to the elements. The algorithm uses an iterative refinement technique. When the pattern assigned to the corresponding clusters, the mean vector m_c is updated by averaging the coordinates of elements of the cluster. Then it provides a new gravity center. The process is iterated until the centers became consistence and algorithm reaches to a minimum of d^2 $(x_i;m_c)$. By considering the mapping function ϕ in kernel space, the similarity measure of K-means becomes as equation 4:

$$d^2(\emptyset(x_i, m_c) = \|\emptyset(x_i) - m_c\|^2, \quad m_c = \frac{1}{N_c}\sum_{je\,C_c}\emptyset(x_i)$$

Eqn 4

Then kernel K-means method first groups similar points in higher dimension and then separates them. By expansion of the equation 5 and replacing dot product by a proper kernel function, the kernel K-means formulation is obtained as:

$$\begin{aligned} d^2(\emptyset(x_i, m_c) = & \langle \emptyset(x_i), \emptyset(x_i)\rangle \\ & + \frac{1}{|N_c|^2}\sum_{j,m\in C_c}\langle \emptyset(x_j), \emptyset(x_m)\rangle \\ & - \frac{2}{|N_c|}\sum_{j,i\in C_c}\langle \emptyset(x_i), \emptyset(x_j)\rangle \\ & = K(x_i, x_i) + \frac{1}{|N_c|^2}\sum_{j,m\in C_c} K(x_i, x_m) \\ & - \frac{2}{|N_c|}\sum_{j,i\in C_c} K(x_i, x_m) \end{aligned}$$

Eqn 5

The main considerations of proposed kernel K-means Change Detection with deep learning algorithm are initialization and optimization of the cost function. It is obvious that one of the limitations of iterative techniques is the initialization issue. Poor initialization can converge in the local minima, and then the algorithm fails. In order to reach a stable and correct grouping, one possible solution is to initialize the algorithm with a training set which are based on the prior analysis.

3.3 Kernel parameter Estimation

RBF kernel related to Gaussian distribution and for this reason have adaptable results. On the other hand, the changes associated to differences between two images and difference kernel contains admissible information about changes. We used RBF kernels for the kernels generate difference kernel equation 2. Then an optimization problem which is the distance between samples and their cluster center, must be solved.

To obtain the difference kernels parameters, two single- and cross-time parameters should be estimated. The Rayleigh coefficient, in term of cluster-related distance in feature space, is presented in equation 6. This coefficient is calculated from the covariance or normalized scatter matrices of data.

$$\sum_i = \frac{\alpha}{M}\sqrt{\sum x_m x_m^T}$$

Eqn 6

$$\{\sigma^{single}, \sigma^{cross}\} = argmin\left\{\frac{\frac{i}{m_c}\sum_{i\in C} d^2(\emptyset^{diff}(x_i), m_c}{\sum_{c\notin q} d^2(m_c, m_q)}\right.$$

Eqn 7

It's worth noting that in order to minimize this expression; the kernel c-means technique is wrapped to test different sets of parameters to cluster the same training samples. For this purpose, we used linear search method to obtain the best parameters. After initialization and estimation of optimum kernel parameters, we can apply optimal kernel K-means algorithm on our data sets. Figure 2 presents the flowchart of main step for our approach which is used in this research.

