2016 23rd International Conference on Pattern Recognition (ICPR)
Cancún Center, Cancún, México, December 4-8, 2016

Misalignment Resilient CCA for Interactive Satellite Image Change Detection

Hichem Sahbi
LTCI CNRS Telecom ParisTech
University Paris-Saclay
Email: hichem.sahbi@telecom-paristech.fr

***Abstract*—Change detection, in multi-temporal satellite imagery, seeks to discover relevant changes and to discard irrelevant ones. This task is usually achieved by modeling accurate decision criteria that capture the user's intention while being resilient to many irrelevant changes including acquisition conditions. Among existing change detection solutions, correlation-based models – such as canonical correlation analysis (CCA) – are particularly successful, but their success is very dependent on the quality of alignments used to train these models.**

In this paper, we introduce a novel interactive change detection algorithm based on a new variant of CCA, referred to as misalignment resilient CCA. Given a small sample of "changes" and "no-changes" labeled by an oracle (user), our method learns transformation matrices that map these data from different input spaces, related to multi-temporal images, into a common latent space which is sensitive to relevant changes while being resilient to irrelevant ones including misalignments. These CCA transformations correspond to the optimum of a particular constrained maximization problem that mixes a new soft-alignment term and a context-based regularization criterion. Extensive experiments conducted in interactive satellite image change detection, show that our misalignment resilient CCA approach is highly effective.

I. Introduction

Satellite image change detection seeks to find instances of relevant changes into a given scene acquired at instance t_1 with respect to the same scene taken at instance $t_0 < t_1$. Early change detection solutions, introduced during the 70's are based on straightforward comparisons of multi-temporal series, using image difference thresholding of vegetation indices, principal components and vector change analysis [1], [2]. One of the major applications in change detection is damage assessment after natural hazards [3]; this consists in finding relevant changes (such as building destructions, etc.) while discarding irrelevant ones (due to occlusions, sensor artifacts and alignment errors, etc.). The success of this task is clearly dependent on the ability to i) attenuate the effect of some irrelevant changes by aligning images[1] and correcting their radiometric effects, etc. (see for instance [4], [5], [6]) and also the ability to ii) make change detection resilient to other irrelevant changes by considering them as a part of scene representation and appearance modeling (e.g. [7], [8]).

In spite of their relative success, these two categories of methods are highly limited by the huge variability due to irrelevant changes resulting into several missing detections and false alarms. Indeed, many existing change detection solutions, mainly correlation-based methods (e.g., [9]), rely on a strong assumption that satellite images are precisely aligned[2], but this may not hold in practice; more precisely, usual and efficient approaches for alignments (that avoid expensive and possibly erroneous digital elevation models), are usually powerless to attenuate 3D residual and deformation effects into multi-temporal images (including parallax, oscillations due to waiving objects such as trees, etc.). Furthermore, usual registration methods (either 2D or 3D) may also fail when scenes are subject to many changes, as interest points in these scenes might not be repeatable through multiple images and hence difficult to match using existing alignment algorithms [10], [11]. As a result, one of the major challenges is to build multi-temporal image representations which are robust against irrelevant changes, such as residual misalignments that may still occur even when satellite images are globally – and reasonably – well registered.

Canonical correlation analysis (CCA) is one of the correlation-based techniques that has been applied to change detection; it consists in building latent representations which are robust – at some extent – to irrelevant changes [12]. Generally speaking, CCA is a machine learning technique that learns statistical correlation between aligned data. It has been widely applied to many other pattern recognition and machine learning tasks including image-to-text annotation and multi-view tracking [13], [14]. Its general principle consists in learning transformations that map aligned data from input spaces – related to different data modalities (such as viewpoints, sensors, etc.) – to latent spaces while maximizing their statistical correlation. It is commonly agreed that the accuracy of CCA is highly dependent on the quality of alignments which are usually subject to errors. In this work, we address the issue of misalignment in canonical correlation analysis and its impact on the particular task of change detection in multi-temporal satellite imagery.

In this paper, we propose a new formulation of canonical correlation analysis referred to as misalignment resilient CCA (MR-CCA); the latter learns transformations that map image

[1] In this paper, the terminology alignment and registration refers to the same concept.

[2] Initially aligned or after achieving automatic alignment.

regions (belonging to the same physical locations) from input images to latent spaces while being *sensitive* to relevant changes and *resilient* to irrelevant ones including misalignment. Our formulation is based on the optimization of an objective function mixing two terms; the first one partly similar to standard CCA seeks to maximize the correlation between aligned "no-changes" in satellite image data (taken at different instants and under different acquisition conditions) while at the same time it reduces the correlation between "changes". The second term is a regularization criterion that diffuses the correlations (and also information about alignments) between neighboring data thereby making CCA context-aware. The novelty of the proposed CCA framework also resides in its ability to model uncertainty when considering aligned and non-aligned data, resulting into better performances as shown through experiments in interactive satellite image change detection.

