Canonical Analysis of Sentinel-1 Radar and Sentinel-2 Optical Data

Allan A. Nielsen[(⊠)] and Rasmus Larsen

DTU Compute – Applied Mathematics and Computer Science,
Technical University of Denmark,
Richard Petersens Plads, Building 321, 2800 Lyngby, Denmark
{alan,rlar}@dtu.dk
http://people.compute.dtu.dk/alan

Abstract. This paper gives results from joint analyses of dual polarimety synthetic aperture radar data from the Sentinel-1 mission and optical data from the Sentinel-2 mission. The analyses are carried out by means of traditional canonical correlation analysis (CCA) and canonical information analysis (CIA). Where CCA is based on maximising correlation between linear combinations of the two data sets, CIA maximises mutual information between the two. CIA is a conceptually more pleasing method for the analysis of data with very different modalities such as radar and optical data. Although a little inconclusive as far as the change detection aspect is concerned, results show that CIA analysis gives conspicuously less noisy appearing images of canonical variates (CVs) than CCA. Also, the 2D histogram of the mutual information based leading CVs clearly reveals much more structure than the correlation based one. This gives promise for potentially better change detection results with CIA than can be obtained by means of CCA.
http://www.imm.dtu.dk/pubdb/p.php?6963.

Keywords: Canonical correlation analysis · Canonical information analysis

1 Introduction

In a preliminary investigation into change detection in data of different modalities this paper looks into canonical analysis of Sentinel-1[1] (S1) radar and Sentinel-2[2] (S2) optical remote sensing data. This kind of analysis is potentially important due to often occurring clouds in optical data and the all-weather acquisition capability of radar. The data are analysed both by means of traditional canonical correlation analysis (CCA) [4] and a more computer intensive method which we have named canonical information analysis (CIA) [13,14].

Earlier we have worked with CCA based change detection in bi-temporal optical data. We termed this method multivariate alteration detection (MAD)

---

[1] https://sentinel.esa.int/web/sentinel/missions/sentinel-1.
[2] https://sentinel.esa.int/web/sentinel/missions/sentinel-2.

[10], and we also published an iterative version termed iteratively re-weighted MAD (IR-MAD or iMAD) [2,9]. The so-called MAD variates containing information on change are the differences between pairs of corresponding canonical variates.

CCA considers second order statistics of the involved variables only and as such it is ideal for Gaussian data. CCA is therefore not necessarily an obvious choice of analysis for data of such different modalities as here with radar and optical data (which have very different genesis and potentially follow very different statistical distributions), but since it is the basis of the widely used MAD methods we have included it here.

The idea in CIA is to replace correlation as the measure of association between variables with the more general information theoretical, entropy based measure mutual information (MI) [1,2,5,8]. Also in this case the change information is found in the differences between corresponding pairs of canonical variates. Since MI between variables is independent of their signs, this differencing is a little more tricky for MI based analysis than for correlation based analysis.

Other workers have dealt with canonical analysis based on mutual information [6,15]. Entropy and mutual information depend on the sample probability density functions of the involved variables and thus on higher order statistics.

In a situation with data of very different modalities the use of MI constitutes a conceptually much more pleasing way of measuring association between variables than the use of correlation.

For an illustrative (toy) example where CCA fails and CIA succeeds, (an example with RGB images covering a busy motorway for traffic surveillance), and an example of joint analysis of weather radar and Meteosat data, see [13].

2   Methods

Here we very briefly sketch the ideas in canonical correlation analysis, canonical information analysis, and basic information theory. In both CCA and CIA linear combinations $U = \boldsymbol{a}^T \boldsymbol{X}$ and $V = \boldsymbol{b}^T \boldsymbol{Y}$ of two sets of stochastic variables, k-dimensional $\boldsymbol{X}$ (here the Sentinel-1 data) and ℓ-dimensional $\boldsymbol{Y}$ (here the Sentinel-2 data), are determined.

2.1   Canonical Correlation Analysis

In canonical correlation analysis first published by Hotelling in 1936 [4] linear combinations which maximise correlation between U and V are found. Correlation considers second order statistics of the involved variables only. It is therefore ideal for Gaussian data. The canonical variates U and V are found by solving a generalised eigenvalue problem.

