



HAL
open science

DEEP-NET INVERSION FOR FRUGAL SATELLITE IMAGE CHANGE DETECTION

Hichem Sahbi, Sebastien Deschamps

► **To cite this version:**

Hichem Sahbi, Sebastien Deschamps. DEEP-NET INVERSION FOR FRUGAL SATELLITE IMAGE CHANGE DETECTION. IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Jul 2023, Pasadena, United States. pp.6470 - 6473, 10.1109/IGARSS52108.2023.10283212 . hal-04274278

HAL Id: hal-04274278

<https://hal.science/hal-04274278v1>

Submitted on 7 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

DEEP-NET INVERSION FOR FRUGAL SATELLITE IMAGE CHANGE DETECTION

Hichem Sahbi¹

Sebastien Deschamps^{1,2}

¹Sorbonne University, CNRS, LIP6, Paris, France

²Therisis Thales, France

ABSTRACT

Change detection in satellite imagery seeks to find occurrences of targeted changes in a given scene taken at different instants. This task has several applications ranging from land-cover mapping, to anthropogenic activity monitoring as well as climate change and natural hazard damage assessment. However, change detection is highly challenging due to the acquisition conditions and also to the subjectivity of changes.

In this paper, we devise a novel algorithm for change detection based on active learning. The proposed method is based on a question & answer model that probes an oracle (user) about the relevance of changes only on a small set of critical images (referred to as virtual exemplars), and according to oracle's responses updates deep neural network (DNN) classifiers. The main contribution resides in a novel adversarial model that allows learning the most representative, diverse and uncertain virtual exemplars (as inverted preimages of the trained DNNs) that challenge (the most) the trained DNNs, and this leads to a better re-estimate of these networks in the subsequent iterations of active learning. Experiments show the out-performance of our proposed deep-net inversion against the related work.

Index Terms— active learning, adversarial virtual exemplars, deep-net inversion, satellite image change detection

1. INTRODUCTION

Satellite image change detection aims at localizing instances of targeted (relevant) changes in scenes acquired at different instants. The wide interest of this task ranges from anthropogenic activity monitoring, to phenology mapping, through natural hazard damage assessment [2, 3]. This task is challenging as satellite images are subject to many pervasive changes resulting from sensors and acquisition conditions (occlusions, radiometric variations and shadows, weather conditions as well as scene content). Early change detection work relies on comparisons of multi-temporal series, via image differences and thresholding, vegetation indices, principal component and change vector analyses [5–8]. Other work either (i) requires a preliminary step that mitigates the effect of pervasive changes using normalization techniques including shadow removal, radiometric correction and by estimating the parameters of sensors for registration, etc.

[10, 11, 13, 14, 38], or (ii) considers these irrelevant variations as a part of scene appearance modeling using statistical machine learning [9, 12, 15–20, 39].

The success of machine learning models in particular, including deep neural networks, is highly reliant on the availability of large collections of hand-labeled reference images that capture the inherent variability of relevant and irrelevant changes as well as the user's targeted relevant changes [1, 4]. However, in practice, labeled data are scarce and their hand-labeling is time and effort demanding, and even when available they suffer from domain-shift as their labeling may not reflect the user's subjectivity and intention. Several solutions seek to make machine learning methods frugal and less labeled-data hungry [23, 34] including few shot [21] and self-supervised learning [35]; however, these methods are agnostic to the users' intention. Hence, solutions based on active learning [22, 24, 25, 27–29, 31–33] are rather more appropriate and consist in *frugally probing* the user (a.k.a. oracle) about the relevance of observed changes, and according to the oracle's responses, train decision criteria that best suit the user's intention and scene acquisition conditions.

