



Learning CCA Representations for Misaligned Data

Hichem Sahbi^(✉)

CNRS, Sorbonne University, Paris, France
hichem.sahbi@sorbonne-universite.fr

Abstract. Canonical correlation analysis (CCA) is a statistical learning method that seeks to build view-independent latent representations from multi-view data. This method has been successfully applied to several pattern analysis tasks such as image-to-text mapping and view-invariant object/action recognition. However, this success is highly dependent on the quality of data pairing (i.e., alignments) and mispairing adversely affects the generalization ability of the learned CCA representations.

In this paper, we address the issue of alignment errors using a new variant of canonical correlation analysis referred to as alignment-agnostic (AA) CCA. Starting from erroneously paired data taken from different views, this CCA finds transformation matrices by optimizing a constrained maximization problem that mixes a data correlation term with context regularization; the particular design of these two terms mitigates the effect of alignment errors when learning the CCA transformations. Experiments conducted on multi-view tasks, including multi-temporal satellite image change detection, show that our AA CCA method is highly effective and resilient to mispairing errors.

Keywords: Canonical correlation analysis
Learning compact representations · Misalignment resilience
Change detection

1 Introduction

Several tasks in computer vision and neighboring fields require *labeled* datasets in order to build effective statistical learning models. It is widely agreed that the accuracy of these models relies substantially on the availability of large labeled training sets. These sets require a tremendous human annotation effort and are thereby very expensive for many large scale classification problems including image/video-to-text (a.k.a captioning) [1–4], multi-modal information retrieval [5], multi-temporal change detection [6, 7], object recognition and segmentation [8, 9], etc. The current trend in machine learning, mainly with the data-hungry deep models [1, 2, 10–12], is to bypass supervision, by making the training of these models totally unsupervised [13], or at least weakly-supervised using: fine-tuning [14], self-supervision [15], data augmentation and game-based models

[16]. However, the hardness of collecting annotated datasets does not *only* stem from assigning accurate labels to these data, but also from aligning them; for instance, in the neighboring field of machine translation, successful training models require accurately aligned bi-texts (parallel bilingual training sets), while in satellite image change detection, these models require accurately georeferenced and registered satellite images. This level of requirement, *both* on the accuracy of labels and *their alignments*, is clearly hard-to-reach; alternative models, that skip the sticky alignment requirement, should be preferred.

Canonical correlation analysis (CCA) [17–20] is one of the statistical learning models that require accurately aligned (paired) multi-view data¹; CCA finds – for each view – a transformation matrix that maps data from that view to a view-independent (latent) representation such that aligned data obtain highly correlated latent representations. Several extensions of CCA have been introduced in the literature including nonlinear (kernel) CCA [21], sparse CCA [22–24], multiple CCA [25], locality preserving and instance-specific CCA [26, 27], time-dependent CCA [28] and other unified variants (see for instance [29, 30]); these methods have been applied to several pattern analysis tasks such as image-to-text [31], pose estimation [21, 26] and object recognition [32], multi-camera activity correlation [33, 34] and motion alignment [35, 36] as well as heterogeneous sensor data classification [37].

The success of all the aforementioned CCA approaches is highly *dependent* on the accuracy of alignments between multi-view data. In practice, data are subject to misalignments (such as registration errors in satellite imagery) and sometimes completely unaligned (as in multi-lingual documents) and this skews the learning of CCA. Excepting a few attempts – to handle temporal deformations in monotonic sequence datasets [38] using canonical time warping [36] (and its deep extension [39]) – none of these existing CCA variants address alignment errors for non-monotonic datasets². Besides CCA, the issue of data alignment has been approached, in general, using manifold alignment [40–42], Procrustes analysis [43] and source-target domain adaption [44] but none of these methods consider resilience to misalignments as a part of CCA design (which is the main purpose of our contribution in this paper). Furthermore, these data alignment solutions rely on a strong “apples-to-apples” comparison hypothesis (*that data taken from different views have similar structures*) which does not always hold especially when handling datasets with heterogeneous views (as text/image data and multi-temporal or multi-sensor satellite images). Moreover, even when data are globally well (re)aligned, some *residual* alignment errors are difficult to handle (such as parallax in multi-temporal satellite imagery) and harm CCA (as shown in our experiments).

In this paper, we introduce a novel CCA approach that handles misaligned data; i.e., it does not require any preliminary step of accurate data alignment.

¹ Multi-view data stands for input data described with multiple modalities such as documents described with text and images.

² Non-monotonic stands for datasets without a “unique” order (such as patches in images).

This is again very useful for different applications where aligning data is very time demanding or when data are taken from multiple sources (sensors, modalities, etc.) which are intrinsically misaligned³. The benefit of our approach is twofold; on the one hand, it models the uncertainty of alignments using a new data correlation term and on the other hand, modeling alignment uncertainty allows us to use not only decently aligned data (if available) when learning CCA, but also the unaligned ones. In sum, this approach can be seen as an extension of CCA on unaligned sets compared to standard CCA (and its variants) that operate only on accurately aligned data. Furthermore, the proposed method is as efficient as standard CCA and its computational complexity grows w.r.t the dimensionality (and not the cardinality) of data, and this makes it very suitable for large datasets.

Our CCA formulation is based on the optimization of a constrained objective function that combines two terms; a correlation criterion and a context-based regularizer. The former maximizes a weighted correlation between data with a high cross-view similarity while the latter makes this weighted correlation high for data whose neighbors have high correlations too (and vice-versa). We will show that optimizing this constrained maximization problem is equivalent to solving an iterative generalized eigenvalue/eigenvector decomposition; we will also show that the solution of this iterative process converges to a fixed-point. Finally, we will illustrate the validity of our CCA formulation on different challenging problems including change detection both on *residually and strongly* misaligned multi-temporal satellite images; indeed, these images are subject to alignment errors due to the hardness of image registration under challenging conditions, such as occlusion and parallax.

The rest of this paper is organized as follows; Sect. 2 briefly reminds the preliminaries in canonical correlation analysis, followed by our main contribution: a novel alignment-agnostic CCA, as well as some theoretical results about the convergence of the learned CCA transformation to a fixed-point (under some constraints on the parameter that weights our regularization term). Section 3 shows the validity of our method both on synthetic toy data as well as real-world problems namely satellite image change detection. Finally, we conclude the paper in Sect. 4 while providing possible extensions for a future work.

2 Canonical Correlation Analysis

Considering the input spaces \mathcal{X}_r and \mathcal{X}_t as two sets of images taken from two modalities; in satellite imagery, these modalities could be two different sensors, or the same sensor at two different instants, etc. Denote $\mathcal{I}_r = \{\mathbf{u}_i\}_i$, $\mathcal{I}_t = \{\mathbf{v}_j\}_j$ as two subsets of \mathcal{X}_r and \mathcal{X}_t respectively; our goal is learn a transformation between \mathcal{X}_r and \mathcal{X}_t that assigns, for a given $\mathbf{u} \in \mathcal{X}_r$, a sample $\mathbf{v} \in \mathcal{X}_t$. The learning of this transformation usually requires accurately *paired* data in $\mathcal{X}_r \times \mathcal{X}_t$ as in CCA.

³ Satellite images – georeferenced with the Global Positioning System (GPS) – have localization errors that may reach 15 m in some geographic areas. On high resolution satellite images (sub-metric resolution) this corresponds to alignment errors/drifts that may reach 30 pixels.