2.2   Canonical Information Analysis

Inspired by canonical correlation analysis, canonical information analysis is a method which replaces the maximisation of correlation with maximisation of

mutual information between the linear combinations U and V. For further details see the next subsection and [11,13,14].

2.3 Basic Information Theory

In 1948 Shannon [12] published his now classical work on information theory. Below, we describe the information theoretical concepts entropy, relative entropy and mutual information for discrete stochastic variables, see also [1,2,5,8].

Entropy. Consider a discrete stochastic variable X with probability density function (pdf) $p(X = x_i)$, $i = 1, \ldots, n$, i.e., the probability of observing a particular realization x_i of stochastic variable X, where n is the number of possible outcomes or the number of bins. Let us look for a measure of information content (or surprise if you like) $h(X = x_i)$ in obtaining that particular realization. If x_i is a very probable value, i.e., $p(X = x_i)$ is high, we receive little information by observing x_i. If on the other hand x_i is a very improbable value, i.e., $p(X = x_i)$ is low, we receive much information by observing x_i. The measure of information content should be a monotonically decreasing function of p. This can be obtained by choosing for example $h \propto 1/p$.

If we observe independent realizations x_i and x_j, i.e., the two-dimensional pdf $p(X = x_i, X = x_j)$ equals the product of the one-dimensional marginal pdfs $p(X = x_i)p(X = x_j)$, we would like the joint information content to equal the sum of the marginal information contents, i.e., $h(X = x_i, X = x_j) = h(X = x_i) + h(X = x_j)$. This can be obtained by transformation by means of the logarithm.

Thus the desired characteristics of the measure of information or surprise can be obtained if we define $h(X = x_i)$ as

$$h(X = x_i) = \ln \frac{1}{p(X = x_i)} = -\ln p(X = x_i). \tag{1}$$

The expectation $H(X)$ of the information measure, i.e., the average amount of information obtained by observing the stochastic variable X, is termed the entropy

$$H(X) = -\sum_{i=1}^{n} p(X = x_i) \ln p(X = x_i). \tag{2}$$

In the limit where p tends to zero and $\ln p$ tends to minus infinity, $-p \ln p$ tends to zero. $H(X) = -\mathrm{E}\{\ln p(X)\}$ is nonnegative. A discrete variable which takes on one value only has zero entropy; a uniform discrete variable has maximum entropy (equal to $\ln n$). For the joint entropy of two discrete stochastic variables X and Y we get

$$H(X, Y) = -\sum_{i,j} p(X = x_i, Y = y_j) \ln p(X = x_i, Y = y_j). \tag{3}$$

Relative Entropy. The relative entropy also known as the Kullback-Leibler divergence [7] between two pdfs $p(X = x_i)$ and $q(X = x_i)$ defined on the same set of outcomes (or bins) is

$$D_{KL}(p, q) = \sum_i p(X = x_i) \ln \frac{p(X = x_i)}{q(X = x_i)}. \tag{4}$$

This is the expectation of the logarithmic difference between p and q. $D_{KL} \geq 0$ with equality for $p(X = x_i) = q(X = x_i)$ only. The relative entropy is not symmetric in p and q (and therefore it is not a metric).

Mutual Information. The extent to which two discrete stochastic variables X and Y are not independent, which is a measure of their mutual information content, may be expressed as the relative entropy or the Kullback-Leibler divergence between the two-dimensional pdf $p(X = x_i, Y = y_j)$ and the product of the one-dimensional marginal pdfs $p(X = x_i)p(Y = y_j)$, i.e.,

$$D_{KL}(p(X, Y), p(X)p(Y)) = \tag{5}$$
$$\sum_{i,j} p(X = x_i, Y = y_j) \ln \frac{p(X = x_i, Y = y_j)}{p(X = x_i)p(Y = y_j)}.$$

This sum defines the mutual information $I(X, Y)$ of the stochastic variables X and Y. Mutual information equals the sum of the two marginal entropies minus the joint entropy

$$I(X, Y) = H(X) + H(Y) - H(X, Y). \tag{6}$$

Unlike the general Kullback-Leibler divergence in (4) this measure is symmetric. Mutual information is always nonnegative, it is zero for independent stochastic variables only.

