ADVERSARIAL LABEL-EFFICIENT SATELLITE IMAGE CHANGE DETECTION

Hichem Sahbi, Sebastien Deschamps

ADVERSARIAL LABEL-EFFICIENT SATELLITE IMAGE CHANGE DETECTION

Hichem Sahbi[1] *Sebastien Deschamps*[1,2]

[1]Sorbonne University, CNRS, LIP6, Paris, France [2]Theresis Thales, France

ABSTRACT

Satellite image change detection aims at finding occurrences of targeted changes in a given scene taken at different instants. This task is highly challenging due to the acquisition conditions and also to the subjectivity of changes. In this paper, we investigate satellite image change detection using active learning. Our method is interactive and relies on a question & answer model which asks the oracle (user) questions about the most informative display (dubbed as virtual exemplars), and according to the user's responses, updates change detections. The main contribution of our method consists in a novel adversarial model that allows frugally probing the oracle only with the most representative, diverse and uncertain virtual exemplars. The latter are learned to challenge (the most) the trained change decision criteria which ultimately leads to a better re-estimate of these criteria in the following iterations of active learning. Conducted experiments show the out-performance of our proposed adversarial display model against related display strategies as well as the related work.

Index Terms— active learning, adversarial virtual exemplars, satellite image change detection

1. INTRODUCTION

Automatic satellite image change detection consists in finding occurrences of *relevant* changes in a given area at an instant t_1 w.r.t. the same area at an earlier instant t_0. This task is useful in different applications, particularly damage assessment after natural hazards, essential to prioritize disaster response and rescues [2, 3]. Change detection is also known to be highly challenging as observed scenes are subject to many irrelevant changes due to sensors, radiometric variations and shadows, weather conditions as well as scene content. Early change detection solutions were based on simple comparisons of multi-temporal signals, via image differences and thresholding, vegetation indices, principal component and change vector analyses [5–8]. Other solutions require preliminary preprocessing techniques that mitigate the effect of irrelevant variations by correcting radiometric changes, occlusions, and by estimating the parameters of sensors for registration, etc. [10, 11, 13, 14, 37] or consider these variations as a part of statistical appearance modeling [9, 12, 15–20].

Among the existing change detection methods, those based on statistical machine learning are particularly effective. However, the success of these methods is bound to the availability of large labeled training sets that comprehensively capture the huge variability in irrelevant changes as well as the user's targeted relevant changes. In practice, labeled sets are scarce [1, 4] and even when available, their labeling may not reflect the user's subjectivity and intention. Several alternative solutions seek to make machine learning methods frugal and less labeled-data hungry [23, 33] including few shot [21] and self-supervised learning [34]; however, these methods are agnostic to the users' intention. Other solutions, based on active learning [22, 24–27, 29–32], are rather more suitable where users label very few examples of relevant and irrelevant changes according to their intention, prior to train and retrain user-dedicated change detection criteria.

In this paper, we devise a novel iterative satellite image change detection solution based on a question & answer model that frugally queries the intention of the user (oracle), and updates change detection results accordingly. These queries are restricted *only* to the most informative subset of exemplars (also referred to as virtual displays) which are *learned* instead of being sampled from the *fixed* pool of unlabeled data. The informativeness of exemplars is modeled with a conditional probability distribution that measures how relevant is a given exemplar in the learned displays given the pool of unlabeled data. Two novel adversarial losses are also used in order to learn the display — with the most diverse, representative, and ambiguous exemplars — that challenge (the most) the previously trained change detection criteria resulting into a better re-estimation of these criteria at the subsequent iterations of change detection. While being adversarial, the two losses are conceptually different from generative adversarial networks (GANs) [36]; GANs aim at producing fake data that mislead the trained discriminators whilst our adversarial losses seek to generate the most informative data for further annotations. In other words, the proposed framework allows to sparingly probe the oracle only on the most representative, diverse and uncertain exemplars that challenge the most the current discriminator, and eventually lead to more accurate ones in the following iterations of change detection. Finally, change detection experiments corroborate these findings and show the effectiveness of our *exemplar and display learning* models against comparative methods.

