# ICA AND KERNEL ICA FOR CHANGE DETECTION IN MULTISPECTRAL REMOTE SENSING IMAGES

*Silvia Marchesi and Lorenzo Bruzzone*

Department of Information Engineering and Computer Science, University of Trento,
Via Sommarive, 14 I-38123, Povo, Trento, Italy, Fax: +39-0461-882093
E-mail: lorenzo.bruzzone@ing.unitn.it

## ABSTRACT

In this paper Principal Component Analysis (PCA), Independent Component Analysis (ICA), and Kernel Independent Component Analysis (KICA) are studied and compared in the framework of unsupervised change detection in multitemporal remote sensing images. Different architectures for using the above-mentioned techniques in change detection are investigated, and their capability to discriminate true changes from the different sources of noise analyzed. Experimental results obtained on a pair of very high geometrical resolution Quickbird images point out the main properties of the different methods when applied to change detection.

***Index Terms***— Change detection, principal component analysis, independent component analysis, kernel independent component analysis, multispectral images, remote sensing.

## 1. INTRODUCTION

Change-detection techniques aim at identifying two different classes in multitemporal images: the class of changed areas and that of unchanged ones. Usually change-detection algorithms compare two images acquired at different times on the same geographical area by assuming that they are similar to each other except for the presence of changes on the ground. However, this assumption is seldom completely satisfied due to differences in atmospheric and sunlight conditions of acquired images, as well as in the sensor acquisition geometry (especially with very high resolution (VHR) images). In order to overcome these problems, change-detection techniques generally implement pre-processing steps, which include image co-registration, radiometric and geometric corrections, and noise reduction. Depending on the kind of sensors considered for image acquisition and on the related geometrical resolution, these steps can result in different complexity. Nonetheless, in real problems pre-processing is often not sufficient to guarantee the ideal condition in which radiometric changes in corresponding pixels on the multitemporal images are associated with true changes on the ground. Usually, residual components of noise (e.g. due to residual radiometric differences, residual misregistration, etc.) result in false alarms in the change-detection maps, which cannot be easily identified in the phase of post-processing.

In this paper, we address the aforementioned problem by exploiting data transformation techniques for separating the different sources of noise from real changes in different components to be selectively exploited in the change-detection phase. In particular, we study the effectiveness of ICA and of its kernelized version, i.e. KICA, as a preliminary step to change detection. These techniques are integrated in standard change-detection methods and their performances are analyzed on different data sets, thus deriving general conclusions on their effectiveness in change-detection applications. Comparisons with the standard PCA method are also given.

## 2. NOTATION AND BACKGROUND

Let us consider two multispectral images $\mathbf{x_1}$ and $\mathbf{x_2}$ acquired on the same geographical area at different times, $t_1$ and $t_2$, respectively. Let $\Omega = \{\omega_n, \Omega_c\}$ be the set of classes to be identified. In particular, $\omega_n$ represents the class of no-changed pixels, while $\Omega_c = \{\omega_{c_1}, ..., \omega_{c_K}\}$ the set of the $K$ possible classes of changes occurred in the considered area. The main objective of the present work is to define techniques capable to identify the $K$ different kinds of changes and to separate them from sources of noise (e.g. registration noise). To this purpose we analyze the effectiveness of ICA and KICA transformation integrated in simple change-detection architectures, and compare their performances with those of the well-known and widely used PCA and the standard change vector analysis (CVA). In the following some background concepts on PCA, ICA, and KICA are reported.

A standard transformation approach to isolate changed areas from unchanged areas is that based on the PCA technique [1]. PCA is a linear transformation which exploits image data second order statistics to extract orthogonal components ordered according to decreasing variances. The transformation can be based on eigenvector analysis of the correlation or of the co-variance matrix. The transformed components are globally uncorrelated under Gaussian hypothesis. PCA can be used in change detection either by applying the transformation separately to single date images, or by applying the transformation jointly to the multitemporal images. In many applications, a subset of the resulting transformed components proved to exhibit a more focused representation of the changed areas than the original spectral channels [1]. However PCA is not suitable for separating information sources from sources of noise that are associated with the complexity of many change-detection problems, especially in VHR images [2], acquired by the last generation sensors.