Obviously we need to estimate marginal as well as joint pdfs to obtain the mutual information estimate in (6). We employ kernel density estimation, which uses N data samples to estimate these pdfs. Mutual information is subsequently estimated using the same N data points. This is possible in practice only due to a very fast estimation of pdfs, see [13].

3   S1 and S2 Data, Frankfurt Airport

Both the radar and the optical data cover the international airport in Frankfurt, Germany. The data are obtained from Google Earth Engine[3] (GEE) [3].

The Sentinel-1 data acquired in instrument Interferometric Wide Swath (IW) mode, is an S1 Ground Range Detected (GRD) scene, processed using the Sentinel-1 Toolbox[4] to generate a calibrated, ortho-corrected product.

---

[3] https://earthengine.google.com, https://developers.google.com/earth-engine.
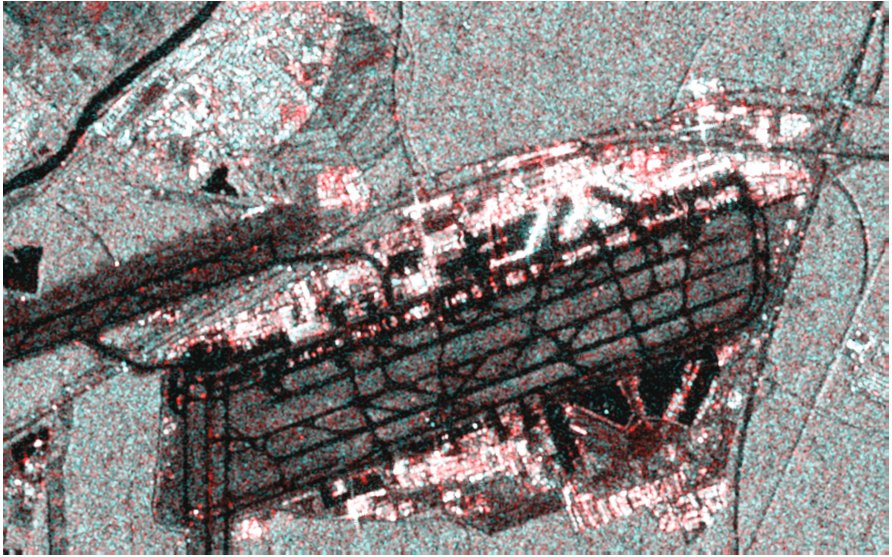[4] https://sentinel.esa.int/web/sentinel/toolboxes/sentinel-1.

Fig. 1. RGB image of Sentinel-1 C-band VV/VH data, VV as R and VH as G and B (i.e., cyan), 10 m pixels, 5 km north-south and 8 Km east-west, Franfurt Airport, Germany, acquired on 15 July 2016. (Color figure online)
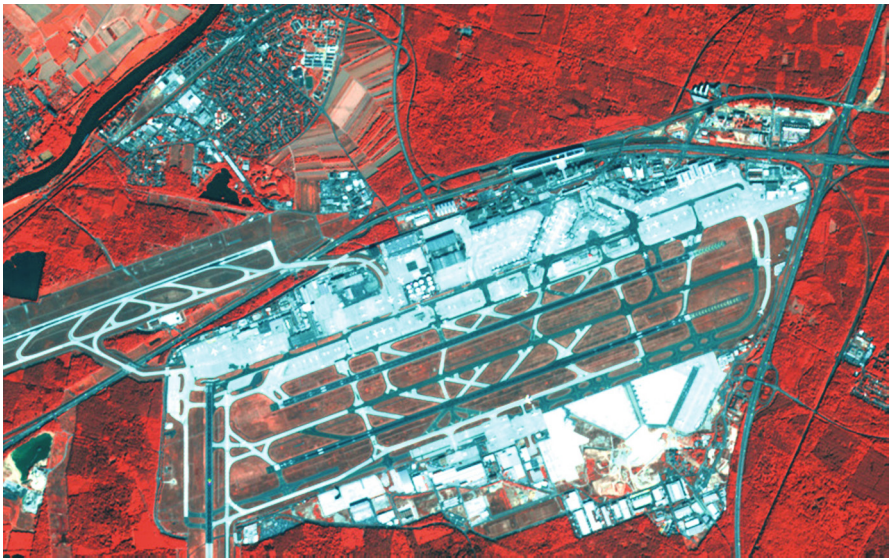

