CHANGE DETECTION WITH HETEROGENEOUS REMOTE SENSING DATA: FROM SEMI-PARAMETRIC REGRESSION TO DEEP LEARNING

Gabriele Moser[1], *Stian N. Anfinsen*[2], *Luigi T. Luppino*[2], *and Sebastiano B. Serpico*[1]

[1] University of Genoa, DITEN Dept., via Opera Pia 11a, Genoa, Italy, gabriele.moser@unige.it
[2] UiT The Arctic University of Norway, Dept. of Physics and Technology, Machine Learning Group

ABSTRACT

Change detection represents a major family of remote sensing image analysis techniques and plays a fundamental role in a variety of applications to environmental monitoring and disaster risk management. However, most change detection methods operate under the assumption that the multitemporal input data have been collected with the same (or very similar) acquisition modality – a possibly critical restriction in several applications. In this paper, the problem and the opportunities of change detection from multitemporal data acquired through heterogeneous modalities are addressed. Methodologically, this is a highly challenging data fusion problem, especially within an unsupervised framework. Here, these challenges and the methodological approaches proposed in the literature, which range from earlier semi-parametric regression to current deep learning architectures, are reviewed. Then, recent fully unsupervised techniques, based on spectral clustering, traditional image regression, and deep image-to-image translation, are briefly described.

Index Terms— Heterogeneous change detection, multimodal data fusion, image regression, image-to-image translation, deep learning, affinity matrix.

1. INTRODUCTION

Multitemporal remote sensing imagery represents a precious information source in many applications to agriculture, forestry, climate change monitoring, urban planning, and disaster risk management. This potential is further enforced by the current availability of data from a variety of space missions with optical (e.g., Sentinel-2, Landsat 8, Pléiades, WorldView-3) and synthetic aperture radar (SAR) payloads (e.g., COSMO-SkyMed Second Generation, TerraSAR-X, RADARSAT-2). In this framework, a family of image analysis methods of major importance is aimed at *change detection (CD)*, i.e., at identifying the pixels that underwent ground changes in between two acquisition times t_1 and t_2 [1, 2].

Most CD methods operate under the assumption that data at t_1 and t_2 have been acquired through the same *modality*, i.e., the same geometrical configuration, spectral bands, radar frequencies, etc [1, 2]. This scenario will be named *homogeneous CD* in the following. A milder statement may be that differences in the acquisition modality are small enough to be mitigated through normalization or co-calibration. Methodologically, this is a natural assumption because it favors that significant variations in the at-sensor signal can be attributed to ground changes rather than to differences in the acquisition modality. However, from an application-oriented perspective, this assumption can be severely restrictive when single-modal data are unavailable or their collection requires unacceptable latency times (e.g., in the case of CD for damage assessment shortly after a natural disaster). Homogeneous CD is a long-studied problem for which many methodological solutions, involving for example Bayesian estimation theory, Markov random fields, segmentation, fuzzy set theory, and kernel machines, have been proposed [1, 3, 2].

When multitemporal data characterized by different acquisition modalities are involved (e.g., a multispectral image at t_1 and a SAR image at t_2 or a C-band SAR image at t_1 and an X-band SAR image at t_2) – a scenario that we shall name *heterogeneous CD* –, a highly challenging data fusion task arises, because one wishes to "compare" observations whose probability distributions intrinsically differ even when they are taken on an unchanged region. The literature of heterogeneous CD is substantially more recent and scarcer than that of homogeneous CD and was essentially initiated by the copula-theoretic work in [4]. Lately, remarkable developments have been obtained through deep learning [5, 6].

In this paper, a general overview of the heterogeneous CD problem and its challenges is presented (see Section 2). The methodological approaches that have been proposed in the literature are reviewed (see Section 3), and recent methods developed by the authors are briefly described (see Section 4).

