after images after MDS projection; Third row represents the result of the double Histogram matching on the images of the second row. Fourth row: difference map; final segmentation result; Fifth row: ground truth.

histogram matched to $\hat{y}^{t_1}_{\bigstar}$ in order to finally obtain $\hat{y}^{t_2}_{\bigstar}$ (see Fig. 1).

Temporal Differentiation and Binarization: At this level, we can apply any monomodal CD method. In our case, we simply generate a difference image by subtracting $\hat{y}_{\star}^{t_1}$ to $\hat{y}_{\star}^{t_2}$ and taking the absolute value to obtain the difference image y^D . Finally y^D is then segmented into two classes to distinguish changes of interest of the land cover. To this end, in order to achieve more robustness, changes are then identified, from the difference image y^D , by combining the results of T = 3 different automatic thresholding algorithms ([24–26]). In this way, this strategy allows us to synergistically integrate multiple different criteria, for which these binary segmentation algorithms have been designed to be optimal in order to further increase the efficiency of our binarization scheme. In



Fig. 2. Optical(NIR)/Optical dataset. From lex. order; image t_1 , t_2 ; difference map; final segmentation result; ground truth.

our application, this binary fusion process is simply achieved by using a median filter using a three dimensional window $W \times W \times T$ whose the first two dimensions are spatial and the third dimension indexes the different binary thresholded maps to be fused.

3. EXPERIMENTAL RESULTS

In order to assess the efficiency of the proposed method to detect different types of land cover changes and to show the strength and the ability of the proposed multimodal CD method to process different remote sensing modalities, we conduct a series of tests on different real multi-source remote sensing imagery data sets. These data sets reflect the three possible change detection conditions in multimodal case. We compare the performance of our method with different state-of-the-art multimodal change detection algorithms recently proposed [5, 21–23]. The different change masks were provided by a photo interpreter.

The first data set consists of one SAR image and one RGB optical image. It shows a piece of the Dongying City in China, before and after a new building construction. The SAR image is acquired by RADARSAT-2 (Jun. 2008) with spatial resolution of 8m. The optical image comes from Google Earth image (Sept. 2012) and its a combination of aerial photography imaging with a satellite imaging (produced respectively by QuickBird and Landsat-7) with a spatial resolution of 4m. After co-registration, they are of the same pixel-resolution 921 \times 593 pixels.

The second dataset is composed of two heterogeneous optical images. It shows the changes of the Mediterranean in Sardinia area (Italy). This dataset is acquired by different sensor specifications, and consists of one TM image

Table 1. Accuracy rate of change detection on the fourth heterogeneous datasets obtained by the proposed method and the state-of-the-art *multimodal* change detectors (supervised and unsupervised) and *monomodal* change detectors.

SAR/Optical Dataset (1)	Accuracy		(Opt	cical(NIR band)/Optical Dataset (2)) Accura	acy	
Proposed method	0.967				Proposed method	0.94	0.942	
Liu <i>et al.</i> [21]	0.976				Zhang $et al.$ [22]	0.975	5	
PCC [21]	0.821			PCC [22]		0.882		
SAR/Optical Dataset (3) Accu		Accura	acy		1-look SAR/5-look SAR Dataset (4)	Accuracy		
Proposed method		0.878		1	Proposed method	0.827		
Jorge et al. [23]		0.844			Chatelain $et al. [5]$	0.732		
Correlation [23]		0.67	70		Correlation [5]	0.521		
Mutual Inf. [23]		0.580			Ratio edge [5]	0.382		



Fig. 3. TSX/Optical dataset. From lex. order; image t_1 , t_2 , difference map; final segmentation result; ground truth.



Fig. 4. SAR 1-look/SAR 5-looks dataset. From lex. order; image t_1, t_2 ; difference map; final segmentation result; ground truth.

and one optical image. The TM image is the near-infrared band of the Landsat-5 (Sept. 1995 with spatial resolution of 30m.). The optical image come from Google Earth (RGB, Jul. 1996, Landsat-5) with the spatial resolution 4m. Af-

ter co-registration they are of same pixel-resolution 412×300 pixels.