2.1 Standard CCA

Assuming centered data in $\mathcal{I}_r, \mathcal{I}_t$, standard CCA (see for instance [19]) finds two projection matrices that map aligned data in $\mathcal{I}_r \times \mathcal{I}_t$ into a latent space while maximizing their correlation. Let $\mathbf{P}_r, \mathbf{P}_t$ denote these projection matrices which respectively correspond to reference and test images. CCA finds these matrices as $(\mathbf{P}_r, \mathbf{P}_t) = \arg \max_{\mathbf{A}, \mathbf{B}} \text{tr}(\mathbf{A}' \mathbf{C}_{rt} \mathbf{B})$, subject to $\mathbf{A}' \mathbf{C}_{rr} \mathbf{A} = \mathbf{I}_u, \mathbf{B}' \mathbf{C}_{tt} \mathbf{B} = \mathbf{I}_v$; here \mathbf{I}_u (resp. \mathbf{I}_v) is the $d_u \times d_u$ (resp. $d_v \times d_v$) identity matrix, d_u (resp. d_v) is the dimensionality of data in \mathcal{X}_r (resp. \mathcal{X}_t), \mathbf{A}' stands for transpose of \mathbf{A} , tr is the trace, \mathbf{C}_{rt} (resp. $\mathbf{C}_{rr}, \mathbf{C}_{tt}$) correspond to inter-class (resp. intra-class) covariance matrices of data in $\mathcal{I}_r, \mathcal{I}_t$, and equality constraints control the effect of scaling on the solution. One can show that problem above is equivalent to solving the eigenproblem $\mathbf{C}_{rt} \mathbf{C}_{tt}^{-1} \mathbf{C}_{tr} \mathbf{P}_r = \gamma^2 \mathbf{C}_{rr} \mathbf{P}_r$ with $\mathbf{P}_t = \frac{1}{\gamma} \mathbf{C}_{tt}^{-1} \mathbf{C}_{tr} \mathbf{P}_r$. In practice, learning these two transformations requires “paired” data in $\mathcal{I}_r \times \mathcal{I}_t$, i.e., aligned data. However, and as will be shown through this paper, accurately paired data are not always available (and also expensive), furthermore the cardinality of \mathcal{I}_r and \mathcal{I}_t can also be different, so one should adapt CCA in order to learn transformation between data in \mathcal{I}_r and \mathcal{I}_t as shown subsequently.

2.2 Alignment Agnostic CCA

We introduce our main contribution: a novel alignment agnostic CCA approach. Considering $\{(\mathbf{u}_i, \mathbf{v}_j)\}_{ij}$ as a subset of $\mathcal{I}_r \times \mathcal{I}_t$ (cardinalities of $\mathcal{I}_r, \mathcal{I}_t$ are not necessarily equal), we propose to find the transformation matrices $\mathbf{P}_r, \mathbf{P}_t$ as

$$\begin{aligned} \max_{\mathbf{P}_r, \mathbf{P}_t} \text{tr}(\mathbf{U}' \mathbf{P}_r \mathbf{P}_t' \mathbf{V} \mathbf{D}) \\ \text{s.t. } \mathbf{P}_r' \mathbf{C}_{rr} \mathbf{P}_r = \mathbf{I}_u \text{ and } \mathbf{P}_t' \mathbf{C}_{tt} \mathbf{P}_t = \mathbf{I}_v, \end{aligned} \quad (1)$$

the non-matrix form of this objective function is given subsequently. In this constrained maximization problem, \mathbf{U}, \mathbf{V} are two matrices of data in $\mathcal{I}_r, \mathcal{I}_t$ respectively, and \mathbf{D} is an (application-dependent) matrix with its given entry \mathbf{D}_{ij} set to the cross affinity or the likelihood that a given data $\mathbf{u}_i \in \mathcal{I}_r$ aligns with $\mathbf{v}_j \in \mathcal{I}_t$ (see Sect. 3.2 about different setting of this matrix). This definition of \mathbf{D} , together with objective function (1), make CCA *alignment agnostic*; indeed, this objective function (equivalent to $\sum_{i,j} \langle \mathbf{P}_r' \mathbf{u}_i, \mathbf{P}_t' \mathbf{v}_j \rangle \mathbf{D}_{ij}$) aims to maximize the correlation between pairs (with a high cross affinity of alignment) while it also minimizes the correlation between pairs with small cross affinity. For a particular setting of \mathbf{D} , the following proposition provides a *special case*.

Proposition 1. Provided that $|\mathcal{I}_r| = |\mathcal{I}_t|$ and $\forall \mathbf{u}_i \in \mathcal{I}_r, \exists! \mathbf{v}_j \in \mathcal{I}_t$ such that $\mathbf{D}_{ij} = 1$; the constrained maximization problem (1) implements standard CCA.

Proof. Considering the non-matrix form of (1), we obtain

$$\text{tr}(\mathbf{U}' \mathbf{P}_r \mathbf{P}_t' \mathbf{V} \mathbf{D}) = \sum_{i,j} \langle \mathbf{P}_r' \mathbf{u}_i, \mathbf{P}_t' \mathbf{v}_j \rangle \mathbf{D}_{ij}, \quad (2)$$

considering a particular order of \mathcal{I}_t such that each sample \mathbf{u}_i in \mathcal{I}_r aligns with a unique \mathbf{v}_i in \mathcal{I}_t we obtain

$$\begin{aligned} \text{tr}(\mathbf{U}'\mathbf{P}_r\mathbf{P}'_t\mathbf{V}\mathbf{D}) &= \sum_{i,j} \langle \mathbf{P}'_r\mathbf{u}_i, \mathbf{P}'_t\mathbf{v}_j \rangle \mathbf{1}_{\{i=j\}} \\ &= \sum_i \langle \mathbf{P}'_r\mathbf{u}_i, \mathbf{P}'_t\mathbf{v}_i \rangle \\ &= \text{tr}(\mathbf{P}'_r(\sum_i \mathbf{u}_i\mathbf{v}'_i)\mathbf{P}_t) \\ &= \text{tr}(\mathbf{P}'_r\mathbf{C}_{rt}\mathbf{P}_t), \end{aligned} \tag{3}$$

with \mathbf{C}_{rt} being the inter-class covariance matrix and $\mathbf{1}_{\{\cdot\}}$ the indicator function. Since the equality constraints (shown in Sect. 2.1) remain unchanged, the constrained maximization problem (1) is strictly equivalent to standard CCA for this particular \mathbf{D} \square

This particular setting of \mathbf{D} is relevant only when data are accurately paired and also when $\mathcal{I}_r, \mathcal{I}_t$ have the same cardinality. In practice, many problems involve unpaired/mispaired datasets with different cardinalities; that’s why \mathbf{D} should be relaxed using affinity between multiple pairs (as discussed earlier in this section) instead of using strict alignments. With this new CCA setting, the learned transformations \mathbf{P}_t , and \mathbf{P}_r generate latent data representations $\phi_t(\mathbf{v}_i) = \mathbf{P}'_t\mathbf{v}_i$, $\phi_r(\mathbf{u}_j) = \mathbf{P}'_r\mathbf{u}_j$ which align according to \mathbf{D} (i.e., $\|\phi_r(\mathbf{v}_i) - \phi_t(\mathbf{u}_j)\|_2$ decreases if \mathbf{D}_{ij} is high and vice-versa). However, when multiple entries $\{\mathbf{D}_{ij}\}_j$ are high for a given i , this may produce noisy correlations between the learned latent representations and may impact their discrimination power (see also experiments). In order to mitigate this effect, we also consider context regularization.