In this paper, we devise a new satellite image change detection algorithm based on frugal training of deep neural networks. Starting from a large unlabeled dataset, the proposed model is interactive and consists in querying the oracle about the labels of *critical* data whose positive impact on the trained deep networks is the most important. These critical data (also dubbed as virtual exemplars) are obtained by *deep-net inversion* using an adversarial loss that allows synthesizing representative and diverse exemplars whose classification scores are also ambiguous. These virtual exemplars are learned – instead of being sampled from a *fixed* pool of unlabeled data – in order to challenge (the most) the previously trained change detection criteria leading to a better re-estimation of these criteria at the subsequent iterations of change detection. Note that the formulation presented in this paper, while being adversarial, is conceptually different from the ones widely used in generative adversarial networks (GANs) [37]; indeed, whereas GANs seek to generate *fake* exemplars that mislead the trained discriminators, our formulation aims instead at generating *critical* exemplars — for further annotations — whose impact on the subsequent learned deep network classifiers is the most noticeable. Differently put, the proposed framework allows

to sparingly probe the oracle only on the most representative, diverse and uncertain exemplars that challenge the current deep network discriminator, and eventually lead to more accurate ones in the following iterations of change detection. Extensive change detection experiments corroborate these claims and show the effectiveness of our *deep-net inversion and exemplar learning* models against comparative methods.

2. PROPOSED METHOD

Let $\mathcal{I}_r = \{p_1, \dots, p_n\}$, $\mathcal{I}_t = \{q_1, \dots, q_n\}$ denote two registered satellite images captured at two instants t_0, t_1 , with $p_i, q_i \in \mathbb{R}^d$. Considering $\mathcal{I} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$, with each $\mathbf{x}_i = (p_i, q_i)$, and $\mathcal{Y} = \{\mathbf{y}_1, \dots, \mathbf{y}_n\}$ the underlying unknown labels; our goal is to train a classifier f — as a deep convolutional network — that predicts the unknown labels in $\{\mathbf{y}_i\}_i$ with $y_i = 1$ if the patch q_i corresponds to a “change” w.r.t. the underlying patch p_i , and $y_i = 0$ otherwise. Training f requires data hand-labeled by an oracle. Our goal is to make the design of f *label-frugal* and as accurate as possible.

2.1. Interactive change detection at a glance

In order to design our change detection algorithm, we consider a question & answer interactive process which first provides the oracle with very few *critical patch pairs* for labeling, and then updates change detection criteria accordingly. In what follows, the subset of critical data, at any iteration t , is referred to as *display*, and denoted as $\mathcal{D}_t \subset \mathcal{I}$, and \mathcal{Y}_t its unknown labels. Starting from a random display \mathcal{D}_0 , we train change detection criteria f_0, \dots, f_{T-1} iteratively according to the subsequent steps

1/ Obtain \mathcal{Y}_t from the oracle, and train $f_t(\cdot)$ on $\cup_{k=0}^t (\mathcal{D}_k, \mathcal{Y}_k)$ using graph convolution networks (GCNs) [30].

2/ Find the next display $\mathcal{D}_{t+1} \subset \mathcal{I} \setminus \cup_{k=0}^t \mathcal{D}_k$. It is clear that a strategy that considers all possible displays $\mathcal{D} \subset \mathcal{I} \setminus \cup_{k=0}^t \mathcal{D}_k$, trains the underlying criteria $f_{t+1}(\cdot)$ on $\mathcal{D} \cup_{k=0}^t \mathcal{D}_k$, and keeps the display with the highest performances is highly combinatorial. Furthermore, it requires labeling each of these displays and this is clearly intractable. In this paper, we rely on active learning display selection strategies, which are rather more appropriate; nevertheless, one should be cautious about these strategies as many of them are shown to be equivalent to (or even worse than) basic strategies that pick data uniformly randomly (see for instance [24] and references within).

As the main contribution of this work, our proposed display selection strategy is different from common ones (see e.g. [36]) and relies on hallucinating virtual exemplars instead of taking them from existing pools of data. This design allows exploring *more flexibly* the uncharted parts of unlabeled data. These exemplars are selected as the most diverse, representative and uncertain data so they *challenge the most* the current change detection criteria, leading to better re-estimate of these criteria in the subsequent iterations of active learning.