2. HETEROGENEOUS CHANGE DETECTION

Let $\mathcal{X}$ and $\mathcal{Y}$ be the images collected at times t_1 and t_2, respectively, and let $\boldsymbol{x}$ and $\boldsymbol{y}$ be two generic data samples drawn on the same spatial location from $\mathcal{X}$ and $\mathcal{Y}$, respectively. Depending on the specifics of each individual method, $\boldsymbol{x}$ and $\boldsymbol{y}$ may include data from either an individual pixel or a patch of neighboring pixels (the latter case is most typical of deep neural methods). If H_0 and H_1 indicate the "no-change" and "change" hypotheses, in the case of homogeneous CD one generally assumes that $\boldsymbol{x}$ and $\boldsymbol{y}$ exhibit the same (or very similar) distribution when conditioned to H_0. This condition is violated in the case of heterogeneous CD, thus hindering the possibility of a direct "comparison" between $\boldsymbol{x}$ and $\boldsymbol{y}$ [4].

Broadly speaking, most heterogeneous CD methods are based on the key idea of transforming $\boldsymbol{x}$ and $\boldsymbol{y}$ so that they exhibit the same (or similar) distributions. This can be accomplished either by a mapping $\boldsymbol{x} \mapsto \phi(\boldsymbol{x}) = \tilde{\boldsymbol{y}}$ and/or $\boldsymbol{y} \mapsto \psi(\boldsymbol{y}) = \tilde{\boldsymbol{x}}$ such that the distributions of the new random vectors $\tilde{\boldsymbol{x}}$ and $\tilde{\boldsymbol{y}}$, when conditioned to H_0, match those of $\boldsymbol{x}$ and $\boldsymbol{y}$, respectively, or by two mappings $\boldsymbol{x} \mapsto \xi(\boldsymbol{x}) = \boldsymbol{z}$ and $\boldsymbol{y} \mapsto \zeta(\boldsymbol{y}) = \boldsymbol{z}'$ so that $\boldsymbol{z}$ and $\boldsymbol{z}'$ has the same distribution given H_0. In the former case $\tilde{\boldsymbol{y}}$ is an estimate of what $\boldsymbol{x}$ would look like if it could be observed through the modality of $\boldsymbol{y}$ (and vice versa). In the latter case, $\boldsymbol{z}$ and $\boldsymbol{z}'$ may not have specific interpretations but are meant to bring the two modalities towards a common domain. From this perspective, heterogeneous CD remarkably relate to domain adaptation, transfer learning, and image-to-image translation [7, 8, 9, 10].

Based on the role of training samples in the determination of the mappings (ϕ, ψ) or (ξ, ζ), methods can be categorized into three broad groups: (i) *fully supervised*, (ii) *partially supervised*, and (iii) *fully unsupervised*. In case (i), training samples are assumed available for the thematic classes at both t_1 and t_2. This scenario involves strong ground truth requirements and is normally not strictly considered heterogeneous CD as it corresponds to a joint supervised classification of the input multitemporal data (i.e., $\boldsymbol{z}$ and $\boldsymbol{z}'$ are class labels), for which effective approaches have been available for long [11, 12, 13]. In case (ii), which encompasses all earlier approaches to heterogeneous CD, a training set is used although it does not cover all classes at all observation times. Most typically, it regards only H_0 with the goal of conveying prior information on what unchanged areas look like through the different modalities. In case (iii), which has been explored only lately, no training data are used. This is both remarkably interesting for applications, because it does not need any ground truth and is not prone to annotation errors, and very challenging methodologically, because no prior is available on the relation between the two modalities. In the next sections we will focus on strict-sense heterogeneous CD through partially supervised and fully unsupervised methods.