A more suitable methodological tool for discriminating sources of noise form true changes (and potentially to distinguish different kinds of change) is the Independent Component Analysis (ICA), which is intrinsically designed for mapping the information sources present in a complex problem in different components. Nonetheless, marginal attention has been devoted to the use of ICA in change detection, without a detailed analysis of its potentialities [3], [4]. The objective of ICA is to extract components with higher-order statistical independence, through a nonlinear transformation function. ICA assumes a statistical model whereby the observed multivariate data are assumed to be linear or nonlinear mixtures of some unknown latent variables. The mixing coefficients are also unknown. The latent variables are nongaussian and mutually independent and they are called the independent components (sources) of the observed data. In particular, an observed data vector $\mathbf{x}=\left(\mathbf{x}_1,...,\mathbf{x}_m\right)$ is modeled by ICA as $\mathbf{x}=A\mathbf{s}$ where $\mathbf{s}$ is a latent vector with independent components and $A$ is the $m$x$m$ matrix of mixing parameters. Given $N$ i.i.d. observations of $\mathbf{x}$, ICA estimates the mixing matrix $A$ and recovers the latent vector $\mathbf{s}$ corresponding to any particular $\mathbf{x}$. ICA is usually applied by introducing proper contrast functions and iterative procedures capable to optimize them. A considerable portion of open literature is dedicated to define contrast functions associated with the estimation of the mixing matrix $A$ by the Maximum Likelihood principle or by minimizing the mutual information between the components. The obtained components $\mathbf{s}$ are statistically as independent as possible. It is worth noting that the goal of independence is stronger than that of uncorrelatedness which can be obtained on the global data distribution with the PCA technique. It follows that ICA can provide more effective decomposition than PCA, especially for non-Gaussian signals.

Kernel ICA is an approach recently introduced in the literature, in which the ICA problem is not solved on the basis of a single nonlinear function, but on an entire reproducing kernel Hilbert space (RKHS) of candidate nonlinear functions [5]. The idea is to project the input space $\mathbf{x}\in R^m$ to a potentially higher dimensional feature space $F$, through a non-linear mapping $\Phi:\mathbf{x}\in R^m\rightarrow\Phi(\mathbf{x})\in F$, so that the nonlinear relation in the input space can be analyzed in a linear way in this feature space. Instead of considering the given learning problem in input space $R^m$, one can deal with $\Phi(\mathbf{x}_1),...,\Phi(\mathbf{x}_m)$ in the feature space $F$ and then finds a linear discriminant function in the feature space $F$. The use of this kernel trick avoids the need to compute the feature vector in $F$ explicitly. It is sufficient to calculate the inner product of two vectors in $F$ with a kernel function $k(\cdot,\cdot)$ so that: $<\Phi(\mathbf{x}_i),\Phi(\mathbf{x}_j)>=k(\mathbf{x}_i,\mathbf{x}_j)$, where $\left(\cdot,\cdot\right)$ denotes the inner product. In the literature several kernel functions exist, like Gaussian or polynomial kernels. In Kernel ICA presented in [5] and used in this work, the $F$-correlation is defined as the maximal correlation between the random variables $f_1(\mathbf{x}_i)$ and $f_2(\mathbf{x}_j)$, where $f_1$ and $f_2$ range over $F$:

$$\rho_F=\max_{f_1,f_2\in F}\operatorname{corr}\left(f_1(\mathbf{x}_i),f_2(\mathbf{x}_j)\right)=\max_{f_1,f_2\in F}\frac{\operatorname{cov}<f_1(\mathbf{x}_i),f_2(\mathbf{x}_j)>}{\left[\operatorname{var}f_1(\mathbf{x}_i)\right]^{1/2}\left[\operatorname{var}f_2(\mathbf{x}_j)\right]^{1/2}} \quad (1)$$

If the variables $\mathbf{x}_i$ and $\mathbf{x}_j$ are independent, then the $F$-correlation is equal to zero. The use of a function space makes it possible to adapt the technique to a variety of sources and thus makes this algorithm more robust to varying source distributions. However, this is obtained at the cost of a significantly increased computational load.