2.3 Context-Based Regularization

For each data $\mathbf{u}_i \in \mathcal{I}_r$, we define a (typed) neighborhood system $\{\mathcal{N}_c(i)\}_{c=1}^C$ which corresponds to the typed neighbors of \mathbf{u}_i (see Sect. 3.2 for an example). Using $\{\mathcal{N}_c(\cdot)\}_{c=1}^C$, we consider for each c an intrinsic adjacency matrix \mathbf{W}_u^c whose $(i, k)^{\text{th}}$ entry is set as $\mathbf{W}_{u,i,k}^c \propto \mathbf{1}_{\{k \in \mathcal{N}_c(i)\}}$. Similarly, we define the matrices $\{\mathbf{W}_v^c\}_c$ for data $\{\mathbf{v}_j\}_j \in \mathcal{I}_t$; extra details about the setting of these matrices are again given in experiments.

Using the above definition of $\{\mathbf{W}_u^c\}_c, \{\mathbf{W}_v^c\}_c$, we add an extra-term in the objective function (1) as

$$\begin{aligned} \max_{\mathbf{P}_r, \mathbf{P}_t} \text{tr}(\mathbf{U}'\mathbf{P}_r\mathbf{P}'_t\mathbf{V}\mathbf{D}) + \beta \sum_{c=1}^C \text{tr}(\mathbf{U}'\mathbf{P}_r\mathbf{P}'_t\mathbf{V}\mathbf{W}_v^c\mathbf{V}'\mathbf{P}_t\mathbf{P}'_r\mathbf{U}\mathbf{W}_u^c) \\ \text{s.t.} \quad \mathbf{P}'_r\mathbf{C}_{rr}\mathbf{P}_r = \mathbf{I}_u \quad \text{and} \quad \mathbf{P}'_t\mathbf{C}_{tt}\mathbf{P}_t = \mathbf{I}_v. \end{aligned} \tag{4}$$

The above right-hand side term is equivalent to

$$\beta \sum_c \sum_{i,j} \langle \mathbf{P}'_r\mathbf{u}_i, \mathbf{P}'_t\mathbf{v}_j \rangle \sum_{k,\ell} \langle \mathbf{P}'_r\mathbf{u}_k, \mathbf{P}'_t\mathbf{v}_\ell \rangle \mathbf{W}_{u,i,k}^c \mathbf{W}_{v,j,\ell}^c$$

the latter corresponds to a neighborhood (or context) criterion which considers that a high value of the correlation $\langle \mathbf{P}'_r \mathbf{u}_i, \mathbf{P}'_t \mathbf{v}_j \rangle$, in the learned latent space, should imply high correlation values in the neighborhoods $\{\mathcal{N}_c(i) \times \mathcal{N}_c(j)\}_c$. This term (via β) controls the sharpness of the correlations (and also the discrimination power) of the learned latent representations (see example in Fig. 2). Put differently, if a given $(\mathbf{u}_i, \mathbf{v}_j)$ is surrounded by highly correlated pairs, then the correlation between $(\mathbf{u}_i, \mathbf{v}_j)$ should be maximized and vice-versa [45, 46].

2.4 Optimization

Considering Lagrange multipliers for the equality constraints in Eq. (4), one may show that optimality conditions (related to the gradient of Eq. (4) w.r.t $\mathbf{P}_r, \mathbf{P}_t$ and the Lagrange multipliers) lead to the following generalized eigenproblem

$$\begin{aligned} \mathbf{K}_{rt} \mathbf{C}_{tt}^{-1} \mathbf{K}_{tr} \mathbf{P}_r &= \gamma^2 \mathbf{C}_{rr} \mathbf{P}_r \\ \text{with } \mathbf{P}_t &= \frac{1}{\gamma} \mathbf{C}_{tt}^{-1} \mathbf{K}_{tr} \mathbf{P}_r, \end{aligned} \quad (5)$$

here $\mathbf{K}_{tr} = \mathbf{K}'_{rt}$ and

$$\begin{aligned} \mathbf{K}_{tr} &= \mathbf{V} \mathbf{D} \mathbf{U}' + \beta \sum_c \mathbf{V} \mathbf{W}_v^c \mathbf{V}' \mathbf{P}_t \mathbf{P}'_r \mathbf{U} \mathbf{W}_u^c \mathbf{U}' \\ &\quad + \beta \sum_c \mathbf{V} \mathbf{W}_v^{c'} \mathbf{V}' \mathbf{P}_t \mathbf{P}'_r \mathbf{U} \mathbf{W}_u^c \mathbf{U}'. \end{aligned} \quad (6)$$

In practice, we solve the above eigenproblem iteratively. For each iteration τ , we fix $\mathbf{P}_r^{(\tau)}, \mathbf{P}_t^{(\tau)}$ (in $\mathbf{K}_{tr}, \mathbf{K}_{rt}$) and we find the subsequent projection matrices $\mathbf{P}_r^{(\tau+1)}, \mathbf{P}_t^{(\tau+1)}$ by solving Eq. (5); initially, $\mathbf{P}_r^{(0)}, \mathbf{P}_t^{(0)}$ 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.

Proposition 2. Let $\|\cdot\|_1$ denote the entry-wise L_1 -norm and $\mathbf{1}_{vu}$ a $d_v \times d_u$ matrix of ones. Provided that the following inequality holds

$$\beta < \gamma_{\min} \times \left(\sum_c \|\mathbf{E}_c \mathbf{1}_{vu} \mathbf{F}'_c\|_1 + \sum_c \|\mathbf{G}_c \mathbf{1}_{vu} \mathbf{H}'_c\|_1 \right)^{-1} \quad (7)$$

with γ_{\min} being a lower bound of the positive eigenvalues of (5), $\mathbf{E}_c = \mathbf{V} \mathbf{W}_v^c \mathbf{V}' \mathbf{C}_{tt}^{-1}$, $\mathbf{F}_c = \mathbf{U} \mathbf{W}_u^c \mathbf{U}' \mathbf{C}_{rr}^{-1}$, $\mathbf{G}_c = \mathbf{V} \mathbf{W}_v^{c'} \mathbf{V}' \mathbf{C}_{tt}^{-1}$ and $\mathbf{H}_c = \mathbf{U} \mathbf{W}_u^{c'} \mathbf{U}' \mathbf{C}_{rr}^{-1}$, the problem in (5), (6) admits a unique solution $\tilde{\mathbf{P}}_r, \tilde{\mathbf{P}}_t$ as the eigenvectors of

$$\begin{aligned} \tilde{\mathbf{K}}_{rt} \mathbf{C}_{tt}^{-1} \tilde{\mathbf{K}}_{tr} \mathbf{P}_r &= \gamma^2 \mathbf{C}_{rr} \mathbf{P}_r \\ \mathbf{P}_t &= \frac{1}{\gamma} \mathbf{C}_{tt}^{-1} \tilde{\mathbf{K}}_{tr} \mathbf{P}_r, \end{aligned} \quad (8)$$

with $\tilde{\mathbf{K}}_{tr}$ being the limit of

$$\mathbf{K}_{tr}^{(\tau+1)} = \Psi(\mathbf{K}_{tr}^{(\tau)}), \quad (9)$$

and $\Psi : \mathbb{R}^{d_v \times d_u} \rightarrow \mathbb{R}^{d_v \times d_u}$ is given as