II. Interactive Change Detection at a Glance

Besides irrelevant changes and acquisition conditions, the challenge in change detection also resides in the difficulty to characterize decision criteria as the frontier between targeted and untargeted changes is highly dependent on the user's intention. Considering this issue, we adopt a user dependent change detection strategy based on relevance feedback (RF)[3]; also known as interactive image search. Our RF method is based on a query & answer model that helps users expressing their intentions and finding their targeted changes effectively in few iterations; as shown in this paper, this model enhances the quality of change detection criteria, by adapting them to input images and users, without systematic/tedious parsing of large satellite images.

Define $\mathcal{I}_0 = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$, $\mathcal{I}_1 = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ as two satellite images (referred to as reference and test images respectively) captured at two different instants. We consider a particular order of patches in $\mathcal{I}_0$, $\mathcal{I}_1$ such that the first $\ell \ll n$ patches in $\mathcal{I}_0$, $\mathcal{I}_1$ are labeled and correctly aligned by an oracle (user). Let $\mathcal{Y} = \{\mathbf{y}_1, \ldots, \mathbf{y}_n\}$ be the labels associated to patches in $\mathcal{I}_1$; again, only the first ℓ labels are known. Our goal is to design a change detection algorithm based on relevance feedback that predicts the unknown labels $\{\mathbf{y}_i\}_{i=\ell+1}^{n}$ with $\mathbf{y}_i = +1$ if the patch $\mathbf{v}_i \in \mathcal{I}_1$ corresponds to a "change" w.r.t its underlying reference patch in $\mathcal{I}_0$; and $\mathbf{y}_i = -1$ otherwise. Let $\mathcal{D}_t \subset \mathcal{I}_1$ be a *display* (subset of patches with $|\mathcal{D}_t| = 16 \ll |\mathcal{I}_1|$ in practice) shown to the oracle at iteration t and let $\mathcal{Y}_t$ be the unknown labels of $\mathcal{D}_t$. Starting from a random display $\mathcal{D}_0$ including representative samples in $\mathcal{I}_1$, we build our RF model by asking the oracle about the relevance of changes in $\mathcal{D}_t$ and by running the following steps for $t = 0, \ldots, T-1$ (In practice, $T = 10$)

i) Label display $\mathcal{D}_t$ with a *known-only-by-the-oracle* function (denoted $\mathcal{C}(.)$) and assign $\mathcal{C}(\mathcal{D}_t)$ to $\mathcal{Y}_t$. As our change detection ground-truth is objective, we assume deterministic oracle functions only.

ii) Select the next display $\mathcal{D}_{t+1} \subset \mathcal{I}_1 - \cup_{k=0}^{t}\mathcal{D}_k$ to label by the oracle using two strategies, closely related to active learning [17], [18]: exploration and exploitation. The former selects data in order to discover new modes of our change detection criteria (denoted $\{f_t\}_t$) while the latter locally refines these criteria. Our display selection strategy, seeks a balance between exploration and exploitation. At $t = 0$, we apply exploration, then at each iteration $t \geq 1$, we select the subsequent display $\mathcal{D}_{t+1}$ depending on how good was the previous display $\mathcal{D}_t$. In practice, we either *keep* the previous action (exploration or exploitation) or we *switch* from exploration to exploitation or vice-versa, depending on a score $\mathcal{S}_t = \sum_{\mathbf{v}\in\mathcal{D}_t} \mathbb{1}_{\{\text{sign}[f_{t-1}(\mathbf{v})]\neq\mathcal{C}(\mathbf{v})\}}$. This score measures how informative is the display $\mathcal{D}_t$ obtained using the previous action[4].

iii) Train a classifier $f_t(.)$ on the ℓ data labeled, so far, $\cup_{k=0}^{t}(\mathcal{D}_k, \mathcal{Y}_k)$ to predict unknown labels $\{\mathbf{y}_i\}_{i=\ell+1}^{n}$. In practice, we use LIBSVM [19] with the triangular kernel [20], in order to build $f_t(.)$; this choice was motivated by the good performance of RF when using the triangular kernel compared to many other kernels (see [20], [21]). Note that SVMs are trained on top of *novel (misalignment resilient) CCA features, and this is the main contribution of this work* (see Section III).

III. Canonical Correlation Analysis

Standard CCA (see for instance [13]) finds two projection matrices that map aligned data in $\mathcal{I}_0 \times \mathcal{I}_1$ into a latent space while maximizing their correlation. Let $\mathbf{P}_u$, $\mathbf{P}_v$ denote these projection matrices which respectively correspond to reference and test images. CCA finds these matrices by maximizing $\mathbf{P}'_v\mathbf{C}_{vu}\mathbf{P}_u$, subject to $\mathbf{P}'_u\mathbf{C}_{uu}\mathbf{P}_u = 1$, $\mathbf{P}'_v\mathbf{C}_{vv}\mathbf{P}_v = 1$; here $'$ stands for matrix transpose, $\mathbf{C}_{vu}$ (resp. $\mathbf{C}_{uu}$, $\mathbf{C}_{vv}$) correspond to the interclass (resp. intraclass) covariance matrices of data in $\mathcal{I}_0, \mathcal{I}_1$, and equality constraints control the effect of scaling on the solution. One can show that problem above is equivalent to solving the eigenproblem $\mathbf{C}_{uv}\mathbf{C}_{vv}^{-1}\mathbf{C}_{vu}\mathbf{P}_u = \gamma^2\mathbf{C}_{uu}\mathbf{P}_u$ with $\mathbf{P}_v = \frac{1}{\gamma}\ \mathbf{C}_{vv}^{-1}\mathbf{C}_{vu}\mathbf{P}_u$. In practice, learning these two transformations requires (aligned) "no-changes" in $\mathcal{I}_1 \times \mathcal{I}_0$, i.e., $\{(\mathbf{u}_i, \mathbf{v}_i)\}_{i=1}^{\ell}$ with $\mathbf{y}_i = -1$. A variant (in [12]; see also experiments) has shown better results when combining both "no-changes" and "changes" in a discriminative setting, in order to find these transformation matrices. However, using only labeled patch pairs, is not enough and *better results are obtained when i) using also unlabeled patch pairs and when ii) considering resilience to alignment errors in these pairs* as described subsequently.