2.2. Virtual exemplars: column-stochastic model

As obtaining labels is usually highly expensive, we train a deep network $f_{t+1}(\cdot)$ on $\mathcal{D}_{t+1} \cup \dots \cup \mathcal{D}_0$ where \mathcal{D}_{t+1} is a subset of critical virtual exemplars (denoted as $\{\mathbf{D}_k\}_{k=1}^K$) whose labeling is frugally obtained from the oracle. In contrast to our previous work [26], we consider a variant which defines for each virtual exemplar \mathbf{D}_k a distribution $\{\mu_{ik}\}_{i=1}^n$ that measures the conditional probability of associating \mathbf{D}_k to the n -training samples. With this variant, the virtual exemplars together with their distributions $\{\mu_{ik}\}_{i,k}$ are obtained by minimizing the following surrogate problem

$$\min_{\mathbf{D}; \mu \in \Omega_c} \text{tr}(d(\mathbf{D}, \mathcal{I}) \mu) + \alpha \left[\frac{1}{K} \mu \mathbf{1}_K \right]^\top \log \left[\frac{1}{K} \mu \mathbf{1}_K \right] + \beta \text{tr}(f(\mathbf{D})^\top \log f(\mathbf{D})) + \gamma \text{tr}(\mu \log \mu^\top), \quad (1)$$

here $\Omega_c = \{\mu : \mu \geq 0; \mu^\top \mathbf{1}_n = \mathbf{1}_K\}$ guarantees the column-stochasticity of the memberships $\{\mu_{ik}\}_i$ instead of row-stochasticity in [26], and $\mathbf{1}_K, \mathbf{1}_n$ denote two vectors of K and n ones respectively. In Eq. 1 the symbol $^\top$ stands for the matrix transpose operator, $\mathbf{D} \in \mathbb{R}^{d \times K}$ is a matrix whose k -th column represents the k -th learned virtual exemplar, $\mu \in \mathbb{R}^{n \times K}$ is a learned matrix whose k -th column corresponds to the conditional probability of assigning \mathbf{D}_k to each of the n input data, and \log is applied entrywise. In Eq. 1, $d(\mathbf{D}, \mathcal{I}) \in \mathbb{R}^{K \times n}$ is the matrix of the euclidean distances between input data in \mathcal{I} and the virtual exemplars in \mathbf{D} whereas $f(\mathbf{D}) \in \mathbb{R}^{2 \times K}$ is the softmax output of the learned network f . The first term in Eq. 1 measures how representative are $\{\mathbf{D}_k\}_k$ w.r.t. the unlabeled samples in \mathcal{I} . This term, as defined, has a major advantage compared to [26]; indeed, due to column-stochasticity of μ , the first term in Eq. 1 reaches (more easily) smaller values than its counterpart in [26] (whose μ is row-stochastic), and this allows producing virtual exemplars closer to the input data and thereby inheriting more accurate labels when the oracle annotates their closest input data. The second term in Eq. 1 also captures diversity, not in the virtual exemplars themselves (as in [26]), but instead in *how training data polarize (or attract) the virtual exemplars*; this term reaches its minimum when all the training samples evenly attract the virtual data. The third and fourth terms are similar to [26]. The former measures the *ambiguity* (or uncertainty) in \mathbf{D} as the entropy of the scoring function (it reaches its smallest value when virtual exemplars in \mathbf{D} are evenly scored w.r.t different classes) while the latter acts as a regularizer which favors uniform conditional probability distribution $\{\mu_{ik}\}_i$. Finally, all these terms are mixed using the coefficients $\alpha, \beta, \gamma \geq 0$.

2.3. Optimization

In order to optimize Eq. 1, we consider, for each change detection cycle t , an EM-like procedure that first fixes the display \mathbf{D} and solves the optimization problem w.r.t. μ as shown in the following proposition.

Proposition 1 The optimality conditions of Eq. (1) lead to

$$\mu^{(\tau+1)} := \hat{\mu}^{(\tau+1)} \text{diag}(\mathbf{1}_n^\top \hat{\mu}^{(\tau+1)})^{-1}, \quad (2)$$

being $\hat{\mu}^{(\tau+1)}$

$$\exp \left\{ -\frac{1}{\gamma} [d(\mathcal{I}, \mathbf{D}^{(\tau)}) + \frac{\alpha}{K} (\mathbf{1}_n + \log \frac{1}{K} \mu^{(\tau)} \mathbf{1}_K) \mathbf{1}_K^\top] \right\}, \quad (3)$$

here $\text{diag}(\cdot)$ maps a vector to a diagonal matrix.