## 3. CHANGE-DETECTION STRATEGIES

The proposed architectures exploits a simple change-detection scheme based on multitemporal images comparison and thresholding, after a preliminary phase based on image transformation. The main idea is that through the transformation of the images it is possible to identify and separate changes from other information sources (noise, unchanged areas, etc.). Then, through a simple analysis of the components that contain change information it is possible to identify true changed areas.

Two different transformation strategies were investigated: i) all the spectral channels of the two multitemporal images were jointly transformed; and ii) the multispectral difference image $\mathbf{x_D}$ (obtained by a simple subtraction of corresponding pixels of the same bands at the two dates, i.e. $\mathbf{x_D}=\mathbf{x_2}-\mathbf{x_1}$ ) was transformed. Concerning the first option, let us denote with $\mathbf{s}_i^t$ the $i$-th component extracted through the transformation method $t$ (with $t=(\text{PCA, ICA, KICA})$). The number of extracted components corresponds to the total number of the spectral channels of the two multitemporal images. Considering the second strategy, let us denote with $\mathbf{s}_{D,i}^t$ the $i$-th component extracted with the three different methods. In this case the number of transformed components is equal to the number of spectral channels of the difference image. In order to extract changed areas from the components obtained with the transformation methods we applied two different procedures: i) identification of relevant single components and thresholding [6]; and ii) application of the Change Vector Analysis (CVA) technique in the polar domain combining pairs of components [7]. In the first case the most relevant single component $\mathbf{s}_i^t$ (or $\mathbf{s}_{D,i}^t$) was chosen according to a visual analysis by the user and then the threshold $T$ that discriminates $\omega_n$ from $\Omega_c$ was automatically retrieved by the expectation-maximization algorithms and the Bayes rule for minimum error [6]. Then the change-detection map was generated assigning all the pixels with value higher than the retrieved threshold to the class of change and the others to the class of no-changed.

In the second case, two relevant transformed components were chosen by the user and the spectral change vectors (SCVs) were computed according to a vector difference operator. The magnitude $\rho$ and the direction $\vartheta$ variables of the SCVs were then exploited in a polar framework for defining the change-detection map through the semi-automatic procedure proposed in [7]. In greater detail, the classes of changed and unchanged pixels can be separated by a threshold $T$ defined along the magnitude variable. In addition, different kinds of changes can be discriminated as SCVs related to them generate different clusters located in different direction ranges far from the origin. It follows that in order to generate the final change-detection map at first the threshold $T$ in the magnitude domain was chosen in an automatic way (as for the previous approach), then the user manually retrieved the couples of thresholds $\vartheta_{K_1}$ and $\vartheta_{K_2}$ in the direction domain for separating the $K$ kinds of change (please refer to [7] for greater details).

## 4. EXPERIMENTAL RESULTS

In order to assess the effectiveness of the proposed methods and to understand which transformation technique is the most suitable for change-detection applications, several experiments were carried out on both medium and VHR multispectral and multitemporal images. For space constraints in the following we report only the results obtained on VHR images. To this purpose, we consider a portion (733×537 pixels) of two Quickbird images acquired by the Quickbird sensor on the Trentino area (Italy) in October 2005 and July 2006. In the pre-processing phase the two images were: i) pan-sharpened by applying the Gram–Schmidt procedure; ii) radiometrically corrected; and iii) co-registered. The final data set was made up of two pan-sharpened multitemporal and multispectral images that have a residual misregistration of about 1 pixel on ground control points. Figure 2.a and 2.b show a real color composition of the pan-sharpened multispectral images $\mathbf{x_1}$ and $\mathbf{x_2}$, respectively. Between the two acquisition dates two kinds of changes occurred: i) new houses were built on rural area (white circles in Figure 2.b - $\omega_{c_1}$); and ii) some roofs in the industrial and urban area were rebuilt (black circles in Figure 2.b - $\omega_{c_2}$). In order to allow one a quantitative evaluation of the effectiveness of the proposed method, a reference map related to the two kinds of changes was defined.