$$\begin{aligned} \Psi(\mathbf{K}_{tr}) = & \mathbf{V}\mathbf{D}\mathbf{U}' + \beta \sum_c \mathbf{V}\mathbf{W}_v^c \mathbf{V}' \mathbf{P}_t \mathbf{P}'_r \mathbf{U} \mathbf{W}_u^{c'} \mathbf{U}' \\ & + \beta \sum_c \mathbf{V}\mathbf{W}_v^{c'} \mathbf{V}' \mathbf{P}_t \mathbf{P}'_r \mathbf{U} \mathbf{W}_u^c \mathbf{U}', \end{aligned} \quad (10)$$

with $\mathbf{P}_t, \mathbf{P}_r$, in (10), being functions of \mathbf{K}_{tr} using (5). Furthermore, the matrices $\mathbf{K}_{tr}^{(\tau+1)}$ in (9) satisfy the convergence property

$$\|\mathbf{K}_{tr}^{(\tau+1)} - \tilde{\mathbf{K}}_{tr}\|_1 \leq L^\tau \|\mathbf{K}_{tr}^{(\tau+1)} - \mathbf{K}_{tr}^{(0)}\|_1, \quad (11)$$

with $L = \frac{\beta}{\gamma_{\min}} (\sum_c \|\mathbf{E}_c \mathbf{1}_{vu} \mathbf{F}'_c\|_1 + \sum_c \|\mathbf{G}_c \mathbf{1}_{vu} \mathbf{H}'_c\|_1)$.

Proof. See appendix

Note that resulting from the extreme sparsity of the *typed* adjacency matrices $\{\mathbf{W}_u^c\}_c, \{\mathbf{W}_v^c\}_c$, the upper bound about β (shown in the sufficient condition in Eq. 7) is loose, and easy to satisfy; in practice, we observed convergence for all the values of β that were tried in our experiments (see the x-axis of Fig. 2).

3 Experiments

In this section, we show the performance of our method both on synthetic and real datasets. The goal is to show the extra gain brought when using our alignment agnostic (AA) CCA approach against standard CCA and other variants.

3.1 Synthetic Toy Example

In order to show the strength of our AA CCA method, we first illustrate its performance on a 2D toy example. We consider 2D data sampled from an ‘‘arc’’ as shown in Fig. 1(a); each sample is endowed with an RGB color feature vector which depends on its curvilinear coordinates in that ‘‘arc’’. We duplicate this dataset using a 2D rotation (with an angle of 180°) and we add a random perturbation field (noise) both to the color features and the 2D coordinates (see Fig. 1). Note that accurate ground-truth pairing is available but, of course, not used in our experiments.

We apply our AA CCA (as well as standard CCA) to these data, and we show alignment results; this 2D toy example is very similar to the subsequent real data task as the goal is to find for each sample in the original set, its correlations and its realignment with the second set. From Fig. 1, it is clear that standard CCA fails to produce accurate results when data is contaminated with random perturbations and alignment errors, while our AA CCA approach successfully realigns the two sets (see again details in Fig. 1).

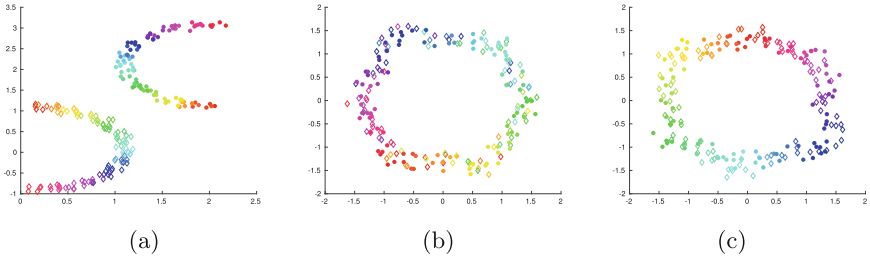


Fig. 1. This figure shows the realignment results of CCA; (a) we consider 100 examples sampled from an “arc”, each sample is endowed with an RGB feature vector. We duplicate this dataset using a 2D rotation (with an angle of 180°) and we add a random perturbation field both to the color features and 2D coordinates. (b) realignment results obtained using standard CCA; note that original data are not aligned, so in order to apply standard CCA, each sample in the first arc-set is paired with its nearest (color descriptor) neighbor in the second arc-set. (c) realignment results obtained using our AA CCA approach; again data are not paired, so we consider a fully dense matrix \mathbf{D} that measures the cross-similarity (using an RBF kernel) between the colors of the first and the second arc-sets. In these toy experiments, β (weight of context regularizer) is set to 0.01 and we use an isotropic neighborhood system in order to fill the context matrices $\{\mathbf{W}_u^c\}_{c=1}^C$, $\{\mathbf{W}_v^c\}_{c=1}^C$ (with $C = 1$) and a given entry $\mathbf{W}_{u,i,k}^c$ is set to 1 iff \mathbf{u}_k is among the 10 spatial neighbors of \mathbf{u}_i . Similarly, we set the entries of $\{\mathbf{W}_v^c\}_{c=1}^C$. **For a better visualization of these results, better to view/zoom the PDF of this paper.**

3.2 Satellite Image Change Detection

We also evaluate and compare the performance of our proposed AA CCA method on the challenging task of satellite image change detection (see for instance [6, 47–50]). The goal is to find instances of relevant changes into a given scene acquired at instance t_1 with respect to the same scene taken at instant $t_0 < t_1$; these acquisitions (at instants t_0, t_1) are referred to as reference and test images respectively. This task is known to be very challenging due to the difficulty to characterize relevant changes (appearance or disappearance of objects⁴) from irrelevant ones such as the presence of cars, clouds, as well as *registration errors*. This task is also practically important; indeed, in the particular important scenario of damage assessment after natural hazards (such as tornadoes, earth quakes, etc.), it is crucial to achieve automatic change detection accurately in order to organize and prioritize rescue operations.

JOPLIN-TORNADOES11 Dataset. This dataset includes 680928 non overlapping image patches (of 30×30 pixels in RGB) taken from six pairs of (reference and test) GeoEye-1 satellite images (of 9850×10400 pixels each).

⁴ This can be any object so there is no a priori knowledge about what object may appear or disappear into a given scene.

This dataset is randomly split into two subsets: *labeled* used for training⁵ (denoted $\mathcal{L}_r \subset \mathcal{I}_r$, $\mathcal{L}_t \subset \mathcal{I}_t$) and *unlabeled* used for testing (denoted $\mathcal{U}_r = \mathcal{I}_r \setminus \mathcal{L}_r$ and $\mathcal{U}_t = \mathcal{I}_t \setminus \mathcal{L}_t$) with $|\mathcal{L}_r| = |\mathcal{L}_t| = 3000$ and $|\mathcal{U}_r| = |\mathcal{U}_t| = 680928 - 3000$. All patches in \mathcal{I}_r (or in \mathcal{I}_t), stitched together, cover a very large area – of about $20 \times 20 \text{ km}^2$ – around Joplin (Missouri) and show many changes after tornadoes that happened in may 2011 (building destruction, etc.) and no-changes (including irrelevant ones such as car appearance/disappearance, etc.). Each patch in \mathcal{I}_r , \mathcal{I}_t is rescaled and encoded with 4096 coefficients corresponding to the output of an inner layer of the pretrained VGG-net [51]. A given test patch is declared as a “change” or “no-change” depending on the score of SVMs trained on top of the learned CCA latent representations.