2. PROPOSED METHOD

Let's consider $\mathcal{I} = \{\mathbf{x}_i = (p_i, q_i)\}_{i=1}^n \subset \mathbb{R}^d$ as a collection of aligned patch pairs taken from two satellite images captured at two different instants t_0, t_1, and let $\mathcal{Y} = \{\mathbf{y}_1, \ldots, \mathbf{y}_n\} \subset \{0, 1\}$ be the underlying unknown labels. Our goal is to train a classifier $f : \mathcal{I} \to \{0, 1\}$ that predicts the unknown labels in $\{\mathbf{y}_i\}_i$ with $\mathbf{y}_i = 1$ if the patch q_i corresponds to a "change" w.r.t. the underlying patch p_i, and $\mathbf{y}_i = 0$ otherwise. Training f requires a subset of hand-labeled data obtained from an oracle. As obtaining these labels is usually highly expensive, the design of f should be *label-frugal* while being as accurate as possible.

2.1. Interactive change detection

Our change detection algorithm is built upon a question & answer iterative process which consists in (i) submitting the *most informative patch pairs* to query their labels from an oracle, and (ii) updating a classifier f accordingly. In the sequel of this paper, the subset of informative images is dubbed as *display*. Let $\mathcal{D}_t \subset \mathcal{I}$ be a display shown to the oracle at iteration t, and $\mathcal{Y}_t$ its unknown labels; considering a random display $\mathcal{D}_0$, we train our change detection criteria iteratively (for $t \in \{0, \ldots, T-1\}$) according to the subsequent steps
1/ Probe the oracle with $\mathcal{D}_t$ to obtain $\mathcal{Y}_t$, and train a support vector machine (SVM) $f_t(.)$ on $\cup_{k=0}^t (\mathcal{D}_k, \mathcal{Y}_k)$; this SVM is trained on top of graph convolution network features [28].
2/ Select $\mathcal{D}_{t+1} \subset \mathcal{I} \backslash \cup_{k=0}^t \mathcal{D}_k$; a strategy that *bruteforces* all the possible displays $\mathcal{D} \subset \mathcal{I} \backslash \cup_{k=0}^t \mathcal{D}_k$, trains the associated classifiers $f_{t+1}(.)$ on $\mathcal{D} \cup_{k=0}^t \mathcal{D}_k$, and maintains the display $\mathcal{D}$ with the highest accuracy is combinatorial and intractable. Besides, collecting labels on each of these displays is also out of reach. In this work, we consider instead display selection strategies, based on active learning, which are rather more tractable; nonetheless, these strategies should be carefully designed as many of them are equivalent to (or worse than) basic strategies that select data uniformly randomly [24].
Our proposed display model, as the main contribution of this paper (shown mainly in section 2.3), is different from usual sampling strategies (see e.g. [35]), and it *flexibly* synthesizes exemplars (also referred to as virtual exemplars) by maximizing diversity, representativity and uncertainty. Diversity seeks to hallucinate exemplars that allow exploring the uncharted parts of unlabeled data whereas representativity makes it possible to register the exemplars, as much as possible, to the input data. Finally, ambiguity locally refines the boundaries of the trained classifiers $f_{t+1}(.)$.

2.2. Virtual exemplar learning: early model

We consider for each sample $\mathbf{x}_i \in \mathcal{I}$ a membership distribution $\{\mu_{ik}\}_{k=1}^K$ that measures the *conditional* probability of assigning $\mathbf{x}_i$ to K-virtual exemplars; the latter constitute the subsequent display $\mathcal{D}_{t+1}$. In contrast to the method in [35] — which relies on ranking values of a *marginal* probability distribution in order to define $\mathcal{D}_{t+1}$ — the early model [1] used in this paper neither requires ranking nor hard thresholding of the memberships $\{\mu_{ik}\}_{k=1}^K$ in order to define $\mathcal{D}_{t+1}$; instead, $\mathcal{D}_{t+1}$ (rewritten for short and in a matrix form as $\mathbf{D}$) together with $\{\mu_{ik}\}_{i,k}$, are found by minimizing

$$\begin{aligned}\min_{\mathbf{D};\mu\in\Omega} \quad & \mathbf{tr}\big(\mu\, d(\mathbf{D},\mathbf{X})\big) \;+\; \alpha\, \big[\tfrac{1}{n}\mathbf{1}_n^\top \mu\big] \log \big[\tfrac{1}{n}\mathbf{1}_n^\top \mu\big]^\top \\ & + \beta\, \mathbf{tr}\big(f(\mathbf{D})^\top \log f(\mathbf{D})\big) \;+\; \gamma\, \mathbf{tr}(\mu^\top \log \mu),\end{aligned} \tag{1}$$