Due to space limitation, details of the proof are omitted and result from the optimality conditions of Eq. 1’s gradients. Once μ optimized (and fixed) the problem is solved w.r.t. \mathbf{D} using stochastic gradient descent. A loss $\mathcal{L}(\mathbf{D})$ is composed of the first and the third term of Eq. 1, then its gradient (w.r.t. \mathbf{D}) is back-propagated through the GCN layers till the input (virtual data) of the learned GCNs using the chain rule. Note that $\mu^{(0)}$ and $\mathbf{D}^{(0)}$ are initially set to random values and, in practice, the procedure converges to an optimal solution (denoted as $\tilde{\mu}, \tilde{\mathbf{D}}$) in few iterations. This solution defines the most *relevant* virtual exemplars of \mathcal{D}_{t+1} (according to criteria in Eq. 1) which are used to train the subsequent classifier f_{t+1} .

3. EXPERIMENTS

Change detection experiments are conducted on the Jefferson dataset. The latter includes 2,200 non-overlapping aligned patch pairs (of 30×30 RGB pixels each) taken from bi-temporal GeoEye-1 satellite images of $2,400 \times 1,652$ pixels with a spatial resolution of 1.65m/pixel. These images have been taken from the area of Jefferson (Alabama) in 2010 and 2011 and correspond to many changes (building destruction, etc.) due to tornadoes as well as no-changes (including irrelevant ones as clouds, occlusions, etc). The ground-truth associated to this dataset includes 2,161 negative pairs (no changes and irrelevant ones) and only 39 positive pairs (relevant changes), so more than 98% of these data correspond to irrelevant changes, and this makes localizing changes very challenging. In all our experiments, we split the whole dataset evenly; (i) for training (our display and learning models) and (ii) for testing performances. As the two classes (changes/no-changes) are highly imbalanced, we measure performances using the equal error rate (EER) for different sampling percentages. At each iteration t , the sampling percentage corresponds to $(\sum_{k=0}^{t-1} |\mathcal{D}_k| / (|\mathcal{I}|/2)) \times 100$ with $|\mathcal{I}| = 2,200$ and $|\mathcal{D}_k|$ set to 16. Smaller EER implies better performances.

Performances & comparison. We compare our proposed method against other related sampling strategies including fixed-pool (random search, maxmin, uncertainty ...) [36] as well as learned-pool methods [26]. Random consists in selecting samples from the pool of unlabeled data while uncertainty consists in collecting, from the same pool, the

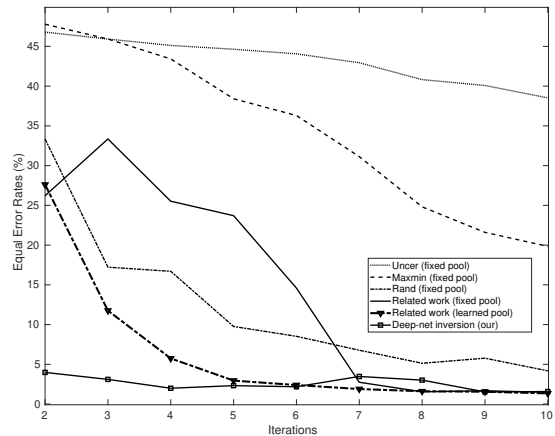


Fig. 1. Comparison of different sampling strategies w.r.t. different iterations and the underlying sampling rates in figure 2. Here Uncer and Rand stand for uncertainty and random sampling respectively. Note that fully-supervised learning achieves an EER of 0.94%. Related works (fixed and learned pools) stand respectively for the methods in [36] and [26].