A. Misalignment Resilient CCA

As discussed earlier, for each iteration of relevance feedback, we assume that only a very small fraction of patches are precisely aligned and labeled by the oracle; for the remaining

[3]RF has been previously studied mainly for multimedia image retrieval [15] and foreground/background separation [16].

[4]a good action should produce a display to correct as many change detection results as possible thereby better refining the subsequent decision criterion.

patches neither precise alignments[5] nor labels are known. Considering $\mathbf{D}$ as an alignment matrix, we assign for each patch pair (entry of $\mathbf{D}$) an alignment score as follows

- If $\mathbf{v}_j \in \mathcal{I}_1$ is labeled by the oracle then the latter selects a unique patch $\mathbf{u}_i \in \mathcal{I}_0$ which is precisely aligned with $\mathbf{v}_j$; as assumed in Section II, we consider a particular order of patches in $\mathcal{I}_0, \mathcal{I}_1$ such that any aligned patch pair $(\mathbf{u}_i, \mathbf{v}_j)$ satisfies $i = j$ (with $i, j \in \{1, \ldots, \ell\}$). Hence, $\mathbf{D}_{ii} = \pm 1$ depending on whether the oracle assigns $(\mathbf{u}_i, \mathbf{v}_i)$ to the "no-change" or "change" class (i.e., $\mathbf{D}_{ii} = -\mathbf{y}_i$) and $\mathbf{D}_{ki} = \mathbf{D}_{ik} = 0$, $\forall k \in \{1, \ldots, \ell\} \backslash i$.
- If $\mathbf{v}_j \in \mathcal{I}_1$ is unlabeled, then for any patch $\mathbf{u}_i \in \mathcal{I}_0$, the entry $\mathbf{D}_{ij}$ is set using a particular similarity function; here $i, j \in \{\ell + 1, \ldots, n\}$. This similarity takes a high value (close to $+1$) if $\mathbf{u}_i$, $\mathbf{v}_j$ are visually similar and spatially close; otherwise $\mathbf{D}_{ij}$ is set to a small value (close to -1). The way the matrix $\mathbf{D}$ is set makes it possible to model the uncertainty about the alignments of unlabeled patch pairs. Extra details about the setting of $\mathbf{D}$ are given in experiments.

Now we introduce our main contribution; a novel misalignment resilient approach for CCA. Considering a small subset $\{(\mathbf{u}_i, \mathbf{v}_i)\}_i \subset \mathcal{I}_0 \times \mathcal{I}_1$ with known labels $\{\mathbf{y}_i\}_i$, we propose to find the transformation matrices $\mathbf{P}_u$, $\mathbf{P}_v$ as

$$\max_{\mathbf{P}_u, \mathbf{P}_v} \ \text{tr}(\mathbf{U}'\mathbf{P}_u\mathbf{P}_v'\mathbf{V}\mathbf{D}) + \beta \sum_c \text{tr}\Big(\mathbf{U}'\mathbf{P}_u\mathbf{P}_v'\mathbf{V}\mathbf{W}_v^c\mathbf{V}'\mathbf{P}_v\mathbf{P}_u'\mathbf{U}\mathbf{W}_u^{c'}\Big)$$
$$\text{s.t.} \quad \mathbf{P}_u'\mathbf{C}_{uu}\mathbf{P}_u = 1 \quad \text{and} \quad \mathbf{P}_v'\mathbf{C}_{vv}\mathbf{P}_v = 1 \tag{1}$$