Fig. 2. RGB image of Sentinel-2 MSI data, band 4 (near-infrared as R), band 3 (red as G), and band 2 (green as B), 10 m pixels, 5 km north-south and 8 Km east-west, Franfurt Airport, Germany, acquired on 12 Sep 2016. (Color figure online)

This processing includes thermal noise removal, radiometric calibration, and terrain correction using Shuttle Radar Topography Mission 30 m (SRTM 30) data. Finally it includes saturating the data (quoting GEE): "Values are then clamped to the 1st and 99th percentile to preserve the dynamic range against anomalous outliers, and quantised to 16 bits." This is to avoid excessive precision loss during conversion from floats to integers for storage. The outliers are usually due to strong reflections from sharp angles on antennas and other man-made objects. The spatial resolution is (range by azimuth) 20 m by 22 m and the pixel spacing is 10 m. The IW data are multi-looked, the number of looks is 5 by 1 and the equivalent number of looks is 4.9. We have 500 rows by 800 columns of 10 m pixels. Figure 1 shows an RGB image of Sentinel-1 C-band VV/VH data acquired on 15 July 2016, VV as red and VH as green and blue (i.e., cyan). The data are log transformed.

The Sentinel-2 data are the near-infrared, red, green and blue channels from the MultiSpectral Instrument (MSI), level-1C processed, 500 rows by 800 columns of 10 m pixels. Figure 2 shows an RGB image of bands 4, 3, and 2 (near-infrared, red, and green). These data are acquired on 12 Sep 2016. The saturation issues mentioned for the S1 data are not present for the S2 data.

Here k-dimensional $\boldsymbol{X}$ in the CCA/CIA analyses is the Sentinel-1 VV and HV data, $k = 2$, and the ℓ-dimensional $\boldsymbol{Y}$ is the Sentinel-2 VNIR data, $\ell = 4$.

4   Results

First: in a change detection setting with data of different modalities, the results obtained of course reflect both change over time and the different kinds of information contained in the different data modalities. It is very difficult to discriminate between the two.

A philosophical question is whether change detection between so different data sources/modalities can be carried out in a meaningful fashion at all. Maybe not, but it is always interesting to try to optimise the joint information content in the data irrespective of their origin. And this is exactly what the CIA method aims to do.

Figures 3 and 4 show the leading (CCA based) canonical variates (CV) for the S1 and the S2 data, respectively.

Figures 5 and 6 show the leading (CIA based) canonical variates for the S1 and the S2 data, respectively. Starting weights $\boldsymbol{a}$ and $\boldsymbol{b}$ are the values obtained by CCA.

As stated above CCA is not necessarily an obvious choice of analysis for data of quite different modalities. Neither is therefore the choice of starting values from CCA a good one. Figure 7 shows the leading (CIA based) canonical variate for the S2 data. Starting weights $\boldsymbol{a}$ and $\boldsymbol{b}$ are equal for each set. Table 1 shows correlation and mutual information for the three solutions, CCA, CIA with starting weights from CCA, and CIA with equal starting weights.

Fig. 3. Leading canonical variate from CCA of Sentinel-1 C-band VV/VH data.