In order to evaluate the performances of change detection, we report the equal error rate (EER). The latter is a balanced generalization error that equally weights errors in “change” and “no-change” classes. Smaller EER implies better performance.

Data Pairing and Context Regularization. In order to study the impact of AA CCA on the performances of change detection – both with residual and relatively stronger misalignments – we consider the following settings for comparison (see also Table 1).

- **Standard CCA:** patches are *strictly* paired by assigning each patch, in the reference image, to a unique patch in the test image (in the same location), so it assumes that satellite images are correctly registered. CCA learning is *supervised* (only labeled patches are used for training) and *no-context* regularization is used (i.e, $\beta = 0$). In order to implement this setting, we consider \mathbf{D} as a diagonal matrix with $\mathbf{D}_{ii} = \pm 1$ depending on whether $\mathbf{v}_i \in \mathcal{L}_t$ is labeled as “no-change” (or “change”) in the ground-truth, and $\mathbf{D}_{ii} = 0$ otherwise.
- **Sup+CA CCA:** this is similar to “standard CCA” with the only difference being β which is set to its “optimal” value (0.01) on the validation set (see Fig. 2).
- **SemiSup CCA:** this setting is similar to “standard CCA” with the only difference being the unlabeled patches which are now added when learning the CCA transformations, and \mathbf{D}_{ii} (on the unlabeled patches) is set to $2\kappa(\mathbf{v}_i, \mathbf{u}_i) - 1$ (score between -1 and $+1$); here $\kappa(\cdot, \cdot) \in [0, 1]$ is the RBF similarity whose scale is set to the 0.1 quantile of pairwise distances in $\mathcal{L}_t \times \mathcal{L}_r$.
- **SemiSup+CA CCA:** this setting is similar to “SemiSup CCA” but context regularization is used (with again β set to 0.01).
- **Res CCA:** this is similar to “standard CCA”, but strict data pairing is *relaxed*, i.e., each patch in the reference image is assigned to multiple patches in the test image; hence, \mathbf{D} is no longer diagonal, and set as $\mathbf{D}_{ij} = \kappa(\mathbf{v}_i, \mathbf{u}_j) \in [0, 1]$ iff $(\mathbf{v}_i, \mathbf{u}_j) \in \mathcal{L}_t \times \mathcal{L}_r$ is labeled as “no-change” in the ground-truth, $\mathbf{D}_{ij} = \kappa(\mathbf{v}_i, \mathbf{u}_j) - 1 \in [-1, 0]$ iff $(\mathbf{v}_i, \mathbf{u}_j) \in \mathcal{L}_t \times \mathcal{L}_r$ is labeled as “change” and $\mathbf{D}_{ij} = 0$ otherwise.

⁵ From which a subset of 1000 is used for validation (as a dev set).

- **Res+Sup+CA CCA:** this is similar to “Res CCA” with the only difference being β which is again set to 0.01.
- **Res+SemiSup CCA:** this setting is similar to “Res CCA” with the only difference being the unlabeled patches which are now added when learning the CCA transformations; on these unlabeled patches $\mathbf{D}_{ij} = 2\kappa(\mathbf{v}_i, \mathbf{u}_j) - 1$.
- **Res+SemiSup+CA CCA:** this setting is similar to “Res+SemiSup CCA” but context regularization is used (i.e., $\beta = 0.01$).

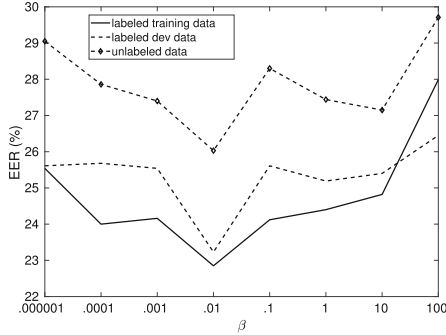


Fig. 2. This figure shows the evolution of change detection performances w.r.t β on labeled training/dev data as well as the unlabeled data. These results correspond to the baseline Sup+CA CCA (under the regime of strong misalignments); we observe from these curves that $\beta = 0.01$ is the best setting which is kept in all our experiments.

Context setting: in order to build the adjacency matrices of the context (see Sect. 2.3), we define for each patch $\mathbf{u}_i \in \mathcal{I}_r$ (in the reference image) an anisotropic (typed) neighborhood system $\{\mathcal{N}_c(i)\}_{c=1}^C$ (with $C = 8$) which corresponds to the eight spatial neighbors of \mathbf{u}_i in a regular grid [52]; for instance when $c = 1$, $\mathcal{N}_1(i)$ corresponds to the top-left neighbor of \mathbf{u}_i . Using $\{\mathcal{N}_c(\cdot)\}_{c=1}^8$, we build for each c an intrinsic adjacency matrix \mathbf{W}_u^c whose $(i, k)^{\text{th}}$ entry is set as $\mathbf{W}_{u,i,k}^c \propto \mathbb{1}_{\{k \in \mathcal{N}_c(i)\}}$; here $\mathbb{1}_{\{\cdot\}}$ is the indicator function equal to 1 iff (i) the patch \mathbf{u}_k is neighbor to \mathbf{u}_i and (ii) its relative position is typed as c ($c = 1$ for top-left, $c = 2$ for left, etc. following an anticlockwise rotation), and 0 otherwise. Similarly, we define the matrices $\{\mathbf{W}_v^c\}_c$ for data $\{\mathbf{v}_j\}_j \in \mathcal{I}_t$.

Impact of AA CCA and Comparison. Table 2 shows a comparison of different versions of AA CCA against other CCA variants under the regime of small residual alignment errors. In this regime, reference and test images are first registered using RANSAC [53]; an exhaustive visual inspection of the overlapping (reference and test) images (after RANSAC registration) shows sharp boundaries in most of the areas covered by these images, but some areas still include residual misalignments due to the presence of changes, occlusions (clouds, etc.) as well as parallax. Note that, in spite of the relative success of RANSAC in registering

Table 1. This table shows different configurations of CCA resulting from different instances of our model. In this table, “Sup” stands for supervised, “SemiSup” for semi-supervised, “CA” for context aware and “Res” for resilient.

Pairing	CCA learning	Context regularization	Designation
strict	supervised	no	Standard CCA
strict	supervised	yes	Sup+CA CCA
strict	semi-sup	no	SemiSup CCA
strict	semi-sup	yes	SemiSup+CA CCA
relaxed	supervised	no	Res CCA
relaxed	supervised	yes	Res+Sup+CA CCA
relaxed	semi-sup	no	Res+SemiSup CCA
relaxed	semi-sup	yes	Res+SemiSup+CA CCA

these images, our AA CCA versions (rows #5–8) provide better performances (see Table 2) compared to the other CCAs (rows #1–4); this clearly corroborates the fact that residual alignment errors remain after RANSAC (re)alignment (as also observed during visual inspection of RANSAC registration). Put differently, our AA CCA method is not an opponent to RANSAC but complementary.