Here $\beta \geq 0$, tr is the matrix trace operator, $\mathbf{U}$, $\mathbf{V}$ are two matrices whose columns correspond to all patches in $\mathcal{I}_0$, $\mathcal{I}_1$ respectively. The first term of this objective function (equivalent to $\sum_{i,j} \langle \mathbf{P}_u'\mathbf{u}_i, \mathbf{P}_v'\mathbf{v}_j \rangle \mathbf{D}_{ij}$) aims to maximize the correlation between aligned patch pairs with negative labels (i.e., "no-changes"), while at the same time, it minimizes the correlation between aligned patch pairs with positive labels (i.e., relevant "changes"). Equivalently, this term also seeks to maximize the correlations between highly similar unlabeled pairs (i.e., with high $\mathbf{D}_{ij}$), and minimize the correlations between dissimilar unlabeled pairs (which are likely to correspond to changes). If one considers only labeled data, then this term is strictly equivalent to discriminant CCA (in [12]) and one may show that it can be rewritten as $\mathbf{P}_v'\mathbf{C}_{vu}^{-}\mathbf{P}_u - \lambda\mathbf{P}_v'\mathbf{C}_{vu}^{+}\mathbf{P}_u$ ($\lambda \in \mathbb{R}^+$); with $\mathbf{C}_{vu}^{-}$ (resp.$\mathbf{C}_{vu}^{+}$) being the covariance matrix of negative (resp. positive) data in $\{(\mathbf{u}_i, \mathbf{v}_i)\}_i$. With this discriminative setting, the learned transformations $\mathbf{P}_u$, $\mathbf{P}_v$ generate latent data representations $\phi_u(\mathbf{u}_i) = \mathbf{P}_u\psi_f(\mathbf{u}_i)$, $\phi_v(\mathbf{v}_i) = \mathbf{P}_v\psi_f(\mathbf{v}_i)$[6], which are robust against irrelevant changes (i.e., $\|\phi_u(\mathbf{u}_i) - \phi_v(\mathbf{v}_i)\|_2 \rightsquigarrow 0$ for $\mathbf{y}_i = -1$) while also being sensitive to relevance changes (i.e., $\|\phi_u(\mathbf{u}_i) - \phi_v(\mathbf{v}_i)\|_2$ is large for $\mathbf{y}_i = +1$). This results into a better discrimination between "changes" and "no-changes" (see also experiments).

Context-based regularization. Without regularization, it is clear that the left-hand side term in the above objective function may produce maps with no spatial coherence. Hence, we define a context-based regularization criterion that considers for each patch $\mathbf{u}_i$ (or $\mathbf{v}_i$), an anisotropic (typed) neighborhood system $\{\mathcal{N}_c(i)\}_{c=1}^8$ which corresponds to the eight spatial neighbors of $\mathbf{u}_i$ in a regular grid; for instance when $c = 1$, $\mathcal{N}_1(i)$ corresponds to the top-left neighbors of $\mathbf{u}_i$. Using $\mathcal{N}_c(.)$, we consider $\mathbf{W}_u^c$ as an intrinsic adjacency matrix whose given entry is defined as $\mathbf{W}_{u,i,k}^c = \mathbb{1}_{\{k \in \mathcal{N}_c(i)\}} \times \mathbf{f}(\mathbf{u}_i, \mathbf{u}_k)$ (with $\mathbf{f}(\mathbf{u}_i, \mathbf{u}_k)$ being a function inversely proportional to the distance between visual features of patches in the i^{th} and k^{th} locations). Similarly, we define the adjacency matrices $\{\mathbf{W}_v^c\}_c$ for patches $\{\mathbf{v}_i\}_i$. Using this definition of $\{\mathbf{W}_u^c\}_c$, $\{\mathbf{W}_v^c\}_c$, the right-hand side term of the objective function (1) is strictly equivalent to $\beta \sum_c \sum_{i,j} \langle \mathbf{P}_u'\mathbf{u}_i, \mathbf{P}_v'\mathbf{v}_j \rangle \sum_{k,\ell} \langle \mathbf{P}_u'\mathbf{u}_k, \mathbf{P}_v'\mathbf{v}_\ell \rangle \mathbf{W}_{u,i,k}^c \mathbf{W}_{v,j,\ell}^c$; the latter corresponds to a neighborhood (or context) criterion which considers that a high value of the correlation $\langle \mathbf{P}_u'\mathbf{u}_i, \mathbf{P}_v'\mathbf{v}_j \rangle$, in the learned latent space, should imply high correlation values in the neighborhoods $\{\mathcal{N}_c(i) \times \mathcal{N}_c(j)\}_c$. Hence, this term enhances the robustness of the correlation between patch pairs in the learned latent space. Put differently, if a given patch pair $(\mathbf{u}_i, \mathbf{v}_j)$ is surrounded by highly correlated pairs, then the correlation between $(\mathbf{u}_i, \mathbf{v}_j)$ should be maximized and vice-versa.

Finally, equality constraints in (1) act as normalization factors that control the effect of scaling on the solution and also avoid null projection matrices.

B. Optimization

Considering Lagrange multipliers for the equality constraints in Eq. (1), one may show that optimality conditions (related to the gradient of Eq. (1) w.r.t $\mathbf{P}_u$, $\mathbf{P}_v$ and the Lagrange multipliers) lead to the following generalized eigenproblem

$$\begin{array}{lll} \mathbf{K}_{uv}\mathbf{C}_{vv}^{-1}\mathbf{K}_{vu}\mathbf{P}_u & = & \gamma^2\mathbf{C}_{uu}\mathbf{P}_u \\ \text{with} \quad \mathbf{P}_v & = & \frac{1}{\gamma}\,\mathbf{C}_{vv}^{-1}\mathbf{K}_{vu}\mathbf{P}_u, \end{array} \tag{2}$$

here $\mathbf{K}_{vu} = \mathbf{K}_{uv}'$ and

$$\begin{array}{lll} \mathbf{K}_{vu} = \mathbf{V}\mathbf{D}\mathbf{U}' & + & \beta \sum_c \mathbf{V}\mathbf{W}_v^c\mathbf{V}'\mathbf{P}_v\mathbf{P}_u'\mathbf{U}\mathbf{W}_u^{c'}\mathbf{U}' \\ & + & \beta \sum_c \mathbf{V}\mathbf{W}_v^{c'}\mathbf{V}'\mathbf{P}_v\mathbf{P}_u'\mathbf{U}\mathbf{W}_u^{c}\mathbf{U}'. \end{array} \tag{3}$$