3. METHODOLOGICAL APPROACHES

3.1. Partially supervised semi/non-parametric methods

Partially supervised approaches to estimate (ϕ, ψ) or (ξ, ζ) have been proposed based on methodological ideas from various areas, with the goal of pushing the transformed data towards a common probability distribution (semi-parametric and non-parametric regression), a common manifold (manifold learning), or a common Hilbert space (kernel machines). The first approach to heterogeneous CD has been the partially supervised method in [4] that combines quantile regression and copula functions to learn ϕ and ψ using a training map for H_0. It integrates nonparametric estimators of the marginals of $\boldsymbol{x}$ and $\boldsymbol{y}$ and parametric copula models of their dependence (hence, an overall semi-parametric method). A limitation is the focus on single-channel data (scalar x and y). In [14], ξ and ζ are modeled by combining estimates of the marginals of $\boldsymbol{x}$ and $\boldsymbol{y}$ with a meta-Gaussian distribution (essentially equivalent to a Gaussian copula). In [8], a nonparametric method is proposed in which ϕ and ψ are modeled by combining kernel regression and a nearest neighbor approach. In [15], a "no-change" manifold is extracted according to the local joint statistics and to models of the noise in the two modalities, so that changes are detected by using a case-specific distance from this manifold. A kernel-based approach is developed in [16] to determine mappings ξ and ζ to a common Hilbert space through a kernel generalization of the canonical correlation analysis (CCA). The optimization of the correlation functional is guided by a set of samples drawn from H_0.

3.2. Fully unsupervised nonparametric methods

Fully unsupervised heterogeneous CD methods have been developed recently on the basis of methodological components rooted in dictionary learning, multidimensional scaling, and deep learning. A goal of most approaches is to automatically identify pairs of samples from $\mathcal{X}$ and $\mathcal{Y}$ that can be used to learn (ϕ, ψ) or (ξ, ζ) and are likely unchanged. The latter condition is crucial to ensure that the resulting transforms emphasize the discrimination between H_0 and H_1.

The first fully unsupervised method has been developed in [17] based on dictionary learning. It determines two coupled dictionaries from $\mathcal{X}$ and $\mathcal{Y}$ through an iterative patch-wise learning process. The pairs of image patches that yield the largest reconstruction error are automatically removed, as they are interpreted as likely drawn from H_1. The resulting dictionaries bring both modalities towards a common code space in which similar sparse codes are expected for patches drawn from H_0. Multidimensional scaling concepts are used in [18] and [19]. In [18] an energy-based model is formulated to encode nonlocal pairwise pixel interactions and compute a transformed feature that preserves local pairwise similarity and its temporal evolution. In [19], multidimensional scaling is combined with local histograms of pixel intensities and of gradient magnitudes and with histogram matching.

3.3. Fully unsupervised deep learning methods

Deep learning, which has been particularly successful in remote sensing image analysis lately, has recently been found effective for heterogeneous CD as well. The rationale is to leverage on the powerful feature learning capabilities of deep neural nets to model (ϕ, ψ) or (ξ, ζ), typically using stacked denoising autoencoders (SDAEs) or generative adversarial networks (GANs). SDAEs consist of a cascade of two networks and are usually aimed at dimensionality reduction, feature extraction, data reconstruction, and regression [20]. A GAN is also made of two interconnected networks, which are trained in a competing fashion with the goal of generating, from an input noise source, samples whose distribution is indistinguishable from that of a target source [20]. Important extensions aimed at image-to-image translation also address the generation from input data conditioned to a separate source (conditional GAN) [21] or a consistent behavior in the translation across two sources (cycle GAN) [22].

In [5, 23, 24], SDAE architectures are proposed for fully unsupervised heterogeneous CD by integrating, in the related loss functions, a weighting on the probability that each sample is drawn from H_1. Network training and update of these weights are performed iteratively. Architectures combining SDAEs and clustering algorithms have been developed to automatically identify (and use for training) pixels that can be assigned to H_0 and H_1 with a certain confidence [25] and to address the detection of two distinct types of change [26].

GAN architectures have been proposed in [6, 27] for heterogeneous CD from optical and SAR data. In [27], a conditional GAN is used to learn a transform ϕ from input optical data $\boldsymbol{x}$ to estimated SAR data $\tilde{\boldsymbol{y}}$, and a further approximation network is trained to map the input SAR data $\boldsymbol{y}$ to a domain in which residual mismatches are minimized. In [6], mappings (ϕ, ψ) that estimate SAR from optical data and vice versa are learned through an architecture that combines cyclic adversarial terms and variational autoencoders [20]. In these methods as well, weights on the chances that each sample is drawn from H_1 are introduced and iteratively updated.