here $\Omega = \{\mu : \mu \geq 0; \mu \mathbf{1}_K = \mathbf{1}_n\}$, $\mathbf{1}_K$, $\mathbf{1}_n$ denote two vectors of K and n ones respectively, $^\top$ is the matrix transpose operator, $\mu \in \mathbb{R}^{n\times K}$ is a learned matrix whose i-th row corresponds to the conditional probability of assigning $\mathbf{x}_i$ to each of the K-virtual exemplars, and $\log$ is applied entrywise. In Eq. 1, $d(\mathbf{D}, \mathbf{X}) \in \mathbb{R}^{K\times n}$ is the matrix of the euclidean distances between input data in $\mathbf{X}$ (i.e., in $\mathcal{I}$) and the virtual exemplars in $\mathbf{D}$ whereas $f(\mathbf{D}) \in \mathbb{R}^{2\times K}$ is a scoring matrix whose columns correspond to the SVM $f_t(.)$ (and its complement) applied to the K exemplars. The first term in Eq. 1, rewritten as $\sum_i \sum_k \mu_{ik} \|\mathbf{x}_i - \mathbf{D}_k\|_2^2$, measures the representativity of the virtual exemplars, and seeks to align the latter with their closest data in $\mathcal{I}$. The second term (equal to $\sum_k [\frac{1}{n}\sum_{i=1}^n \mu_{ik}] \log[\frac{1}{n}\sum_{i=1}^n \mu_{ik}]$) captures the *diversity* of the generated virtual data as the entropy of the probability distribution of the underlying memberships; this measure is minimal when input data are assigned to different virtual exemplars, and vice-versa. The third criterion (equivalent to $\sum_k \sum_c [f_c(\mathbf{D})]_k^\top [\log f_c(\mathbf{D})]_k$) measures the *ambiguity* (or uncertainty) in $\mathcal{D}_{t+1}$ as the entropy of the scoring function; it reaches its smallest value when exemplars in $\mathcal{D}_{t+1}$ are evenly scored w.r.t different classes. Finally, the fourth term acts as a regularizer, and equality and inequality constraints guarantee that the memberships $\{\mu_{ik}\}_k$ form a probability distribution.

2.3. Virtual exemplar learning: surrogate model

The formulation proposed in the aforementioned section, in spite of being relatively effective (as shown later in experiments), has a major drawback: it combines heterogeneous terms (entropy and distance based criteria) whose mixing hyperparameters are difficult to optimize[1]. In what follows, we consider a surrogate objective function which considers only homogeneous (distance based) criteria excepting the regularizer, and this turns out to be highly effective as shown later in experiments. With this variant, the virtual exemplars together with their distributions $\{\mu_{ik}\}_{i,k}$ are now obtained as

$$\begin{aligned}\min_{\mathbf{D};\mu\in\Omega} \quad & \mathbf{tr}\big(\mu\, d(\mathbf{D},\mathbf{X})\big) \;+\; \tfrac{\alpha}{2}\, \big\|\tfrac{1}{n}\mathbf{1}_n^\top \mu - \tfrac{1}{K}\mathbf{1}_K^\top\big\|_F^2 \\ & + \tfrac{\beta}{2}\, \big\|f(\mathbf{D}) - \tfrac{1}{2}\mathbf{1}_2\mathbf{1}_K^\top\big\|_F^2 \;+\; \gamma\, \mathbf{tr}(\mu^\top \log \mu),\end{aligned} \tag{2}$$

here $\|.\|_F^2$ is the Frobenius norm. The first and fourth terms of the above objective function remain unchanged while the second and third terms, despite being different, their impact

[1] This requires trying many hyperparameteres, inducing the underlying displays, and labeling them by the oracle. This is intractable so labeling should be sparingly achieved.

is strictly equivalent *when considered separately*. Indeed, the second term, equal to $\sum_k (\frac{1}{n}\sum_i \mu_{ik} - \frac{1}{K})^2$, aims at assigning the same probability mass to the virtual exemplars (and this maximizes entropy) whilst the third term, equal to $\sum_k \sum_c (f_c(\mathbf{D}_k) - \frac{1}{2})^2$, has also an equivalent behavior w.r.t. its counterpart in Eq. 1, namely, it favors virtual exemplars whose SVM scores are the closest to $\frac{1}{2}$ (i.e., near the SVM decision boundary), and this also maximizes entropy. This similar behavior is also corroborated through the observed performances of these terms when taken individually[2]; nevertheless, the joint combination of these homogeneous terms (through Eq. 2) provides a more pronounced gain compared to their heterogeneous counterparts (in Eq. 1).