In practice, we solve the above eigenproblems iteratively. For each iteration τ, we fix $\mathbf{P}_u^{(\tau)}$, $\mathbf{P}_v^{(\tau)}$ (in $\mathbf{K}_{vu}$, $\mathbf{K}_{uv}$) and we find the subsequent projection matrices $\mathbf{P}_u^{(\tau+1)}$, $\mathbf{P}_v^{(\tau+1)}$ by solving Eq. (2); initially, $\mathbf{P}_u^{(0)}$, $\mathbf{P}_v^{(0)}$ are set using projection matrices of standard CCA. This process continues till a fixed point is reached. In practice, convergence to a fixed point is observed in less than five iterations (see Fig. 1).

As shown above, this new CCA formulation produces latent descriptors by combining both labeled and unlabeled data in a semi-supervised setting. In spite of this (and in contrast to related semi-supervised methods, mainly transductive ones)

[5]We assume that reference and test images are roughly (not precisely) registered.

[6]with $\psi_f(\mathbf{u}_i)$ being a feature vector associated to $\mathbf{u}_i$, see experiments.

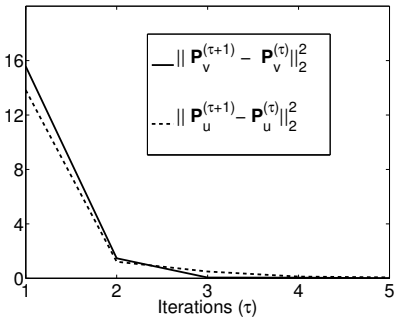

Fig. 1. This figure shows an example of the convergence of our iterative algorithm w.r.t τ. The same behavior occurred in all our experiments.

it is tractable and also inductive; it can be applied to new unseen data efficiently and without retraining the whole CCA transformations (in contrast to transductive methods). Indeed, the complexity of solving the above generalized eigenproblems depends mainly on the dimension d of data which is relatively small (i.e., $d \ll n$; see the setting of d in experiments).

IV. Experiments

A. Database and Evaluation Protocol

We evaluate the performance of our interactive (CCA based) change detection method on a database of $4,400$ patch pairs, taken from two registered (reference and test) GeoEye-1 satellite images of $2,400 \times 1,652$ pixels; the registration (between reference and test images) is given with some residual errors due to occlusions and other alignment errors. These two images correspond to the same area of Jefferson (Alabama) taken respectively in 2010 and in 2011 with many relevant changes due to tornados that happen in april 2011 (building destruction, etc.) and no-changes (including irrelevant ones such as clouds, etc.). Both reference and test images are processed in order to extract the database of $4,400$ non overlapping patches, each one includes 30×30 pixels in RGB. The underlying ground truth contains $4,275$ negative patch pairs ("no-changes" including irrelevant ones) and only 125 positive patch pairs (relevant changes), so less than 3% of these patches correspond to relevant changes.

Each patch (in reference and test images) is encoded with $d = 200$ coefficients corresponding to its projection on the 200 principal axes of principal component analysis (PCA). These principal axes of PCA were estimated using all patches of the reference image and capture more than 95% of the statistical variance of the data. Afterwards, the i^{th} patch in the test image is described (w.r.t the reference image) either without CCA as i) $\psi_f(\mathbf{v}_i) - \psi_f(\mathbf{u}_i)$ with $\psi_f(\mathbf{v}_i)$ being the projection of $\mathbf{v}_i$ using PCA or as ii) $\phi_v(\psi_f(\mathbf{v}_i)) - \phi_u(\psi_f(\mathbf{u}_i))$ when the CCA latent representations $\phi_u(.)$, $\phi_v(.)$ are considered. Performances are reported using equal error rate (EER) of learned classifiers $\{f_t\}_{t=1}^T$ on unlabeled data of $\mathcal{I}_1$. EER is the balanced generalization error that equally weights errors in "change" and "no-change" classes. Smaller EER implies better performance.

In all the subsequent experiments, only labeled patch pairs are correctly aligned by the oracle without errors while unlabeled pairs are subject to alignment errors.