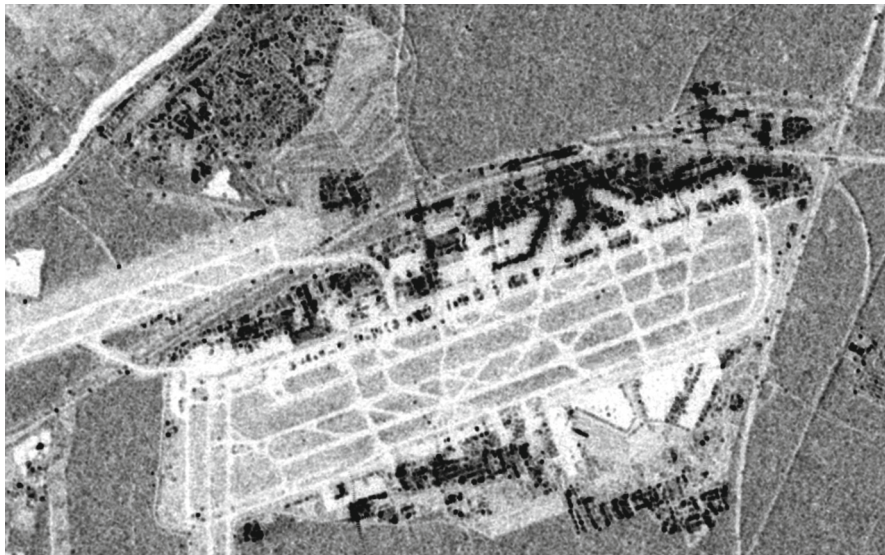**Fig. 4.** Leading canonical variate from CCA of Sentinel-2 VNIR data.

Fig. 5. Leading canonical variate from CIA of Sentinel-1 C-band VV/VH data, starting weights (*a* and *b*) from CCA.
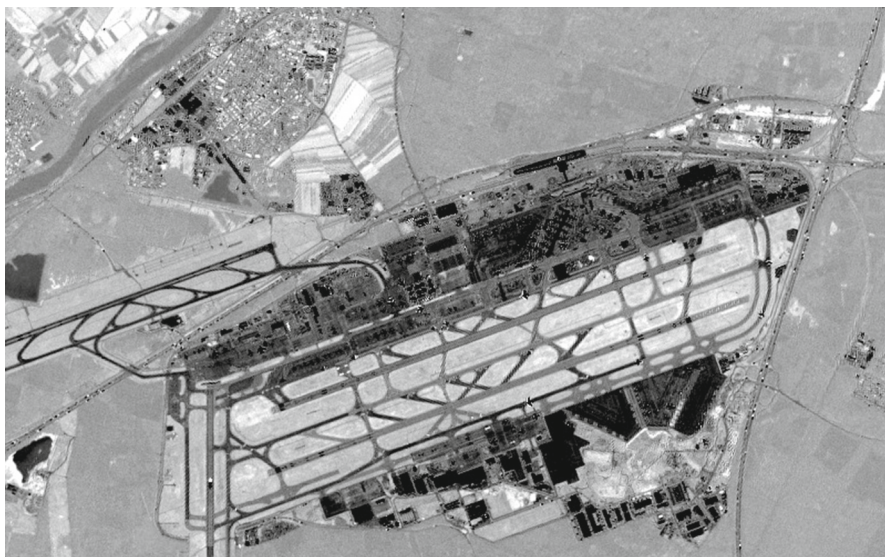

Fig. 6. Leading canonical variate from CIA of Sentinel-2 VNIR data, starting weights (*a* and *b*) from CCA.

Fig. 7. Leading canonical variate from CIA of Sentinel-2 VNIR data, equal starting weights (*a* and *b*) for each set.

Table 1. Correlation and mutual information for the three solutions examined.

	CCA	CIA (cca)	CIA (equal)
Corr	0.4484	0.0305	−0.3297
MI	0.0453	0.2569	0.3183

Figures 8 and 9 show the development for mutual information and the weights (*a* and *b*) over the iterations for the CIA method with both starting conditions.

Figure 10 shows 2D histograms of the leading CVs for both CCA (left) and CIA with equal starting weights (right). CIA based CVs clearly reveal more structure in the data than the CCA CVs. The saturation issue mentioned above is clearly seen, especially in the CCA based S1 CV.

Although the weights in $a = [a_1 \, a_2]^T$ are not equal for the two different sets of starting values there is no visual difference between the two CIA based S1 CVs. Therefore only one of them is shown. On the other hand the two S2 CVs are very different. The solution based on equal weights have a higher MI and much more structure in both built-up areas in the town of Kelsterbach to the north of the airport and in wooded regions.