Proposition 1 *The optimality conditions of Eqs. 1, 2 lead to*

$$\begin{array}{rcl} \mu^{(\tau+1)} & := & \mathbf{diag}\big(\hat{\mu}^{(\tau+1)}\mathbf{1}_K\big)^{-1}\,\hat{\mu}^{(\tau+1)}, \\ \mathbf{D}^{(\tau+1)} & := & \hat{\mathbf{D}}^{(\tau+1)}\,\mathbf{diag}\big(\mathbf{1}_n'\mu^{(\tau)}\big)^{-1}, \end{array} \tag{3}$$

with $\hat{\mu}^{(\tau+1)}$, $\hat{\mathbf{D}}^{(\tau+1)}$ *being respectively for Eq. (1)*

$$\begin{array}{l} \exp\big(-\frac{1}{\gamma}[d(\mathbf{X},\mathbf{D}^{(\tau)}) + \frac{\alpha}{n}\mathbf{1}_n(\mathbf{1}_K^\top + \log\frac{1}{n}\mathbf{1}_n^\top\mu^{(\tau)})]\big) \\ \mathbf{X}\,\mu^{(\tau)} + \beta\sum_c \nabla_v f_c(\mathbf{D}^{(\tau)}) \circ (\mathbf{1}_d\,[\log f_c(\mathbf{D}^{(\tau)})]' + \mathbf{1}_d\mathbf{1}_K'), \end{array} \tag{4}$$

and for Eq. (2)

$$\begin{array}{l} \exp\big(-\frac{1}{\gamma}[d(\mathbf{X},\mathbf{D}^{(\tau)}) + \frac{\alpha}{n}\mathbf{1}_n(\frac{1}{n}\mathbf{1}_n^\top\mu^{(\tau)} - \frac{1}{K}\mathbf{1}_K^\top)]\big) \\ \mathbf{X}\,\mu^{(\tau)} + \beta\sum_c \nabla_v f_c(\mathbf{D}^{(\tau)}) \circ (\mathbf{1}_d\,[f_c(\mathbf{D}^{(\tau)}) - \frac{1}{2}\mathbf{1}_K^\top]), \end{array} \tag{5}$$

here $\circ$ *stands for the Hadamard product and* $\mathbf{diag}(.)$ *maps a vector to a diagonal matrix.*

In view of space, details of the proof are omitted and result from the optimality conditions of Eq. 1's and Eq. 2's gradients. 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 subsequent display $\mathcal{D}_{t+1}$ used to train f_{t+1}. Note also that α and β are set to make the impact of the underlying terms equally proportional, and this corresponds to $\alpha = \frac{1}{K}$ and $\beta = \frac{1}{2\times K}$. Finally, since γ acts as scaling factor that controls the shape of the exponential function, its setting is iteration-dependent and proportional to the input of that exponential (i.e., $\log(\hat{\mu}^{(\tau+1)})$), so in practice $\gamma = \frac{1}{nK}\|\log(\hat{\mu}^{(\tau+1)})\|_1$.

3. EXPERIMENTS

Change detection experiments are conducted on the Jefferson dataset. The latter includes $2,200$ 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 were taken from the area of Jefferson (Alabama) in 2010 and 2011 with many changes (building destruction, etc.) due to tornadoes as well as no-changes (including irrelevant ones as clouds). This dataset includes 2,161 negative pairs (no/irrelevant changes) and only 39 positive pairs (relevant changes), so less than 2% of these data correspond to relevant changes, and this makes their localization very challenging. In our experiments, we split the whole dataset evenly; one half to train our display and learning models while the remaining half to measure accuracy. As changes/no-changes classes are highly off-balanced, we measure accuracy using the equal error rate (EER) for different sampling percentages defined — at each iteration t — as $(\sum_{k=0}^{t-1}|\mathcal{D}_k|/(|\mathcal{I}|/2)) \times 100$ with $|\mathcal{I}| = 2,200$ and $|\mathcal{D}_k|$ set to 16. Smaller EER implies better performances.