B. Change detection performances: CCA baselines

Depending on the setting of the alignment matrix $\mathbf{D}$ and parameter β in Eq. (1), we define the following baselines for comparison. All these baselines use only the labeled patch pairs in order to set $\mathbf{D}$ (i.e., unlabeled patch pairs have their entries in $\mathbf{D}$ set to 0).

i) No-CCA (N-CCA): RF-based change detection is achieved on top of PCA features (see again Section IV-A), so the CCA transformations (ϕ_u and ϕ_v) correspond to identity functions.

ii) Standard CCA (S-CCA): this CCA version is obtained by using only negative labeled data (i.e., "no-changes"). Consequently, $\mathbf{D}$ is a diagonal matrix with $\mathbf{D}_{ii} = +1$ iff the aligned patch pair $(\mathbf{u}_i, \mathbf{v}_i)$ is labeled as "no-change" and $\mathbf{D}_{ii} = 0$ otherwise. In this setting, no context-based regularization is used (i.e., $\beta = 0$).

iii) Discriminant (D-CCA): this variant is obtained by using both positive and negative labeled data (i.e., "changes" and "no-changes") to define $\mathbf{D}$; the latter is a diagonal matrix with $\mathbf{D}_{ii} = \pm 1$ depending on whether the aligned patch pair $(\mathbf{u}_i, \mathbf{v}_i)$ is labeled as "no-change" or "change" by the oracle. Again, no context-based regularization is used (i.e., $\beta = 0$).

iv) Context-Aware Discriminant CCA (CAD-CCA): the setting of this CCA version is similar to D-CCA with the only difference being $\beta \neq 0$. Again, only labeled data are used to define $\mathbf{D}$ (i.e., the left-hand side term in Eq. 1) while both labeled and unlabeled data are used to define the right-hand side term (i.e., context-based regularization).

Fig. (2, bottom; top four curves) shows the RF performance without CCA and with (standard, discriminant and context-aware discriminant) CCA w.r.t the iteration number t. Note that all the results of this paper were obtained by averaging EERs of 10 RF runs, each one corresponds to a random setting of display $\mathcal{D}_0$ (see again Section II). These EERs decrease as t increases and reach their smallest values at the end of the iterative process, i.e., when decision criteria $\{f_t(.)\}_t$ are well trained/adapted to the reference and test images, and this happens after 10 iterations only. Fig. (2, top) shows the EERs of our RF algorithm (after 10 iterations) built upon context-aware discriminant CCA; these EERs globally decrease as β increases/reaches intermediate values and EERs increase again for larger values of β. From all these observations, it is clear that RF (when combined with context-aware CCA) has a clear gain compared to the other settings.

C. Impact of MR-CCA

In order to study the impact of MR-CCA on the performances of change detection – both with residual and relatively stronger misalignments – we consider the following settings for comparison

i) CAD-CCA (baseline 1): among the baselines discussed in Section IV-B, we keep only context-aware discriminant

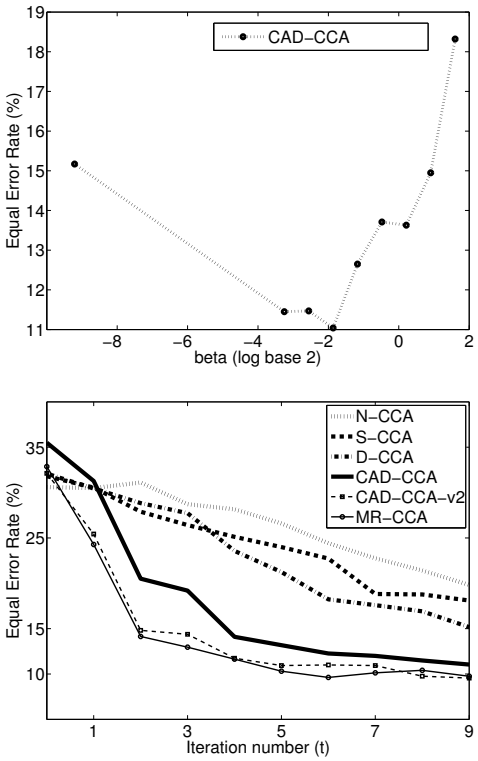

Fig. 2. (Top) This figure shows the evolution of EER w.r.t β; these values correspond to EER performances at the end of the RF process (i.e., when $t = 9$). Note that $\beta \rightarrow 0$ is equivalent to discriminant CCA. (Bottom) Evolution of change detection results w.r.t the iteration number (t), for different versions of CCA (for CAD-CCA, CAD-CCA-v2 and MR-CCA, β is set to 0.15). All these results are obtained by averaging EERs of 10 RF runs. For CAD-CCA-v2, $\sigma_f = 10$ while for MR-CCA $\sigma_f = 10$ and $\sigma_s = 1.0$. In these results, reference and test images are subject to residual alignment errors.

CCA (CAD-CCA) as it provides the best change detection performances; therefore, we use exactly the same setting for the matrix $\mathbf{D}$ and (the best value of) parameter β that controls context regularization (as shown in the caption of Fig. 2).

ii) CAD-CCA-v2 (baseline 2): the goal of this CAD-CCA variant is to study the impact of combining both labeled and unlabeled data (when defining $\mathbf{D}$) *without* being resilient to residual alignment errors in the unlabeled data. As reference and test images are globally aligned (with some residual errors), each unlabeled patch $\mathbf{v}_i \in \mathcal{I}_1$ is assumed to match $\mathbf{u}_i \in \mathcal{I}_0$ in the same 2D location, though this assumption does not hold all the time in practice. Hence, diagonal entries $\{\mathbf{D}_{ii}\}_i$ of labeled data (i.e., $i \in \{1, \ldots, \ell\}$) are set to ± 1 following the same protocol as CAD-CCA while entries $\{\mathbf{D}_{ii}\}_i$ of unlabeled data $(\mathbf{u}_i, \mathbf{v}_i)$ (i.e., $i \in \{\ell+1, \ldots, n\}$) are defined as $\{\mathbf{D}_{ii} = 2\exp(-\frac{\|\psi_f(\mathbf{v}_i)-\psi_f(\mathbf{u}_i)\|_2^2}{\sigma_f}) - 1\}_i$; again $\psi_f(.)$ stands for the PCA features and σ_f is a scale factor.