(a) (b)

Figure 2: Real color composition of images of the Trento city (Italy) acquired by the Quickbird VHR multispectral sensor in: (a) October 2005; and (b) July 2006 (changes occurred between the two acquisition dates appear in black and white circles).

At first PCA, ICA, and KICA were separately applied to: i) the original images originating 8 components (as Quickbird images are made up of four spectral channels – blue, green, red and near infrared), and ii) the difference image generating 4 components. Concerning the Kernel ICA different algorithms (i.e. the Kernel Canonical Correlation Analysis KCCA, and the Kernel Generalized Variance KGV [5]) were tested with different parameter values (i.e. different kernels, different values of spread σ for Gaussian kernels, different orders for polynomial kernels, and different values for regularization parameter). However, on the considered data set, different combinations involved very similar results. In the following we report only the analysis conducted using KCCA algorithm with Gaussian kernel having σ=0.5 and regularization parameter set to 0.002.

All the components were visually analyzed in order to choose the most significant ones. Then the two change-detection strategies described in Section III were applied. In details, according to the first strategy, the selected components were thresholded and the change-detection map was derived. It is worth noting that this results in a single change-detection map for each kind of change (i.e. each component). Concerning the second strategy the CVA was applied to two significant components of PCA, ICA and KICA, and thresholds applied to both magnitude and direction of SCVs in order to generate the final change-detection map. This procedure allowed us to obtain a change-detection map in which the two different kinds of changes are reported and isolated from the noise components (which are contained in other ICA and KICA components). Figure 3 reports an example of the change-detection maps obtained with the ICA transformation. A quantitative analysis was performed by comparing the obtained change-detection maps with the reference maps. In addition, a comparison with the results obtained by applying the standard CVA technique to the original spectral channels was carried out.

**Table I. Change-detection results obtained by the proposed approaches in terms of kappa accuracy**

| *Transf. method* | $\omega_{c_1}$ | $\omega_{c_2}$ | $\omega_{c_1}$ *and* $\omega_{c_2}$ |
|---|---|---|---|
| **ICA** | 0.774 | 0.788 | 0.747 |
| **KICA** | 0.771 | 0.795 | 0.808 |
| **PCA** | 0.658 | 0.549 | 0.733 |
| **CVA standard** | - | - | 0.493 |

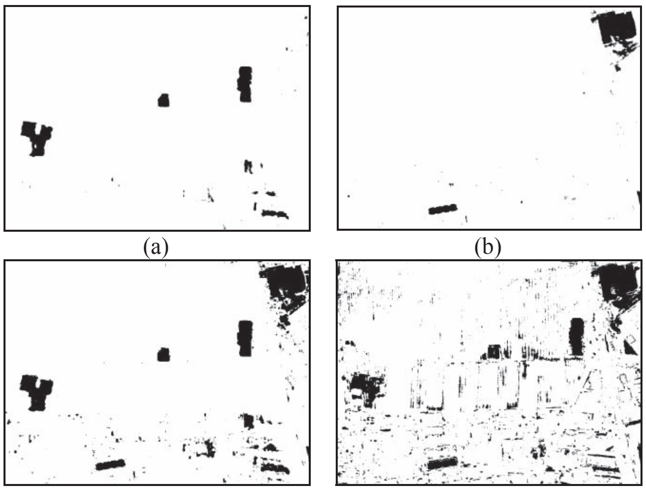


Figure 3: Change-detection maps obtained by the ICA components computed on the eight original spectral channels of the Quickbird images. (a) Thresholding of the 6th component; (b) Thresholding of the 1st component; and (c) thresholding of the magnitude of the SCVs obtained from the 1st and the 6th components. And (d) change-detection map obtained through the standard CVA on the original images.