These results also show that when reference and test images are globally well aligned (with some residual errors; see Table 2), the gain in performance is dominated by the positive impact of alignment resilience; indeed, the impact of the unlabeled data is not always consistent (#5,6 vs. #7,8 resp.) in spite of being positive (in #1,2 vs. #3,4 resp.) while the impact of context regularization is globally positive (#1,3,5,7 vs. #2,4,6,8 resp.). This clearly shows that, under the regime of small residual errors, the use of labeled data is already enough in order to enhance the performance of change detection; the gain comes essentially

Table 2. This table shows change detection EER (in %) on labeled (training and validation) and unlabeled sets under the residual error regime. When context regularization (referred to as CA in this table) is used, β is set to 10^{-2} .

#	Configurations	Labeled (train)	Labeled (dev)	Unlabeled
1	Standard CCA	14.91	15.18	12.81
2	Sup+CA CCA	12.95	14.90	11.44
3	SemiSup CCA	11.26	12.80	11.18
4	SemiSup+CA CCA	12.57	11.82	09.96
5	Res CCA	05.81	04.97	05.38
6	Res+Sup+CA CCA	06.35	05.53	05.55
7	Res+SemiSup CCA	08.60	08.74	08.33
8	Res+SemiSup+CA CCA	08.77	08.60	06.94

Table 3. This table shows change detection EER (in %) on labeled (training and validation) and unlabeled sets under the strong error regime. When context regularization (referred to as CA in this table) is used, β is set to 10^{-2} .

#	Configurations	Labeled (train)	Labeled (dev)	Unlabeled
1	Standard CCA	25.63	25.61	28.44
2	Sup+CA CCA	22.85	23.23	26.03
3	SemiSup CCA	22.31	23.58	24.99
4	SemiSup+CA CCA	25.74	25.40	25.47
5	Res CCA	16.42	14.34	19.67
6	Res+Sup+CA CCA	16.55	16.80	19.90
7	Res+SemiSup CCA	19.01	19.24	19.55
8	Res+SemiSup+CA CCA	23.71	21.55	26.76

from alignment resilience with a marginal (but clear) positive impact of context regularization.

In order to study the impact of AA 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; precisely, each pixel is shifted along a direction whose x–y coordinates are randomly set to values between 15 and 30 pixels. These shifts are sufficient in order to make the quality of alignments used for CCA very weak so the different versions of CCA, mentioned earlier, become more sensitive to alignment errors (EERs increase by more than 100% in Table 3 compared to EERs with residual alignment errors in Table 2). With this setting, AA CCA is clearly more resilient and shows a substantial relative gain compared to the other CCA versions.

3.3 Discussion

Invariance: resulting from its misalignment resilience, it is easy to see that our AA CCA is *de facto* robust to local deformations as these deformations are strictly equivalent to local misalignments. It is also easy to see that our AA CCA may achieve invariance to similarity transformations; indeed, the matrices used to define the spatial context are translation invariant, and can also be made rotation and scale invariant by measuring a “characteristic” scale and orientation of patches in a given satellite image. For that purpose, dense SIFT can be used to recover (or at least approximate) the field of orientations and scales, and hence adapt the spatial support (extent and orientation) of context using the characteristic scale, in order to make context invariant to similarity transformations.

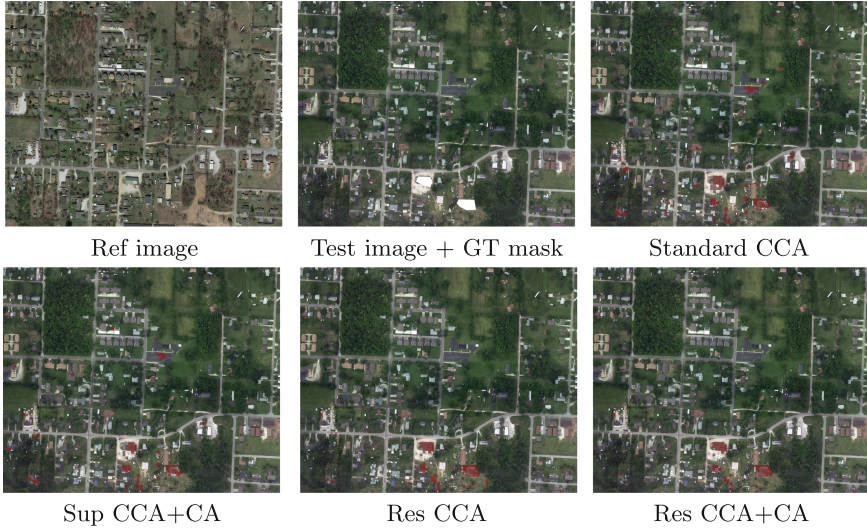


Fig. 3. These examples show the evolution of detections (in red) for four different settings of CCA; as we go from top-right to bottom-right, change detection results get better. CCA acronyms shown below pictures are already defined in Table 1. (Color figure online)

Computational Complexity: provided that VGG-features are extracted (offline) on all the patches of the reference/test images, and provided that the adjacency matrices of context are precomputed⁶, and since the adjacency matrices $\{\mathbf{W}_u^c\}_c, \{\mathbf{W}_v^c\}_c$ are very sparse, the computational complexity of evaluating Eq. (6) and solving the generalized eigenproblem in Eq. (5) both reduce to $O(\min(d_u^2 d_v, d_v^2 d_u))$, here d_u, d_v are again the dimensions of data in \mathbf{U}, \mathbf{V} respectively; hence, this complexity is very equivalent to standard CCA which also requires solving a generalized eigenproblem. Therefore, the gain in the accuracy of our AA CCA is obtained without any overhead in the computational complexity that remains dependent on dimensionality of data (which is, in practice, smaller compared to the cardinality of our datasets) (see also Fig. 3).

4 Conclusion

We introduced in this paper a new canonical correlation analysis method that learns projection matrices which map data from input spaces to a latent common space where unaligned data become strongly or weakly correlated depending on their cross-view similarity and their context. This is achieved by optimizing a

⁶ Note that the adjacency matrices of the spatial neighborhood system can be computed offline once and reused.

criterion that mixes two terms: the first one aims at maximizing the correlations between data which are likely to be paired while the second term acts as a regularizer and 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 projections while being resilient to alignment errors. Extensive experiments show the substantial gain of our CCA method under the regimes of residual and strong alignment errors.

As a future work, our CCA method can be extended to many other tasks where alignments are error-prone and when context can be exploited in order to recover from these alignment errors. These tasks include “text-to-text” alignment in multilingual machine translation, as well as “image-to-image” matching in multi-view object tracking.

A Appendix (Proof of Proposition 2)

We will prove that Ψ is L -Lipschitzian, with $L = \frac{\beta}{\gamma_{\min}} (\sum_c \|\mathbf{E}_c \mathbf{1}_{vu} \mathbf{F}'_c\|_1 + \sum_c \|\mathbf{G}_c \mathbf{1}_{vu} \mathbf{H}'_c\|_1)$. For ease of writing, we omit in this proof the subscripts t, r in \mathbf{K}_{tr} (unless explicitly required and mentioned).