iii) MR-CCA: the goal of this MR-CCA variant is to study the impact of combining both labeled and unlabeled data (when defining $\mathbf{D}$) *while now being* resilient to residual alignment errors in the unlabeled data. We assume that each unlabeled patch $\mathbf{v}_i \in \mathcal{I}_1$ matches one of the neighboring patches $\{\mathbf{u}_j\}_{j \in \cup_c \mathcal{N}_c(i)} \subset \mathcal{I}_0$; *but this match is unknown*. Hence, diagonal entries $\{\mathbf{D}_{ii}\}_i$ of labeled data (i.e., $i \in \{1, \ldots, \ell\}$) are again set to ± 1 following the same protocol as CAD-CCA while non-diagonal entries $\{\mathbf{D}_{ij}\}_{ij}$ of unlabeled data $\{(\mathbf{u}_i, \mathbf{v}_j)\}_{ij}$ (with $i,\ j \in \{\ell+1, \ldots, n\}$) are defined as

$$\Big\{\mathbf{D}_{ij} = 2\exp(-\frac{\|\psi_f(\mathbf{v}_j) - \psi_f(\mathbf{u}_i)\|_2^2}{\sigma_f}).\exp(-\frac{\|\psi_s(\mathbf{v}_j) - \psi_s(\mathbf{u}_i)\|_2^2}{\sigma_s}) - 1\Big\}_{ij}, \quad (4)$$

here $\psi_s(\mathbf{v}_j)$ stands for spatial (2D) coordinates of $\mathbf{v}_j$ and σ_s corresponds to a scale factor. With Eq. (4), we model the *uncertainty* of matches in the unlabeled data.

Fig. (2, bottom) shows a comparison of MR-CCA against all other baselines including S-CCA, D-CCA and CAD-CCA (with its variant CAD-CCA-v2). According to these results, the three CCA versions (CAD-CCA, CAD-CCA-v2 and MR-CCA) outperform N-CCA, S-CCA and D-CCA versions with a slight advantage of MR-CCA over CAD-CCA and CAD-CCA-v2. This clearly shows that when reference and test images are globally well aligned (with some residual errors), the gain in performances is dominated by the positive impact of context regularization; indeed, the impact of misalignment resilience – in spite of being positive – is relatively marginal when patches are globally well aligned.

In order to study the impact of MR-CCA w.r.t stronger alignment errors (i.e. w.r.t a more challenging setting), we apply a relatively strong motion field to all the pixels in the reference image $\mathcal{I}_0$; precisely, each pixel is shifted along a direction whose x-y coordinates are randomly set to values between 5 and 15 pixels. These shifts are sufficient in order to make the quality of alignments[7] used for CCA very weak so the different versions of CCA, mentioned earlier, become more sensitive to alignment errors (EERs increase by more than 60% in Fig. 3 compared to EERs without strong alignment errors in Fig. 2, bottom). With this setting, MR-CCA is clearly more resilient and shows a substantial relative gain compared to both CAD-CCA and CAD-CCA-v2 (see again Fig. 3).

V. Conclusion

We introduced in this paper a novel change detection algorithm based on relevance feedback and a new variant of CCA referred to as misalignment resilient CCA (MR-CCA). The latter learns transformation matrices that map data from their original input spaces into a latent space where aligned data become strongly or weakly correlated depending on their labels. This is achieved by optimizing an objective function mixing two terms: the first one relies on a discriminative setting that maximizes (resp. minimizes) correlations between "no changes" (resp. "changes"). The

[7]Initially, patch pairs in reference and test images $\mathcal{I}_0$, $\mathcal{I}_1$ were globally well registered (with only small residual errors). When applying the motion field to $\mathcal{I}_0$, these patch pairs become strongly misaligned; however, as the oracle selects labeled data while also providing precisely their alignments, only unlabeled data remain strongly misaligned.