The most significant results are reported in Table I. In particular, the table contains the results (in terms of kappa accuracy) obtained by thresholding the 6th component for ICA and the 4th component for KICA and PCA for extracting the change related to new houses (second column), and the 1st component (ICA), the 7th (KICA) and the 2nd (PCA) for identifying the change related to rebuilt roofs (third column). Furthermore, the accuracies obtained by thresholding the magnitude and direction of the SCVs obtained by both components for detecting both kinds of changes are reported (fourth column).

Comparing the numerical results yielded by ICA and KICA with the ones obtained with PCA and standard CVA, one can observe that ICA and KICA, independently from the change-detection strategy applied, involved change-detection maps with higher accuracies. This is mainly due to a considerable reduction of false alarms. With respect to the PCA, the first two techniques show better capabilities in separating the sources associated with true changes from sources of noise. Concerning the standard CVA algorithm applied to the original spectral channels, one can observe that false alarms due to different kinds of noise (i.e. registration noise, radiometric variations) cannot be eliminated in the generation of the final change-detection map (see Fig. 3.d), even if the spectral channels are accurately selected. It is worth nothing that the proposed techniques based on image transformation followed by simple change-detection algorithms allow one to generate different change-detection maps for different kinds of change (see Fig. 3.a and 3.b) according to the selected transformed components.

## 5. CONCLUSION

In this paper different data transformation techniques (ICA, KICA and PCA) have been exploited for change-detection purposes. The capability of each technique to separate the information sources associated with true changes (and to differentiate among different changes) from those associated with noise was investigated. Results obtained on different remote sensing images confirmed the effectiveness of the ICA and the KICA techniques in separating the different sources in the change-detection process. In particular, it is possible to conclude that the ICA resulted in the best tradeoff between complexity and change-detection accuracy, while the PCA obtained the poorest change detection results. Kernel ICA achieved results only slightly better that the ICA, but the computational complexity required from this approach is very high compared to the one required by ICA.

## 6. REFERENCES

[1] T. Fung and E. LeDrew, “Application of principal component analysis to Change detection,” *Photogrammetric Engineering and Remote Sensing*, Vol. 53, No. 12, pp. 1649-1658, Dec. 1987.

[2] F. Bovolo, L. Bruzzone and S. Marchesi, “Analysis and Adaptive Estimation of the Registration Noise Distribution in Multitemporal VHR Images”, *IEEE Transaction on Geoscience and Remote Sensing*, in press.

[3] J. Zhong and R. Wang, “Multi-temporal remote sensing change detection based on independent component analysis,” *International Journal of Remote Sensing*, Vol. 27, No. 10, pp. 2055-2061, May 2006.

[4] M. Ceccarelli and A. Petrosino, “Unsupervised Change Detection in Multispectral Images based on Independent Component Analysis,” *Proceedings of the 2006 International Workshop on Imagining Systems and Techniques*, pp. 54-59, April 2006.

[5] F.R. Bach and M.I. Jordan, “Kernel independent component analysis,” *The Journal of Machine Learning Research*, Vol. 3, pp. 1-48, March 2003.

[6] L. Bruzzone and D.F.Prieto, “Automatic analysis of the difference image for unsupervised change detection,” *IEEE Transaction on Geoscience and Remote Sensing*, Vol. 38, No. 3, pp. 1171-1182, May 2000.

[7] F. Bovolo and L. Bruzzone, “A theoretical framework for unsupervised change detection based on change vector analysis in polar domain,” *IEEE Transaction on Geoscience and Remote Sensing*, Vol. 45, No. 1, pp. 218-236, 2007.