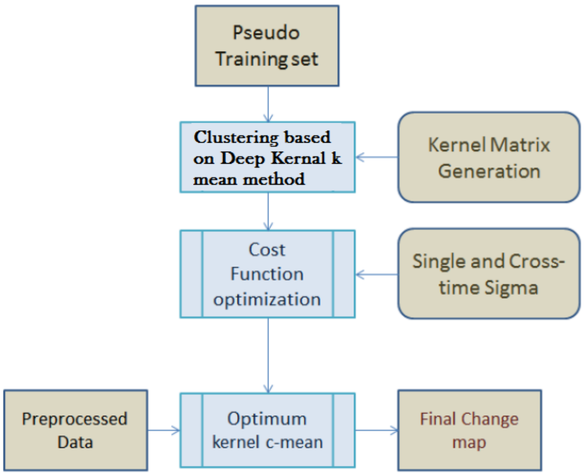

Figure 2: Overview for the main step of the proposed approach

4. Experimental Results

4.1. Datasets

Change happened in Dubai's Islands between 2000 and 2012 figure 4: These images are provided by Earth Resources Observation and Science (EROS) as part of U.S. Geological Survey (USGS). Each image is of size 1600 X 1600 pixels. The images were accured on 12 November 2000 and 27 November 2012. The city of Dubai is situated along the Persian Gulf in the United Arab Emirates. These Landsat 7 images show the area in 2000 and 2012, and give a remarkable view of the changes that have taken place, both on land and in the water. In 2001, work began to create artificial archipelagos along the shoreline of Dubai. The Palm Jebel Ali and smaller Palm Jumeirah are two Palm Islands that can be seen in the 2012 imagery. North from the Palm Islands is a group of smaller islands, created in the rough shape of a world map. Known as The World, this small area has created an additional 144 miles of shoreline.

Figure 3: Dubai-Islands- 2000 and 2012

4.2 Experimental Setup

In order to test the sensitivity to different sizes of the pseudo training set, many situations have been considered. After experimental evaluation, results only on a single set size are reported. A plateau effect in accuracy is observed for growing set sizes. The smallest set reaching this upper limit is considered for analyses. The sample is taken in such a way that the number of pixels is large enough to cover data variability but small enough to allow fast computations, regulated by the size of the kernel matrix. To have robust statistical estimates, 10 different realizations of the pseudo training set are considered followed by as many independent runs of the algorithm. Average and standard deviation of accuracies are reported for evaluation purposes.

The proposed approaches are tested versus the linear counterpart of the considered mappings (both resulting in standard k-means on the difference image) providing a baseline accuracy, and against two automatic change detection methods: the standard CVA [15] and the approach presented in [16]. The former puts a threshold in the magnitude distribution. The latter relies on a patch-based PCA transformation of the difference of intensities images and it clusters changes using binary standard k means. See [16] for details. Note that, for fair comparison, the free parameters of this method are chosen by minimizing the classification error on a labelled validation set. Moreover, since this approach is designed for intensity images (thus unidimensional), an investigation has been carried out in order to select the best input: among single band differences and the magnitude, the latter resulted in higher accuracies and has thus been retained.

For all the nonlinear cases, Gaussian RBF kernels are adopted. This choice is motivated by their interpretability, recast as a local distance (similarity) and by good performances that make this kernel function the most used in many fields. RBF bandwidths are optimized in the interval $\sigma \varepsilon$ [0:1; 10].

4.3 Result

Change happened in Dubai's islands: The changes detected in Dubai City after the successful execution of this system is shown in figure 4. The training set are composed by 1000 random pixels out of which 500 pixels each represented the change region and the no change distribution.

Figure 4: Changes detected in Dubai, UAE

By using nonlinear clustering and in particular by adopting the better representation provided by the Ker.Diff.RBF, small deviations to the no change class can be clustered as unchanged pixels, making the detection of large deviations more accurate. In this case, both the CVA and the method of [16] suffered from the non-normalizable differences that appear on the island, that make the standard difference image and the magnitude measure suffer from ambiguity in the representation. Therefore, methods relying solely on the magnitude are supposed to provide worse results than other approaches, as the latter fully exploit the information content of the data. By using deep learning in ker.diff.RBF method the performance was increased considerably. The steps in the execution were simplified by the layered approach.

Table II: Quantitative comparison of the results

Method/Accuracy	False Alarm	Misdetection	kappa	OA
K mean clustering	15.22	5.98	0.7633	88.45
Diff.Lin.	8.34	4.52	0.8741	93.22
Diff.RBF.	7.58	3.80	0.8824	94.52
Ker. Diff.RBF.	5.87	1.64	0.9250	96.13

In the table I and II, the false alarm rate and misdetection rate consequently show the detected changes which are not actual changes and the un-changed pixels which the algorithm considers them as changed pixels. These elements of confusion matrices and also total accuracy and kappa coefficient show improvement of the kernel-based algorithm in accurate detection of changes.