display whose classifier scores are the most ambiguous (i.e., similar across different classes). Maxmin greedily samples data in \mathcal{D}_{t+1} ; each sample in $\mathbf{x}_i \in \mathcal{D}_{t+1} \subset \mathcal{I} \setminus \cup_{k=0}^t \mathcal{D}_k$ is taken by *maximizing its minimum distance w.r.t. $\cup_{k=0}^t \mathcal{D}_k$* , leading to the most different samples in \mathcal{D}_{t+1} . We further compare our display model against [36] which consists in assigning marginal membership probabilities (instead of conditional ones) to the whole unlabeled set and selecting the display with the highest membership values. We also show the performance of virtual exemplar learning model in [26] which relies on support vector machines (SVMs) and row-stochastic formulation as already discussed in section 2.2. Finally, we show a lower bound on EER performances using a fully supervised setting which builds a unique classifier on top of the full training set whose annotation is taken from the ground-truth. The EER performances of all these settings are shown in Figure 1 through different iterations and sampling rates, as well as the aforementioned display selection strategies. From the EERs, most of the comparative methods are powerless to spot the change class sufficiently well; indeed, random and maxmin capture diversity during early change detection iterations but they are less effective later, and uncertainty lacks diversity. The comparative fixed-pool work [36] (which mixes diversity, representativity and uncertainty on a fixed pool) captures diversity and refines better change detections, however, it suffers from the rigidity of the sampled data (especially at the early iterations), while the comparative learned-pool method [26] (which mixes the same terms, and learns pools of data with SVMs and row-stochastic memberships) is more flexible, but it is less effective than our proposed deep-net inversion method particularly at highly frugal regimes.

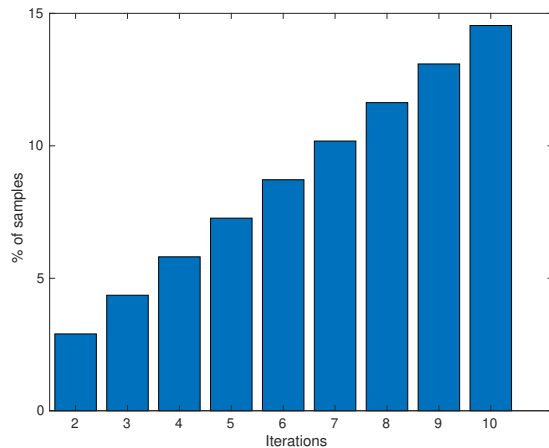


Fig. 2. This figure shows sample rates for each iteration of active learning.

4. CONCLUSION

We introduce in this paper a satellite image change detection algorithm based on active learning. The particularity of our contribution resides in a new adversarial learning model based on deep neural network (DNN) inversion. The latter allows to flexibly train small sets of critical data (dubbed as virtual exemplar) whose labels are frugally obtained from the oracle. These virtual exemplars correspond to the most diverse, representative and uncertain data that challenge (the most) the current DNNs leading to better re-estimate of these DNNs at the subsequent iterations of active learning. Extensive experiments show the out-performance of our virtual exemplar learning model against different baselines, including fixed and learned pools, as well as the related work.

