

## KERNEL-COMPOSITION FOR CHANGE DETECTION IN MEDIUM RESOLUTION REMOTE SENSING DATA

Andreas Ch. Braun, Uwe Weidner, Stefan Hinz

Institute of Photogrammetry and Remote Sensing  
KIT - Karlsruhe Institute of Technology  
Englerstr. 7, 76131 Karlsruhe  
andreas.ch.braun@kit.edu, uwe.weidner@kit.edu, stefan.hinz@kit.edu  
<http://www.ipf.kit.edu>

Commission VII/5

**KEY WORDS:** Change Detection, Classification, Landuse, multispectral, multitemporal, Landsat

### ABSTRACT:

A framework for multitemporal change detection based on kernel-composition is applied to a multispectral-multitemporal classification scenario, evaluated and compared to traditional change detection approaches. The framework makes use of the fact that images of different points in time can be used as input data sources for kernel-composition - a data fusion approach typically used with kernel based classifiers like support vector machines (SVM). The framework is used to analyze the growth of a limestone pit in the Upper Rhine Graben (West Germany). Results indicate that the highest accuracy rates are produced by the kernel based framework. The approach produces the least number of false positives and gives the most convincing overall impression.

### 1 INTRODUCTION

Although the availability of modern remote sensing datasets increases – e.g. hyperspectral, high resolution optical satellite imagery, interferometric synthetic aperture radar – many change detection applications continue to require traditional datasets. The fact that e.g. Landsat data are available since 1972 makes them a valuable source of information for four entire decades. They can be seen as a way to recover information on past environmental conditions which are not observable any more by field campaigns. However, traditional methods like post-classification change detection based on overlaying classification maps raise accuracy issues (Serra et al., 2003). Therefore, more sophisticated change detection methods have been proposed in literature. Some approaches model the probability of transition from one class to another. Another approach is change detection based on kernel-composition (Camps-Valls et al., 2006b). These issues are exemplified on a change detection application from Upper Rhine Graben. Changes in the landuse are outlined with special focus on the construction of a limestone pit which continuously grows replacing near-natural ecosystems. Kernel based classifiers – like the well know support vector machine (SVM) (Boser et al., 1992), (Cortes and Vapnik, 1995) – work on kernel matrices. These kernel matrices represent the similarity between data points in high dimensional feature spaces (reproducing kernel Hilbert spaces, RKHS). The SVM chooses the most suitable points by optimizing a target function on the kernel choosing only a few training data points as SVs. These points are used to define a separating hyperplane which is usually non-linear in the input space. The conceptual advantage of kernel-composition is, that different kernel functions can be combined e.g. by addition, thus performing data fusion in the RKHS during classification. This circumstance is usually employed for data fusion (Tuia and Camps-Valls, 2009). However, it can also be employed for change detection. In (Camps-Valls et al., 2008), (Camps-Valls et al., 2006a) well know change detection techniques – like image differencing – are adapted for kernel-composition based approaches. An entire framework for change detection and multitemporal classification is presented. This framework is evalu-



Figure 1: Limestone pit near Heidelberg

ated and compared against traditional approaches herein. We exemplify this framework analyzing the growth of a limestone pit near the city of Heidelberg between 2001 and 2005. The kernel based approaches are compared against traditional approaches like post-classification change detection and an approach based on stacking features before classification. Although traditional approaches provide valuable information on the overall growth of the limestone pit, kernel based approaches provide more robust results which are visually more convincing and quantitatively more accurate. The remainder of the paper is organized as follows. Sec.2 provides the mathematical foundations of the methods used. Sec.3 gives an overview of relevant literature in change detection and kernel-composition methods. The results of the comparison are presented in Sec.4 and discussed in Sec.5. Finally, Sec.6 concludes the contribution.