Given two matrices $\mathbf{K}^{(2)}, \mathbf{K}^{(1)}$, we have $\|\mathbf{K}^{(2)} - \mathbf{K}^{(1)}\|_1 = (*)$ with

$$(*) = \beta \left\| \sum_c \mathbf{V} \mathbf{W}_v^c \mathbf{V}' (\mathbf{P}_t^{(1)} \mathbf{P}'_r^{(1)} - \mathbf{P}_t^{(0)} \mathbf{P}'_r^{(0)}) \mathbf{U} \mathbf{W}_u^c \mathbf{U}' + \sum_c \mathbf{V} \mathbf{W}_v^{c'} \mathbf{V}' (\mathbf{P}_t^{(1)} \mathbf{P}'_r^{(1)} - \mathbf{P}_t^{(0)} \mathbf{P}'_r^{(0)}) \mathbf{U} \mathbf{W}_u^c \mathbf{U}' \right\|_1. \quad (12)$$

Using Eq. (5), one may write

$$\mathbf{P}_t \mathbf{P}'_r = \frac{1}{\gamma} \mathbf{C}_{tt}^{-1} \mathbf{K}_{tr} \mathbf{C}_{rr}^{-1}, \quad (13)$$

which also results from the fact that $\mathbf{K}_{rt} \mathbf{C}_{tt}^{-1} \mathbf{K}_{tr}$ is Hermitian and \mathbf{C}_{rr} is positive semi-definite. By adding the superscript τ in $\mathbf{P}_t, \mathbf{P}'_r, \gamma, \mathbf{K}_{tr}$ (with $\tau = 0, 1$), omitting again the subscripts t, r in \mathbf{K}_{tr} and then plugging (13) into (12) we obtain

$$(*) = \beta \left\| \sum_c \mathbf{E}_c \left(\frac{1}{\gamma^{(1)}} \mathbf{K}^{(1)} - \frac{1}{\gamma^{(0)}} \mathbf{K}^{(0)} \right) \mathbf{F}'_c + \sum_c \mathbf{G}_c \left(\frac{1}{\gamma^{(1)}} \mathbf{K}^{(1)} - \frac{1}{\gamma^{(0)}} \mathbf{K}^{(0)} \right) \mathbf{H}'_c \right\|_1 \\ \leq \frac{\beta}{\gamma_{\min}} \left\| \sum_c \mathbf{E}_c (\mathbf{K}^{(1)} - \mathbf{K}^{(0)}) \mathbf{F}'_c \right\|_1 + \frac{\beta}{\gamma_{\min}} \left\| \sum_c \mathbf{G}_c (\mathbf{K}^{(1)} - \mathbf{K}^{(0)}) \mathbf{H}'_c \right\|_1, \quad (14)$$

here γ_{\min} is the lower bound of the eigenvalues of (5) which can be derived (see for instance [54]). Considering $\mathbf{K}_{k,\ell}$ as the $(k, \ell)^{\text{th}}$ entry of \mathbf{K} , we have

$$\begin{aligned}
 (*) &\leq \frac{\beta}{\gamma_{\min}} \sum_{i,j} \left| \sum_{k,\ell} (\mathbf{K}_{k,\ell}^{(1)} - \mathbf{K}_{k,\ell}^{(0)}) \sum_c \mathbf{E}_{c,i,k} \mathbf{F}_{c,j,\ell} \right| \\
 &\quad + \frac{\beta}{\gamma_{\min}} \sum_{i,j} \left| \sum_{k,\ell} (\mathbf{K}_{k,\ell}^{(1)} - \mathbf{K}_{k,\ell}^{(0)}) \sum_c \mathbf{G}_{c,i,k} \mathbf{H}_{c,j,\ell} \right| \\
 &\leq \frac{\beta}{\gamma_{\min}} \sum_{i,j} \sum_{k,\ell} |\mathbf{K}_{k,\ell}^{(1)} - \mathbf{K}_{k,\ell}^{(0)}| \sum_c |\mathbf{E}_{c,i,k} \mathbf{F}_{c,j,\ell}| \\
 &\quad + \frac{\beta}{\gamma_{\min}} \sum_{i,j} \sum_{k,\ell} |\mathbf{K}_{k,\ell}^{(1)} - \mathbf{K}_{k,\ell}^{(0)}| \sum_c |\mathbf{G}_{c,i,k} \mathbf{H}_{c,j,\ell}| \\
 &\leq \frac{\beta}{\gamma_{\min}} \sum_{k,\ell} |\mathbf{K}_{k,\ell}^{(1)} - \mathbf{K}_{k,\ell}^{(0)}| \\
 &\quad \times \left(\sum_{i,j} \sum_{k,\ell,c} |\mathbf{E}_{c,i,k} \mathbf{F}_{c,j,\ell}| + \sum_{i,j} \sum_{k,\ell,c} |\mathbf{G}_{c,i,k} \mathbf{H}_{c,j,\ell}| \right) \\
 &\quad \text{(as } \sum_i |a_i| \cdot |b_i| \leq \sum_{i,j} |a_i| \cdot |b_j|, \forall \{a_i\}_i, \{b_j\}_j \subset \mathbb{R}) \\
 &= L \|\mathbf{K}^{(1)} - \mathbf{K}^{(0)}\|_1,
 \end{aligned}$$

$$\text{with } L = \frac{\beta}{\gamma_{\min}} \left(\sum_c \|\mathbf{E}_c \mathbf{1}_{vu} \mathbf{F}'_c\|_1 + \sum_c \|\mathbf{G}_c \mathbf{1}_{vu} \mathbf{H}'_c\|_1 \right) \quad \square$$

References

1. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: *Advances in Neural Information Processing Systems*, pp. 1097–1105 (2012)
2. Russakovsky, O., et al.: ImageNet large scale visual recognition challenge. *Int. J. Comput. Vis.* **115**(3), 211–252 (2015)
3. Boujemaa, N., Fleuret, F., Gouet, V., Sahbi, H.: Visual content extraction for automatic semantic annotation of video news. In: *The proceedings of the SPIE Conference, San Jose, CA*, vol. 6 (2004)
4. Wang, L., Sahbi, H.: Directed acyclic graph kernels for action recognition. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 3168–3175 (2013)
5. Tollari, S., et al.: A comparative study of diversity methods for hybrid text and image retrieval approaches. In: Peters, C., et al. (eds.) *CLEF 2008. LNCS*, vol. 5706, pp. 585–592. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-04447-2_72
6. Hussain, M., Chen, D., Cheng, A., Wei, H., Stanley, D.: Change detection from remotely sensed images: from pixel-based to object-based approaches. *ISPRS J. Photogrammetry Remote Sens.* **80**, 91–106 (2013)
7. Bourdis, N., Marraud, D., Sahbi, H.: Spatio-temporal interaction for aerial video change detection. In: *2012 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 2253–2256. IEEE (2012)
8. Sahbi, H., Boujemaa, N.: Coarse-to-fine support vector classifiers for face detection. In: null, p. 30359. IEEE (2002)