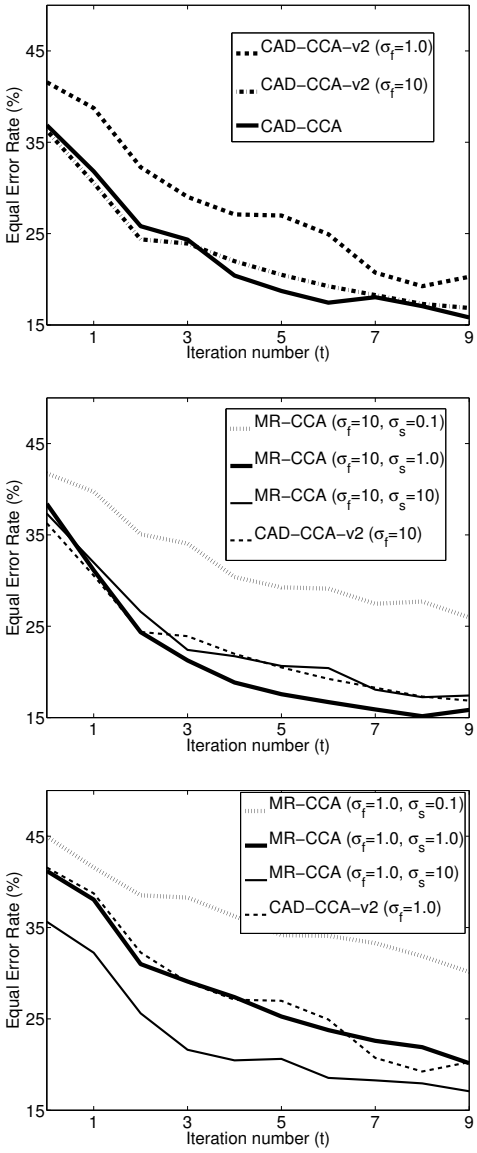

Fig. 3. This figure shows a comparison of our RF method using CAD-CCA, CAD-CCA-v2 and MR-CCA for different settings of σ_f and σ_s; again β is set to 0.15. As already discussed, and in contrast to Fig. 2, reference and test patch pairs are subject to stronger alignment errors.

second term acts as a regularizer that makes correlations spatially smooth and provides us with robust context-aware latent representations. Our method considers both labeled and unlabeled data when learning the CCA transformations while being resilient to alignment errors in the unlabeled data. As shown through extensive experiments, the relative gain of our MR-CCA method is substantial especially when reference and test satellite images are subject to alignment errors.

Even though applied to the particular problem of satellite image change detection, our MR-CCA can be extended to many other pattern recognition tasks where alignments are error-prone – and when context can be exploited in order to recover from these alignment errors. These tasks include "image-to-text" mapping in video annotation, "text-to-text" alignment in multilingual machine translation, as well as "image-to-image" matching in multi-view object tracking.

ACKNOWLEDGMENT

This work was supported in part with a grant from the research agency ANR (Agence Nationale de la Recherche) under the MLVIS project, ANR-11-BS02-0017.

REFERENCES

[1] R.J. 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.

[2] G. Moser and S. Serpico, "Generalized minimum error thresholding for unsupervised change detection from amplitude sar imagery," *IEEE TGRS*, vol. 44, no. 10, pp. 2972–2983, 2006.

[3] 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.

[4] 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.

[5] C. Benedek and T. Sziranyi, "Bayesian foreground and shadow detection in uncertain frame rate surveillance videos," *IEEE Trans on Image Processing*, vol. 17, no. 4, pp. 608–621, 2008.

[6] N. Bourdis, D. Marraud, and H. Sahbi, "Constrained optical flow for aerial image change detection," in *International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE, 2011, pp. 4176–4179.

[7] L. Bruzzone and D. Fernandez-Prieto, "Automatic analysis of the difference image for unsupervised change detection," *IEEE TGRS*, vol. 38, no. 3, pp. 1171–1182, 2000.

[8] Pollard, "Comprehensive 3d change detection using volumetric appearance modeling," *Phd, Brown University*, 2009.

[9] J. Im, J. Jensen, and J. Tullis, "Object-based change detection using correlation image analysis and image segmentation," *International Journal of Remote Sensing*, vol. 29, no. 2, pp. 399–423, 2008.

[10] Martin A Fischler and Robert C Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[11] T. Kim and Y-Jo. Im, "Automatic satellite image registration by combination of matching and random sample consensus," *IEEE TGRS*, vol. 41, no. 5, pp. 1111–1117, 2003.

[12] H. Sahbi, "Discriminant canonial correlation analysis for interactive satellite image change detection," *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2015.

[13] D. Hardoon, S. Szedmak, and J. Shawe-Taylor, "Canonical correlation analysis: An overview with application to learning methods," *Neural computation*, vol. 16, no. 12, pp. 2639–2664, 2004.

[14] M. Ferecatu and H. Sahbi, "Multi-view object matching and tracking using canonical correlation analysis," in *Image Processing (ICIP), 2009 16th IEEE International Conference on*. IEEE, 2009, pp. 2109–2112.

[15] X.S. Zhou and T.S. Huang, "Relevance feedback in image retrieval: A comprehensive review," *CVPR - CBAIVL Workshop*, 2006.

[16] A. Morde, X. Ma, and S. Guler, "Learning a background model for change detection," *CVPR Workshop*, 2012.

[17] F. Bach, "Active learning for misspecified generalized linear models," *NIPS*, vol. 19, 2006.

[18] H. Sahbi, "Relevance feedback for satellite image change detection," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1503–1507, 2013.

[19] C-C Chang and C-J Lin, "Libsvm: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, pp. 27, 2011.

[20] F. Fleuret and H. Sahbi, "Scale invariance of support vector machines based on the triangular kernel," *The third international workshop on statistical and computational theories of vision (part of ICCV)*, 2003.

[21] M. Ferecatu, "Image retrieval with active relevance feedback using both visual and keyword-based descriptors," *PhD thesis, Versailles University*, 2005.