4. RECENT AFFINITY MATRIX-BASED METHODS

4.1. Affinity-based non-parametric image regression

As discussed above, a major challenge of fully unsupervised heterogeneous CD is to capture the relation between the two modalities over unchanged areas, while simultaneously emphasizing the discrimination of changed areas. On one hand, the flexibility of nonparametric – traditional and deep – approaches allows powerful regression models to be formulated. On the other hand, the lack of training information on H_0 or H_1 makes the learning of these models particularly complex.

An approach to address this challenge has recently been proposed in [28] based on the combination of nonparametric regression and spectral clustering concepts. With the goal of estimating ϕ and ψ in a fully unsupervised manner, the key idea is to leverage on the information captured by local affinity matrices to automatically determine a pseudo-training set $\mathcal{T}_0$ composed of pixel pairs that likely belong to H_0. The affinity matrix of a set of pixels provides a graph-theoretic characterization of their spatial structure and interrelations [29, 30]. The rationale of the approach in [28] is that, if a change occurs on some pixels, these interrelations are expected to change as well, quite regardless of the acquisition modality.

Computationally, for each pair p of spatially corresponding patches in $\mathcal{X}$ and $\mathcal{Y}$, fully connected graphs are defined within these patches, the affinity matrices $A_p^{\mathcal{X}}$ and $A_p^{\mathcal{Y}}$ associated with these graphs are computed, and the Frobenius distance between $A_p^{\mathcal{X}}$ and $A_p^{\mathcal{Y}}$ is evaluated. A pixel is inserted into $\mathcal{T}_0$ according to the statistics of the Frobenius distances obtained on all patches containing it. The resulting set $\mathcal{T}_0$ is used to train a multioutput support vector machine [31], a Gaussian process regression [32], a random forest [33], or the regression algorithm in [8] to learn ϕ and ψ. Case-specific algorithms, based on the Hellinger distance between probability distributions and on k-nearest neighbors, are integrated in the method to make sure that $\mathcal{T}_0$ is representative of "no-change" across the whole image and to adaptively tune the kernel width used to compute the affinities.

Experimental validation shown in [28] with Landsat 5 TM, EO-1 ALI, Landsat 8 OLI, and Sentinel-1A data pointed out the capability of this approach, applied with all aforementioned regression methods, to compute representative estimates of the image of each acquisition date through the modality of the other date and to derive accurate change maps. A special interest was noted for random forest, owing to its appealing computational properties. These results suggest the effectiveness of local affinity matrices in characterizing the spatial structure of multimodal data set for fully unsupervised heterogeneous CD. Details can be found in [28].

4.2. Deep image-to-image translation with affinity prior

Motivated by the powerful regression capabilities of deep neural networks and by the aforementioned potential of affinity matrix concepts for fully unsupervised heterogeneous CD, methods based on their combination and aimed at learning (ϕ, ψ) and (ξ, ζ) have recently been proposed in [34, 35]. The key idea is to formalize heterogeneous CD through a deep image-to-image translation problem and to incorporate affinity matrix information to condition the regression process to H_0. This idea is formalized by integrating into the loss functions associated with image translation a pixelwise prior computed as a function of the local affinity matrices. In this framework, architectures involving cyclically consistent autoencoders, adversarial networks, and CCA are developed. Details and experimental results can be found in [34, 35].

5. CONCLUSION

In this paper, the topic of heterogeneous CD from multitemporal multimodal remote sensing data has been addressed. It is a topic of great potential, especially in relation to the currently available wealth of satellite imagery. Yet, it is methodologically very challenging due to the intrinsic difficulty in modeling the dependence among highly heterogeneous data sources, especially in fully unsupervised scenarios. With the goal of outlining the main background concepts and challenges of this highly promising research area, previous methodological solutions have been reviewed, ranging from the first formulation as semi-parametric regression to current deep image-to-image translation, and recent techniques, based on the integration of local affinity matrices into traditional or deep image regression, have been briefly discussed.