For both sets of starting values for *a* and *b* we observe the following:

– It is obvious, that the MI based solution is far less noisy for both the S1 and the S2 data and that it reveals more structure.
– The different appearance of the leading S1 CIA CV (compared to the S1 CCA CV) must be due mainly to the drastic change in a_1 associated with the S1 VV data. Apart from the conspicuously less noisy visual appearance,
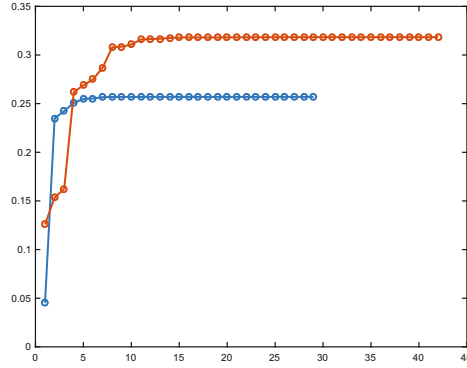
Fig. 8. Development of mutual information over the CIA iterations starting with the CCA solution (in blue) and starting with equal weights for each set (in red). (Color figure online)
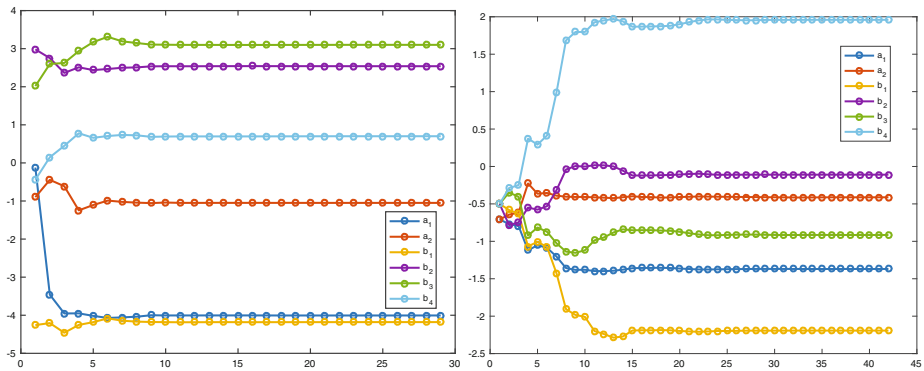

Fig. 9. Development of the weights $[a_1, a_2]^T$ for S1 data and $[b_1, \ldots, b_4]^T$ for S2 data over the CIA iterations starting with the CCA solution for each set (left), and equal weights for each set (right). For the left plot, especially coefficients b_1 and b_4 associated with the S2 blue and near-infrared bands change drastically and b_4 changes sign. Also in this case the coefficient a_1 for S1 VV changes drastically albeit less than b_1 and b_4. For the right plot, especially coefficients b_1 and b_4 associated with the S2 blue and near-infrared bands change drastically and b_4 changes sign. Also in this case the coefficient a_1 for S1 VV changes drastically albeit less than b_1 and b_4. (Color figure online)

we note that the leading S2 CIA CV looks very different from the CCA CV. This is especially true for most taxiways (but not for runways), aprons and other impervious regions, which have high values/appear bright in the CCA solution and have low values/appear dark in the two CIA solutions. Also the wooded areas surrounding the airport have low values/appear dark in the CCA solution whereas they have intermediate or high values/appear gray or bright in the two CIA solutions. This must be due to the change of sign in b_4 associated with the near-infrared channel of the S2 MSI data.
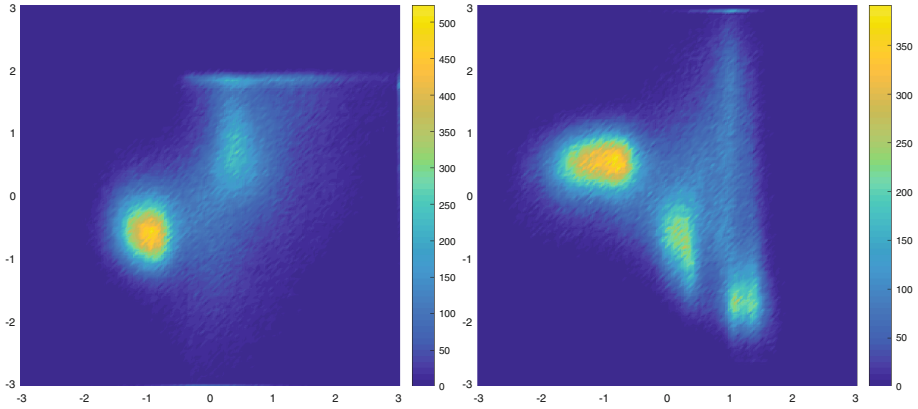
Fig. 10. 2D histograms for leading canonical variates, correlation based (left) and mutual information based with equal starting weights (right). The CVs all have mean value zero and variance one; the axes are stretched between minus and plus three standard deviations. The 2D histogram of the leading MI based CVs clearly reveals more structure in the data than the correlation based one. Also, we see that the MI based CVs are less sensitive to the saturation of the S1 data.