rep	div	amb	1	2	3	4	5	6	7	8	9	10	AUC.
✗	✗	✓	47.81	27.29	11.15	7.97	8.18	7.31	7.97	7.94	7.50	7.90	14.10
			47.81	18.71	11.23	7.96	8.17	7.28	7.58	7.88	7.49	7.90	13.21
✗	✓	✗	47.81	18.72	11.24	7.97	8.18	7.29	7.59	7.88	7.50	7.90	13.21
			47.81	18.71	11.23	7.96	8.17	7.28	7.58	7.88	7.49	7.90	13.21
✓	✗	✗	47.81	35.98	16.86	6.52	4.98	2.67	2.03	1.80	1.45	1.30	12.14
			47.81	36.75	29.26	8.61	4.27	2.37	2.34	1.68	1.45	1.27	13.59
✓	✗	✓	47.81	40.40	23.86	9.56	7.65	5.75	5.47	6.12	4.40	5.72	15.67
			47.81	36.75	29.26	8.61	4.27	2.37	2.34	1.68	1.45	1.27	13.59
✗	✓	✓	47.81	27.29	11.15	7.97	8.18	7.31	7.97	7.94	7.50	7.90	14.10
			47.81	28.94	12.39	9.12	7.05	6.94	7.05	7.09	7.25	6.93	14.06
✓	✓	✗	47.81	29.84	17.63	6.21	4.40	2.70	1.98	1.92	1.65	1.52	11.57
			47.81	38.20	23.73	9.36	7.67	5.67	4.66	4.31	3.28	2.59	14.73
✓	✓	✓	47.81	27.61	11.76	5.74	2.95	2.39	1.89	1.61	1.55	1.34	10.47
			47.81	32.56	**9.88**	**4.54**	**2.71**	**2.00**	**1.56**	**1.21**	**1.10**	**1.08**	**10.44**
Samp%			1.45	2.90	4.36	5.81	7.27	8.72	10.18	11.63	13.09	14.54	-

Table 1. This table shows an ablation study of our display model. Here rep, amb and div stand for representativity, ambiguity and diversity respectively. These results are shown for different iterations $t = 0, \ldots, T-1$ (Iter) and the underlying sampling rates (Samp) again defined as $(\sum_{k=0}^{t-1}|\mathcal{D}_k|/(|\mathcal{I}|/2)) \times 100$. The AUC (Area Under Curve) corresponds to the average of EERs across iterations.

Ablation Study & Comparison. We first study the impact of each term of our objective functions separately, and then we consider them jointly and all combined. Note that the regularization term is always kept as it allows to obtain the closed form solutions shown in proposition 1. Table 1 illustrates the impact of these terms where EER performances are shown for each configuration using both the early and the surrogate models respectively. As expected, when using the terms separately, their impact on change detection performances are very similar (and sometimes identical), and this results from their equivalence. However, when jointly combining these terms, their impact is different, and this results from their heterogeneity in Eq. 1 and homogeneity in Eq. 2 which makes the setting of the underlying mixing hyper-parameters in the latter configuration more effective. From these results, we also observe the highest impact of representativity+diversity especially at the earliest iterations of change detection, whilst the impact of ambiguity term raises later in order to locally refine the decision criteria. Finally, we compare our model against fixed-pool strategies including random search, maxmin and uncertainty as well as a fixed-pool combination of representativity, diversity and uncertainty[3]. From figure 1, most of the comparative methods are powerless to spot the change class sufficiently well. Indeed, while random and maxmin capture diversity during early iterations of change detection, they are less effective in refining the learned

[2]This equivalence of performances is also due to the fact that no tuning is necessary for the underlying weights when terms are used individually.

[3]Due to limited space, details about comparative strategies can be found in [35].