6. REFERENCES

[1] A. Singh, "Review article digital change detection techniques using remotely-sensed data," *Int. J. Rem. Sens.*, vol. 10, pp. 989–1003, 1989.

[2] L. Bruzzone and F. Bovolo, "A novel framework for the design of change-detection systems for very-high-resolution remote sensing images," *Proc. of the IEEE*, vol. 101, pp. 609–630, 2013.

[3] F. Melgani, G. Moser, and S. B. Serpico, "Unsupervised change-detection methods for remote-sensing images," *Optical Engineering*, vol. 41, pp. 3288–3297, 2002.

[4] G. Mercier, G. Moser, and S. B. Serpico, "Conditional copulas for change detection in heterogeneous remote sensing images," *IEEE Trans. Geosci. Rem. Sens.*, vol. 46, pp. 1428–1441, 2008.

[5] 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 Trans. Neural Net. Learn. Systems*, vol. 29, pp. 545–559, 2016.

[6] M. Gong, X. Niu, T. Zhan, and M. Zhang, "A coupling translation network for change detection in heterogeneous images," *Int. J. Rem. Sens.*, vol. 40, no. 9, pp. 3647–3672, 2019.

[7] D. Marcos, R. Hamid, and Devis Tuia, "Geospatial correspondences for multimodal registration," in *Proc. CVPR*, 2016, pp. 5091–5100.

[8] Z. Liu, G. Li, G. Mercier, Y. He, and Q. Pan, "Change detection in heterogenous remote sensing images via homogeneous pixel transformation," *IEEE Trans. Image Process.*, vol. 27, pp. 1822–1834, 2018.

[9] D. Tuia, D. Marcos, and G. Camps-Valls, "Multi-temporal and multi-source remote sensing image classification by nonlinear relative normalization," *ISPRS J. Photogramm. Rem. Sens.*, vol. 120, pp. 1–12, 2016.

[10] S. H. Khan, X. He, F. Porikli, and M. Bennamoun, "Forest change detection in incomplete satellite images with deep neural networks," *IEEE Trans. Geosci. Rem. Sens.*, vol. 55, pp. 5407–5423, 2017.

[11] Philip H. Swain, "Bayesian classification in a time-varying environment.," *IEEE Trans Syst Man Cybern*, vol. SMC-8, pp. 879–883, 1978.

[12] L. Bruzzone, D. F. Prieto, and S. B. Serpico, "A neural-statistical approach to multitemporal and multisource remote-sensing image classification," *IEEE Trans. Geosci. Rem. Sens.*, vol. 37, pp. 1350–1359, 1999.

[13] F. Melgani and S. B. Serpico, "A Markov random field approach to spatio-temporal contextual image classification," *IEEE Trans. Geosci. Rem. Sens.*, vol. 41, pp. 2478–2487, 2003.

[14] B. Storvik, G. Storvik, and R. Fjortoft, "On the combination of multisensor data using meta-gaussian distributions," *IEEE Trans. Geosci. Rem. Sens.*, vol. 47, pp. 2372–2379, 2009.

[15] J. Prendes, M. Chabert, F. Pascal, A. Giros, and J.-Y. Tourneret, "A new multivariate statistical model for change detection in images acquired by homogeneous and heterogeneous sensors," *IEEE Trans. Image Process.*, vol. 24, pp. 799–812, 2015.

[16] M. Volpi, G. Camps-Valls, and D. Tuia, "Spectral alignment of multi-temporal cross-sensor images with automated kernel canonical correlation analysis," *ISPRS J. Photogramm. Rem. Sens.*, vol. 107, pp. 50–63, 2015.

[17] M. Gong, P. Zhang, L. Su, and J. Liu, "Coupled dictionary learning for change detection from multisource data," *IEEE Trans. Geosci. Rem. Sens.*, vol. 54, pp. 7077–7091, 2016.