5. REFERENCES

- [1] P. Vo and H. Sahbi. "Transductive kernel map learning and its application to image annotation." *BMVC*. 2012.
- [2] D. Brunner, G. Lemoine, and L. Bruzzone, Earthquake damage assessment of buildings using vhr optical and sar imagery, *IEEE TGRS*, 2010.
- [3] H. Gokon et al. A method for detecting buildings destroyed by the 2011 tohoku earthquake and tsunami using multitemporal terrasars-x data, *GRSL*, vol. 12, no. 6, pp. 1277–1281, 2015.
- [4] Q. Oliveau and H. Sahbi. "Learning attribute representations for remote sensing ship category classification." *IEEE JSTARS* 10(6) 2017.
- [5] J. Deng et al. PCA-based land-use change detection and analysis using multitemporal and multisensor satellite data, *IJRS*, 29 (16), 2008.
- [6] R. Radke, S. Andra, O. Al-Kofahi, and B. Roysam, Image change detection algorithms: A systematic survey, *IEEE TIP*, 14(3), 2005.
- [7] S. Liu, L. Bruzzone, F. Bovolo, M. Zanetti, and P. Du, Sequential spectral change vector analysis for iteratively discovering and detecting multiple changes in hyperspectral images, *TGRS*, 53(8), pp. 4363–4378, 2015.
- [8] G. Chen, G. J. Hay, L. M. Carvalho, and M. A. Wulder, Object-based change detection, *IJRS*, vol. 33, no. 14, pp. 4434–4457, 2012.
- [9] H. Sahbi. "Interactive satellite image change detection with context-aware canonical correlation analysis." *IEEE GRSL*, (14)5, 2017.
- [10] J. Zhu et al. Reducing mis-registration and shadow effects on change detection in wetlands, *PERS*, vol. 77, no. 4, pp. 325–334, 2011.
- [11] A. Fournier, P. Weiss, L. Blanc-Fraud, and G. Aubert, A contrast equalization procedure for change detection algorithms: applications to remotely sensed images of urban areas, In *ICPR*, 2008
- [12] H. Sahbi. "Relevance feedback for satellite image change detection." *IEEE ICASSP*, 2013.
- [13] Carlotto, Detecting change in images with parallax, In *Society of Photo-Optical Instrumentation Engineers*, 2007
- [14] S. Leprince et al. Automatic and precise orthorectification, coregistration, and subpixel correlation of satellite images, application to ground deformation measurements, *TGRS*, vol. 45, no. 6, pp. 1529–1558, 2007.
- [15] Pollard, Comprehensive 3d change detection using volumetric appearance modeling, Phd, Brown University, 2009.
- [16] A. A. Nielsen, The regularized iteratively reweighted mad method for change detection in multi-and hyperspectral data, *IEEE Transactions on Image processing*, vol. 16, no. 2, pp. 463–478, 2007.
- [17] C. Wu, B. Du, and L. Zhang, Slow feature analysis for change detection in multispectral imagery, *TGRS*, vol. 52, no. 5, pp. 2858–2874, 2014
- [18] N. Bourdis, D. Marraud and H. Sahbi. "Constrained optical flow for aerial image change detection." in *IEEE IGARSS*, 2011.
- [19] J. Im, J. Jensen, and J. Tullis, Object-based change detection using correlation image analysis and image seg, *IJRS*, 29(2), 399–423, 2008.
- [20] N. Bourdis, D. Marraud, and H. Sahbi, Spatio-temporal interaction for aerial video change detection, in *IGARSS*, 2012, pp. 2253–2256
- [21] Vinyals et al., Matching networks for one shot learning. 2016.
- [22] Dasgupta, Sanjoy. "Analysis of a greedy active learning strategy." *Advances in neural information processing systems* 17 (2004).
- [23] H. Sahbi. "Coarse-to-fine deep kernel networks." *IEEE ICCV-W*, 2017.
- [24] Burr, Settles. "Active learning." *Synthesis Lectures on Artificial Intelligence and Machine Learning* 6.1 (2012).
- [25] He, Tianxu, et al. "An active learning approach with uncertainty, representativeness, and diversity." *The Scientific World Journal* (2014).
- [26] H. Sahbi and S. Deschamps. "Learning Virtual Exemplars for Label-Efficient Satellite Image Change Detection." In *IEEE IGARSS*, pp. 1197-1200, 2022.
- [27] Joshi et al., Multi-class active learning for image classification. 2009.
- [28] Settles & Craven. An analysis of active learning strategies for sequence labeling tasks. 2008.
- [29] Houthby et al., Bayesian active learning for classification and preference learning. 2011.
- [30] H. Sahbi. "Learning connectivity with graph convolutional networks." In *ICPR* 2021.
- [31] Campbell & Broderick, Automated scalable Bayesian inference via Hilbert coresets. 2019.
- [32] Gal et al., Deep bayesian active learning with image data. 2017
- [33] Pang, Kunkun, et al. "Meta-learning transferable active learning policies by deep reinforcement learning." *arXiv preprint arXiv:1806.04798* (2018).
- [34] M. Jiu and H. Sahbi. "Laplacian deep kernel learning for image annotation." *IEEE ICASSP*, 2016.
- [35] A. Kolesnikov, X. Zhai, L. Beyer. Revisiting Self-Supervised Visual Representation Learning. *IEEE/CVF CVPR*, 2019, pp. 1920-1929
- [36] H. Sahbi, S. Deschamps, A. Stoian. Frugal Learning for Interactive Satellite Image Change Detection. *IEEE IGARSS*, 2021.
- [37] Creswell, Antonia, et al. "Generative adversarial networks: An overview." *IEEE Signal Processing Magazine* 35.1 (2018): 53-65.
- [38] N. Bourdis, D. Marraud and H. Sahbi. "Camera pose estimation using visual servoing for aerial video change detection." *IEEE IGARSS* 2012.
- [39] H. Sahbi, "Misalignment resilient cca for interactive satellite image change detection." 2016 23rd International Conference on Pattern Recognition (ICPR). IEEE, 2016.