9. Li, X., Sahbi, H.: Superpixel-based object class segmentation using conditional random fields. In: 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1101–1104. IEEE (2011)
10. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
11. Goodfellow, I., Bengio, Y., Courville, A.: Deep learning. Book in preparation for MIT Press (2016). <http://www.deeplearningbook.org>
12. Jiu, M., Sahbi, H.: Nonlinear deep kernel learning for image annotation. IEEE Trans. Image Process. **26**(4), 1820–1832 (2017)
13. Erhan, D., Bengio, Y., Courville, A., Manzagol, P.A., Vincent, P., Bengio, S.: Why does unsupervised pre-training help deep learning? J. Mach. Learn. Res. **11**, 625–660 (2010)
14. Yosinski, J., Clune, J., Bengio, Y., Lipson, H.: How transferable are features in deep neural networks? In: Advances in Neural Information Processing Systems, pp. 3320–3328 (2014)
15. Doersch, C., Zisserman, A.: Multi-task self-supervised visual learning. arXiv preprint [arXiv:1708.07860](https://arxiv.org/abs/1708.07860) (2017)
16. Richter, S.R., Vineet, V., Roth, S., Koltun, V.: Playing for data: ground truth from computer games. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016, Part II. LNCS, vol. 9906, pp. 102–118. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46475-6_7
17. Hotelling, H.: Relations between two sets of variates. Biometrika **28**(3/4), 321–377 (1936)
18. Anderson, T.W.: An Introduction to Multivariate Statistical Analysis, vol. 2. Wiley, New York (1958)
19. Hardoon, D.R., Szedmak, S., Shawe-Taylor, J.: Canonical correlation analysis: an overview with application to learning methods. Neural Comput. **16**(12), 2639–2664 (2004)
20. Sahbi, H.: Interactive satellite image change detection with context-aware canonical correlation analysis. IEEE Geosci. Remote Sens. Lett. **14**(5), 607–611 (2017)
21. Melzer, T., Reiter, M., Bischof, H.: Appearance models based on kernel canonical correlation analysis. Patt. Recogn. **36**(9), 1961–1971 (2003)
22. Hardoon, D.R., Shawe-Taylor, J.: Sparse canonical correlation analysis. Mach. Learn. **83**(3), 331–353 (2011)
23. Witten, D.M., Tibshirani, R.J.: Extensions of sparse canonical correlation analysis with applications to genomic data. Stat. Appl. Genet. Mol. Biol. **8**(1), 1–27 (2009)
24. Zhang, Z., Zhao, M., Chow, T.W.: Binary-and multi-class group sparse canonical correlation analysis for feature extraction and classification. IEEE Trans. Knowl. Data Eng. **25**(10), 2192–2205 (2013)
25. Vía, J., Santamaría, I., Pérez, J.: A learning algorithm for adaptive canonical correlation analysis of several data sets. Neural Netw. **20**(1), 139–152 (2007)
26. Sun, T., Chen, S.: Locality preserving cca with applications to data visualization and pose estimation. Image Vis. Comput. **25**(5), 531–543 (2007)
27. Zhai, D., Zhang, Y., Yeung, D.Y., Chang, H., Chen, X., Gao, W.: Instance-specific canonical correlation analysis. Neurocomputing **155**, 205–218 (2015)
28. Yger, F., Berar, M., Gasso, G., Rakotomamonjy, A.: Adaptive canonical correlation analysis based on matrix manifolds. arXiv preprint [arXiv:1206.6453](https://arxiv.org/abs/1206.6453) (2012)
29. De la Torre, F.: A unification of component analysis methods. In: Handbook of Pattern Recognition and Computer Vision, pp. 3–22 (2009)

30. Sun, J., Keates, S.: Canonical correlation analysis on data with censoring and error information. *IEEE Trans. Neural Netw. Learn. Syst.* **24**(12), 1909–1919 (2013)
31. Sun, L., Ji, S., Ye, J.: Canonical correlation analysis for multilabel classification: a least-squares formulation, extensions, and analysis. *IEEE Trans. Patt. Anal. Mach. Intell.* **33**(1), 194–200 (2011)
32. Haghighat, M., Abdel-Mottaleb, M.: Low resolution face recognition in surveillance systems using discriminant correlation analysis. In: 2017 12th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2017), pp. 912–917. IEEE (2017)
33. Ferecatu, M., Sahbi, H.: Multi-view object matching and tracking using canonical correlation analysis. In: 2009 16th IEEE International Conference on Image Processing (ICIP), pp. 2109–2112. IEEE (2009)
34. Loy, C.C., Xiang, T., Gong, S.: Multi-camera activity correlation analysis. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2009, pp. 1988–1995. IEEE (2009)
35. Zhou, F., Torre, F.: Canonical time warping for alignment of human behavior. In: *Advances in Neural Information Processing Systems*, pp. 2286–2294 (2009)
36. Zhou, F., De la Torre, F.: Generalized canonical time warping. *IEEE Trans. Patt. Anal. Mach. Intell.* **38**(2), 279–294 (2016)
37. Kim, T.K., Cipolla, R.: Canonical correlation analysis of video volume tensors for action categorization and detection. *IEEE Trans. Patt. Anal. Mach. Intell.* **31**(8), 1415–1428 (2009)
38. Fischer, B., Roth, V., Buhmann, J.M.: Time-series alignment by non-negative multiple generalized canonical correlation analysis. *BMC Bioinform.* **8**(10), S4 (2007)
39. Trigeorgis, G., Nicolaou, M., Zafeiriou, S., Schuller, B.: Deep canonical time warping for simultaneous alignment and representation learning of sequences. *IEEE Trans. Patt. Anal. Mach. Intell.* **40**(5), 1128–1138 (2017)
40. Ham, J., Lee, D.D., Saul, L.K.: Semisupervised alignment of manifolds. In: *AISTATS*, pp. 120–127 (2005)
41. Lafon, S., Keller, Y., Coifman, R.R.: Data fusion and multicue data matching by diffusion maps. *IEEE Trans. Patt. Anal. Mach. Intell.* **28**(11), 1784–1797 (2006)
42. Wang, C., Mahadevan, S.: Manifold alignment using procrustes analysis. In: *Proceedings of the 25th International Conference on Machine Learning*, pp. 1120–1127. ACM (2008)
43. Luo, B., Hancock, E.R.: Iterative procrustes alignment with the EM algorithm. *Image Vis. Comput.* **20**(5), 377–396 (2002)
44. Feuz, K.D., Cook, D.J.: Collegial activity learning between heterogeneous sensors. *Knowl. Inf. Syst.* 1–28 (2017)
45. Sahbi, H., Li, X.: Context-based support vector machines for interconnected image annotation. In: Kimmel, R., Klette, R., Sugimoto, A. (eds.) *ACCV 2010, Part I*. LNCS, vol. 6492, pp. 214–227. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-19315-6_17
46. Sahbi, H.: CNRS-TELECOM ParisTech at imageCLEF 2013 scalable concept image annotation task: winning annotations with context dependent SVMs. In: *CLEF (Working Notes)* (2013)
47. Sahbi, H.: Discriminant canonical correlation analysis for interactive satellite image change detection. In: *IGARSS*, pp. 2789–2792 (2015)
48. Bourdis, N., Denis, M., Sahbi, H.: Constrained optical flow for aerial image change detection. In: 2011 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 4176–4179 (2011)

49. Bourdis, N., Denis, M., Sahbi, H.: Camera pose estimation using visual servoing for aerial video change detection. In: 2012 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 3459–3462 (2012)
50. Sahbi, H.: Relevance feedback for satellite image change detection. In: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1503–1507. IEEE (2013)
51. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. CoRR abs/1409.1556 (2014)
52. Thiemert, S., Sahbi, H., Steinebach, M.: Applying interest operators in semi-fragile video watermarking. In: Security, Steganography, and Watermarking of Multimedia Contents VII, vol. 5681, International Society for Optics and Photonics, pp. 353–363 (2005)
53. Kim, T., Im, Y.J.: Automatic satellite image registration by combination of matching and random sample consensus. IEEE Trans. Geosci. Remote Sens. **41**(5), 1111–1117 (2003)
54. Lu, L.Z., Pearce, C.E.M.: Some new bounds for singular values and eigenvalues of matrix products. Ann. Oper. Res. **98**(1), 141–148 (2000)