5. Conclusion

A kernel-based deep learning approach to unsupervised change detection has been presented. By exploiting a proper initialization, kernel k-means clustering is used to learn the representatives for the two classes of interest. Issues related to the estimation of the kernel parameters have been tackled by optimizing a geometrical criterion favoring dense groups and distant cluster centers showing a minimum when a convenient representation for clustering is found.

When estimating the similarity of the difference image in feature spaces performances are strongly improved with respect to approaches based on the difference image in the input space. This indicates that a better representation can be obtained by considering single and cross-time relationships among multitemporal pixels. The consequent strong decrease in false alarm rate and slight improvement in detection makes this approach the most accurate. Experimental comparisons showed that relying on the magnitude only the correct discrimination of the changes becomes a difficult task. This is related to the ambiguity of the measure, as well as to its unidimensional representation hiding useful information. On the contrary, by detecting changes in higher dimensional feature spaces, the multitemporal information is unfolded into clusters that are easily detectable. Such approach also reduces the preprocessing corrections, since the single time information is considered separately and regularized by the cross-similarity of the scenes.

Further research is required to investigate the inclusion of contextual and multiscale approaches, as well as multisource information. They can be included in the process by combining specific kernel functions. Type of changes and their direction (e.g. the spectral change vector angles) could also be considered to extend the proposed framework to multiclass change detection.

References

[1] A. Singh, Digital change detection techniques using remotely-sensed data, Int. J. Remote Sens., vol. 10(6), pp. 9891003, 1989.
[2] Michele Volpi, Devis Tuia, Gustavo Camps-Valls, and Mikhail Kanevski, Unsupervised Change Detection with Kernels
[3] L. Bruzzone and D. F. Prieto, Automatic analysis of the difference image for unsupervised change detection, IEEE Trans. Geosci. Remote Sens., vol. 38, no. 3, pp. 11711182, 2000.
[4] G. Camps-Valls and L. Bruzzone, Eds., Kernel Methods for Remote Sensing Data Analysis, J. Wiley & Sons, 2009.
[5] M. Volpi, D. Tuia, G. Camps-Valls, and M. Kanevski, Unsupervised change detection by kernel clustering, in SPIE, Im. Signal Proc. Remote Sensi., L. Bruzzone, Ed., 2010, vol.7830.
[6] J. Shawe-Taylor and N. Cristianini, Kernel Methods for Pattern Analysis, Cambridge University Press, 2004.
[7] F. Bovolo, G. Camps-Valls, and L. Bruzzone, A support vector domain method for change detection in multitemporal images, Pattern Recogn. Lett., vol. 31, no. 10, pp. 11481154,2010.
[8] Karl Ni, Ryan Prenger, Learning Features in Deep Architectures with Unsupervised Kernel k-Means
[9] SINGH, A., 1989, Digital change detection techniques using remotely-sensed data, International Journal of Remote Sensing, 10(6), pp.989-1003.
[10] DEER, P., 1999, Digital Change Detection Techniques: Civilian and Military Application Published in the UK (Taylar & Francis Ltd).
[11] LI, D., SUI, H. and XIAO, P., 2002, Automatic Change Detection of Geo-spatial Data from Imagery, International Archives of Photogrammetry and Remote Sensing, Xian, China, Vol.XXXIV, Part 2, pp.245-251.
[12] LU D., MAUSEL, P., BRONDIZIO, E. and MORAN,E., 2004, Change detection techniques, International Journal of Remote Sensing, 25(12), pp.2365-2407.
[13] RICHARD, J., RADKE, SRINIVAS, A., OMAR, A.K. and BADRINATH, R., 2005, Image Change Detection Algorithms: A Systematic Survey, IEEE Transactions on Image Processing, 14(3), pp.294-307.
[14] SUI, H.G., 2002, Automatic change detection for road-networks base on features, Ph.D dissertation, Wuhan University, China
[15] F. Bovolo and L. Bruzzone, A theoretical framework for unsupervised change detection based on change vector analysis in polar domain, IEEE Trans. Geosci. Remote Sens., vol. 45, no. 1, pp. 218236, 2006.
[16] T. Celik, Unsupervised change detection in satellite images using principal component analysis and k-means clustering, IEEE Geosci. Remote Sensi. Lett., vol. 6, no. 4, pp. 772776, 2009.