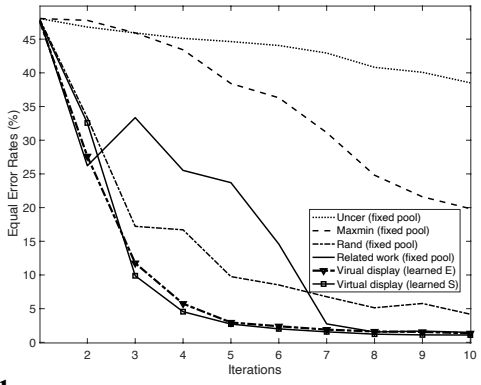

Fig. 1. This figure shows a comparison of different sampling strategies w.r.t. different iterations (Iter) and the underlying sampling rates in table 1 (Samp). Here Uncer and Rand stand for uncertainty and random sampling respectively. Note that fully-supervised learning achieves an EER of 0.94%. Related work stands for the method in [35] while "learned E and S" stand respectively for the early model in section 2.2 and the new surrogate model in section 2.3.

decision functions whilst uncertainty overcomes this issue, it lacks diversity. The fixed-pool combination [35] captures diversity and refines better the learned classifiers, however, it suffers from rigidity of the selected data. In contrast, our proposed learned-pool method (particularly the surrogate model) leads to an extra gain in performances at high frugal regimes.

4. CONCLUSION

We introduce in this paper a new interactive satellite image change detection algorithm. The proposed solution is based on active learning and consists in training both a classifier and a subset of data (referred to as virtual display) resulting into more flexible model. These virtual displays are learned by maximizing their representativity, diversity and uncertainty, and this provides a more effective adversarial model that challenges the current learned classifiers leading to more effective subsequent ones. All these findings are corroborated through extensive experiments conducted on the challenging task of interactive satellite image change detection which show the out-performance of the proposed virtual display models against different baselines as well as the related work.

5. REFERENCES

[1] H. Sahbi and S. Deschamps. "Learning Virtual Exemplars for Label-Efficient Satellite Image Change Detection." In IEEE IGARSS, 2022.

[2] D. Brunner, G. Lemoine, and L. Bruzzone, Earthquake damage assessment of buildings using vhr optical and sar imagery, IEEE Trans. Geosc. Remote Sens., vol. 48, no. 5, pp. 2403–2420, 2010.

[3] H. Gokon et al. A method for detecting buildings destroyed by the 2011 tohoku earthquake and tsunami using multitemporal terrasar-x data, GRSL, vol. 12, no. 6, pp. 1277–1281, 2015.

[4] Q. Oliveau and H. Sahbi. "Learning attribute rep for remote sensing ship category classification." IEEE JSTARS 10.6 (2017): 2830-2840.

[5] J. Deng, K. Wang, Y. Deng, and G. Qi, PCA-based land-use change detection and analysis using multitemporal and multisensor satellite data, IJRS, vol. 29, no. 16, pp. 4823–4838, 2008.

[6] R. Radke, S. Andra, O. Al-Kofahi, and B. Roysam, Image change detection algorithms: A systematic survey, IEEE Trans. on Im Proc, vol. 14, no. 3, pp. 294–307, 2005.

[7] S. Liu et al. Sequential spectral CVA for iteratively discovering and detecting multiple changes in hyperspectral images, TGRS, 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 et al. A contrast equalization procedure for change detection algorithms: applications to RS 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 TIP, 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] Joshi et al., Multi-class active learning for image classification. 2009.

[27] Settles & Craven. An analysis of active learning strategies for sequence labeling tasks. 2008.

[28] H. Sahbi. "Learning connectivity with graph convolutional networks." In ICPR 2021.

[29] Houlsby et al., Bayesian active learning for classification and preference learning. 2011.

[30] Campbell & Broderick, Automated scalable Bayesian inference via Hilbert coresets. 2019.

[31] Gal et al., Deep bayesian active learning with image data. 2017

[32] Pang, Kunkun, et al. "Meta-learning transferable active learning policies by deep reinforcement learning." arXiv:1806.04798 (2018).

[33] M. Jiu and H. Sahbi. "Laplacian deep kernel learning for image annotation." IEEE ICASSP, 2016.

[34] A. Kolesnikov, X. Zhai, L. Beyer. Revisiting Self-Supervised Visual Representation Learning. IEEE/CVF CVPR, 2019, pp. 1920-1929

[35] H. Sahbi, S. Deschamps, A. Stoian. Frugal Learning for Interactive Satellite Image Change Detection. IEEE IGARSS, 2021.

[36] Creswell, Antonia, et al. "Generative adversarial networks: An overview." IEEE Signal Processing Magazine 35.1 (2018): 53-65.

[37] N. Bourdis, D. Marraud and H. Sahbi. "Camera pose estimation using visual servoing for aerial video change detection." IEEE IGARSS 2012.