5   Conclusions

Although one may not be able to perform change detection between so different data sources/modalities in a meaningful fashion (here with radar and optical data), it is certainly interesting to try to optimise the joint information content in the data irrespective of their origin. And this is exactly what CIA aims to do.

Starting with equal weights for each set in the search for the CIA solution in our case is better than starting with weights from traditional CCA.

In this preliminary study of canonical analysis of data with different modalities, the results are a little inconclusive as far as the change detection aspect is concerned. Still, results show that CIA gives conspicuously less noisy appearing CV images than CCA. Also, a 2D histogram of the leading MI based CVs clearly reveals more structure in the data than the correlation based one. Finally, the MI based CVs are less sensitive to the saturation of the S1 data. Thus MI based canonical analysis reveals more signal and structure than correlation based analysis, and it gives promise for potentially better change detection results.

References

1. Bishop, C.M.: Pattern Recognition and Machine Learning. Springer, New York (2007)
2. Canty, M.J.: Image Analysis, Classification and Change Detection in Remote Sensing. With Algorithms for ENVI/IDL and Python, 3rd edn. Taylor & Francis, CRC Press (2014)

3. Google Earth Engine Team: Google Earth Engine: A planetary-scale geo-spatial analysis platform (2015). https://earthengine.google.com
4. Hotelling, H.: Relations between two sets of variates. Biometrika **XXVIII**, 321–377 (1936)
5. Hyvärinen, A., Karhunen, J., Oja, E.: Independent Component Analysis. Wiley, New York (2001)
6. Karasuyama, M., Sugiyama, M.: Canonical dependency analysis based on squared-loss mutual information. Neural Netw. **34**, 46–55 (2012)
7. Kullback, S., Leibler, R.A.: On information and sufficiency. Ann. Math. Stat. **22**(1), 79–86 (1951)
8. Mackay, D.J.C.: Information Theory, Inference and Learning Algorithms. Cambridge University Press, Cambridge (2003)
9. Nielsen, A.A.: The regularized iteratively reweighted MAD method for change detection in multi- and hyperspectral data. IEEE Trans. Image Process. **16**(2), 463–478 (2007). http://www2.imm.dtu.dk/pubdb/p.php?4695
10. Nielsen, A.A., Conradsen, K., Simpson, J.J.: Multivariate alteration detection (MAD) and MAF postprocessing in multispectral, bitemporal image data: New approaches to change detection studies. Remote Sens. Environ. **64**(1), 1–19 (1998). http://www2.imm.dtu.dk/pubdb/p.php?1220
11. Nielsen, A.A., Vestergaard, J.S.: Change detection in bi-temporal data by canonical information analysis. In: 8th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp), Annecy, France (2015). http://www.imm.dtu.dk/pubdb/p.php?6888, Matlab code at https://github.com/schackv/cia
12. Shannon, C.E.: A mathematical theory of communication. Bell Syst. Techn. J. **27**(3), 379–423, 623–656 (1948)
13. Vestergaard, J.S., Nielsen, A.A.: Canonical information analysis. ISPRS J. Photogrammetry Remote Sens. **101**, 1–9 (2015). http://authors.elsevier.com/a/1QAnN3I9x1EeMt, http://www.imm.dtu.dk/pubdb/p.php?6270, Matlab code at https://github.com/schackv/cia
14. Vestergaard, J.S.: Interpretation of images from intensity, texture and geometry. Ph.D. thesis, Technical University of Denmark (2015). http://orbit.dtu.dk
15. Yin, X.: Canonical correlation analysis based on information theory. J. Multivar. Anal. **91**, 161–176 (2004)