[18] R. Touati and M. Mignotte, "An energy-based model encoding nonlocal pairwise pixel interactions for multisensor change detection," *IEEE Trans. Geosci. Rem. Sens.*, vol. 56, pp. 1046–1058, 2018.

[19] R. Touati, M. Mignotte, and M. Dahmane, "Change detection in heterogeneous remote sensing images based on an imaging modality-invariant MDS representation," in *Proc. ICIP*, 2018, pp. 3998–4002.

[20] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*, MIT Press, 2016.

[21] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. CVPR*, 2017.

[22] J.-Y. Zhu, T. Park, P. Isola, and A. A Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. CVPR*, 2017, pp. 2223–2232.

[23] W. Zhao, Z. Wang, M. Gong, and J. Liu, "Discriminative feature learning for unsupervised change detection in heterogeneous images based on a coupled neural network," *IEEE Trans. Geosci. Rem. Sens.*, 2017.

[24] T. Zhan, M. Gong, J. Liu, and P. Zhang, "Iterative feature mapping network for detecting multiple changes in multi-source remote sensing images," *ISPRS J. Photogramm. Rem. Sens.*, vol. 146, pp. 38–51, 2018.

[25] T. Zhan, M. Gong, X. Jiang, and S. Li, "Log-based transformation feature learning for change detection in heterogeneous images," *IEEE Geosci. Rem. Sens. Lett.*, vol. 15, no. 9, pp. 1352–1356, 2018.

[26] L. Su, M. Gong, P. Zhang, M. Zhang, J. Liu, and H. Yang, "Deep learning and mapping based ternary change detection for information unbalanced images," *Patt. Rec.*, vol. 66, pp. 213–228, 2017.

[27] X. Niu, M. Gong, T. Zhan, and Y. Yang, "A conditional adversarial network for change detection in heterogeneous images," *IEEE Geosci. Rem. Sens. Lett.*, vol. 16, no. 1, pp. 45–49, 2018.

[28] L. T. Luppino, F. M. Bianchi, G. Moser, and S. N. Anfinsen, "Unsupervised image regression for heterogeneous change detection," *IEEE Trans. Geosci. Rem. Sens.*, vol. 57, pp. 9960–9975, 2019.

[29] A. Y. Ng, M. I. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," in *Advances in Neural Information Processing Systems 14*, T. G. Dietterich, S. Becker, and Z. Ghahramani, Eds., pp. 849–856. MIT Press, 2002.

[30] Stian Normann Anfinsen, Robert Jenssen, and Torbjørn Eltoft, "Spectral clustering of polarimetric sar data with wishart-derived distance measures," in *Proc. POLinSAR*, 2007, vol. 7, pp. 1–9.

[31] D. Tuia, J. Verrelst, L. Alonso, F. Pérez-Cruz, and G. Camps-Valls, "Multioutput support vector regression for remote sensing biophysical parameter estimation," *IEEE Geosci Rem. Sens. Lett.*, vol. 8, pp. 804–808, 2011.

[32] C. E. Rasmussen, "Gaussian processes in machine learning," in *Advanced lectures on machine learning*, pp. 63–71. Springer, 2004.

[33] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, pp. 5–32, 2001.

[34] Luppino, L. T. and Kampffmeyer, M. and Bianchi, F. M. and Moser, G. and Serpico, S. B. and Jenssen, R. and Anfinsen, S. N., "Deep Image Translation with an Affinity-Based Change Prior for Unsupervised Multimodal Change Detection," in *IGARSS 2020, submitted*.

[35] Figari Tomenotti, F. and Luppino, L. T. and Hansen, M. A. and Bianchi, F. M. and Moser, G. and Serpico, S. B. and Jenssen, R. and Anfinsen, S. N., "Heterogeneous Change Detection with Self-Supervised Deep Canonically Correlated Autoencoders," in *IGARSS 2020, submitted*.