The third heterogeneous data set consists of one optical image and one SAR image. It shows the area of Toulouse (FR), with a size of 4404×2604 pixels. The SAR image is taken by the TerraSAR-X satellite on Feb. 2009 before a building construction. The optical image is captured by the Pleiades satellite on Jul. 2013 after the construction of a building. The optical image have a resolution of 2m. The TSX image was co-registered and re-sampled by [7] to match the optical image resolution.

The fourth multimodal dataset is composed of two heterogeneous SAR images. It shows the area of Gloucester (UK) before and after a flooding event, with a size of 762×292 pixels and with a pixel resolution of 40m. The before and after images are captured by the RADARSAT satellite with different number of looks. The numbers of looks for the before and after SAR image is one-look image (Sept. and Oct. 2000) and five-looks.

In all the experimental results, we have considered $N_w = 7$, $q_g = 10$, $q_l = 40$ and W = 7 (see Section 2).

Table 1 summarizes the different change detection accuracy rates obtained by our approach with a comparison with other state of the art approaches. We can see that the different changed-unchanged detection binary map results match fairly the different regions present in the ground truth, and that the most changed regions for the different imagery modalities are well recognized by our strategy (see Figs. 1-4).

4. CONCLUSION

In this work, the applicability of a new multimodal change detection strategy, in remote sensing, is presented. This one is based on an imaging modality-invariant transformation that projects the two multisensor satellite images to a common feature space in which the bi-temporal images share the same statistical properties and thus on which any simple monomodal change detection methods can be applied. Qualitative and quantitative results show that the proposed method offers a good compromise between simplicity of the implementation and reliability. Indeed, this method consistently performs well on different types of input satellite images and showing different kind of changes.

5. REFERENCES

- L. Bruzzone and D. Prieto, "An adaptive semiparametric and context-based approach to unsupervised change detection in multitemporal remote-sensing images," *IEEE Transactions on Image Processing*, vol. 11, no. 4, pp. 452– 466, 2002.
- [2] R. Hedjam, M. Kalacska, M. Mignotte, H. Z. Nafchi, and M. Cheriet, "Iterative classifiers combination model for change detection in remote sensing imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 12, pp. 6997–7008, December 2016.
- [3] R. Touati and M. Mignotte, "A multidimensional scaling optimization and fusion approach for the unsupervised change detection problem in remote sensing images," in Sixth International Conference on Image Processing Theory, Tools and Applications, IPTA 2016, Oulu, Finland, December 2016, pp. 1–6.
- [4] G. Mercier, G. Moser, and S. Serpico, "Conditional copulas for change detection in heterogeneous remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 5, pp. 1428–1441, May 2008.
- [5] F. Chatelain, J. Y. Tourneret, and J. Inglada, "Change detection in multisensor sar images using bivariate gamma distributions," *IEEE Transactions on Image Processing*, vol. 17, no. 3, pp. 249–258, March 2008.
- [6] J. Prendes, M. Chabert, F. Pascal, A. Giros, and J. Tourneret, "A new multivariate statistical model for change detection in images acquired by homogeneous and heterogeneous sensors," *IEEE Transactions on Im*age Processing, vol. 24, no. 3, pp. 799–812, March 2015.
- [7] J. Prendes, M. Chabert, F. Pascal, A. Giros, and J.-Y. Tourneret, "Change detection for optical and radar images using a Bayesian nonparametric model coupled with a Markov random field," in *Proc. IEEE International Conference on Acoustic, Speech, and Signal Processing, ICASSP'15*, Brisbane, Australia, April 2015.
- [8] R. Touati and M. Mignotte, "An energy-based model encoding non-local pairwise pixel interactions for multisensor change detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 1, February 2018.
- [9] N. Merkle, P. F. S. Auer, and R. Muller, "On the possibility of conditional adversarial networks for multisensor image matching," in *Proceedings of IGARSS 2017*, *IGARSS 2017*, Fort Worth, Texas, USA, July 2017, pp. 1–4.
- [10] J. Liu, M. Gong, K. Qin, and P. Zhang, "A deep convolutional coupling network for change detection based on heterogeneous optical and radar images." *IEEE Transactions Neural Netw. Learn. Syst.*, 2018, in press.
- [11] V. Alberga, "Similarity measures of remotely sensed multi-sensor images for change detection applications," *Remote Sensing*, vol. 1, no. 3, pp. 122–143, 2009.
- [12] D. Brunner, G. Lemoine, and L. Bruzzone, "Earthquake damage assessment of buildings using vhr optical and sar imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 5, pp. 2403–2420, May 2010.

- [13] R. Touati, M. Mignotte, and M. Dahmane, "A new change detector in heterogeneous remote sensing imagery," in *Proceedings of 7th IEEE International Conference on Image Processing Theory, Tools and Applications, IPTA 2017*, Montreal, Canada, 2017.
- [14] C. Wu, B. Du, and L. Zhang, "Slow feature analysis for change detection in multispectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 5, pp. 2858–2874, 2014.
- [15] Z. g. Liu, G. Mercier, J. Dezert, and Q. Pan, "Change detection in heterogeneous remote sensing images based on multidimensional evidential reasoning," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 1, pp. 168–172, January 2014.
- [16] M. Mignotte, "MDS-based multiresolution nonlinear dimensionality reduction model for color image segmentation," *IEEE Trans. on on Neural Networks*, vol. 22, no. 3, pp. 447–460, March 2011.
- [17] W. Torgerson, "Multidimensional scaling: I. theory and method," *Psychometrika*, vol. 17, pp. 401–419, 1952.
- [18] T. F. Cox and M. A. A. Cox, "Multidimensional scaling," Chapman & Hall, London, 1994.
- [19] C. Faloutsos and K.-I. Lin, "FastMap: A Fast Algorithm for Indexing, Data-Mining and Visualization of Traditional and Multimedia Datasets," in *Proceedings of the ACM SIGMOD International Conference on Management of Data*, San Jose, California, june 1995, pp. 163–174.
- [20] D. Shapira, S. Avidan, and Y. Hel-Or, "Multiple histogram matching," in *IEEE International Conference on Image Processing*, *ICIP'13*, September 2013, pp. 2269– 2273.
- [21] J. Liu, M. Gong, K. Qin, and P. Zhang, "A deep convolutional coupling network for change detection based on heterogeneous optical and radar images," *IEEE Transactions on Neural Networks and Learning Systems*, vol. PP, no. 99, pp. 1–15, 2017.
- [22] P. Zhang, M. Gong, L. Su, J. Liu, and Z. Li, "Change detection based on deep feature representation and mapping transformation for multi-spatial-resolution remote sensing images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 116, pp. 24–41, 2016.
- [23] J. Prendes, M. Chabert, F. Pascal, A. Giros, and J. Tourneret, "Performance assessment of a recent change detection method for homogeneous and heterogeneous images," *Revue Française de Photogrammétrie* et de Télédétection, pp. 163–174, 2015.
- [24] J. Kapur, P. Sahoo, and A. Wong, "A new method for gray-level picture thresholding using the entropy of the histogram," *Computer Vision, Graphics, and Image Processing*, vol. 29, no. 3, pp. 273–285, 1985.
- [25] G. Zack, W.E.Rogers, and S. Latt, "Automatic Measurement of Sister Chromatid Exchange Frequency," *Journal* of Histochemistry and Cytochemistry, vol. 25, no. 7, pp. 741–753, 1977.
- [26] J. Yen, F. Chang, and S. Chang, "A new criterion for automatic multilevel thresholding," *IEEE Transactions* on *Image Processing*, vol. 4, no. 3, pp. 370–378, 1995.