## 2 MATHEMATICAL FOUNDATIONS

Within this section, the mathematical foundations of the main methods used will be given. A high number of profound introductions into the SVM problem (Burges, 1998), (Ivanciuc, 2007), (Zhang, 2001), (Schölkopf and Smola, 2002), (Camps-Valls and Bruzzone, 2009) and valuable reviews on the application of SVM in remote sensing (Mountrakis et al., 2010), (Plaza et al., 2009) have been published. For this reason, the foundations of SVM and state-of-the-art application examples will not be given exhaustively, but strictly focused on the kernel-composition problem.

### 2.1 Kernel matrices and the SVM problem

Given a data set  $X$  with  $n$  data points, kernel matrices are the result of kernel functions applied over all  $n^2$  tuples of data (Shawe-Taylor and Cristianini, 2004). The outcome of a kernel function  $K_{x_i, x_j} = f_\delta(x_i, x_j)$  is a similarity measure for the two training data  $x_i$  and  $x_j$  depending on some distance metric  $\delta$ . Usually,  $\delta$  is the Euclidean distance (Mercier and Lennon, 2003). However, kernel functions can be modified by e.g. by introducing different similarity measures (Amari and Wu, 1999). For instance, (Mercier and Lennon, 2003) and (Honeine and Richard, 2010) use the spectral angle as a similarity measure for hyperspectral SVM classification. To model complex distributions of the training data in the feature space,  $f_\delta$  is usually some non-linear function. The most frequently applied family of non-linear functions are Gaussian radial basis functions (RBF) (Schölkopf et al., 1997). The closer two points are found in the feature space, the higher is their resulting kernel value. Given these facts, the kernel matrix simply represents the similarity between the points of the training data set. To understand how the kernel matrix is used in SVM classification, it is helpful not to look at the primal, but the dual formulation of the SVM problem (Ivanciuc, 2007). The dual problem is given by Eq.1

$$\begin{aligned} \text{maximize : } & \sum_i \lambda_i - \frac{1}{2} \sum_i \sum_j \lambda_i \lambda_j y_i y_j K(x_i, x_j) \quad (1) \\ & \lambda_i > 0 \quad \forall \text{ Support Vectors} \\ & \lambda_i = 0 \quad \forall \text{ other points} \\ & y_i \in [-1, +1] \end{aligned}$$

The Lagrange multipliers  $\lambda_i$  are only greater than zero for the support vectors. These are usually identified by sequential minimal optimization (Platt, 1998). Hence, only training data which are both SVs contribute to the solution of Eq.1 (for all other cases,  $\lambda_i \lambda_j = 0$  setting the second part of Eq.1 to zero). The class labels  $y_i$  are in  $[-1, 1]$ . Since the second part of Eq.1 is subtracted, only points with different class labels can maximize the term (their product  $y_i y_j = -1$  renders the second part positive). The problem is therefore maximized, if points are chosen as SVs which have different class labels but are found close to each other in the feature space (thus yielding a high value in the kernel matrix  $K(x_i, x_j)$ ). Thus, the similarity values of the kernel matrix are used for finding the best suited training points as SVs. By setting the  $\lambda_i$  of all other points to zero, a sparse solution is found which only depends on the SVs.

### 2.2 Kernel-Composition

As can be seen in Eq.1, the training data  $x_i$  do not enter directly into the SVM problem. In contrast, the data are represented by kernels  $K(x_i, x_j)$ . According to Mercer's theorem (Mercer, 1909), valid kernel functions can be combined, e.g. through addition, to form new valid kernels. From there, different sources

of information on the same training data can be fused through simple arithmetical operations (Camps-Valls et al., 2006b). For instance,  $K_C(x_i^C, x_j^C) = K_A(x_i^A, x_j^A) + K_B(x_i^B, x_j^B)$  fuses the information domains  $A$  and  $B$  on the training data  $x_i, x_j$  and forms a new kernel  $K_C$ . Within the original framework on kernel-composition for data fusion (Camps-Valls et al., 2006b), the following fusion approaches are published.

$$K_C(x_i^C, x_j^C) = K_A(x_i^A, x_j^A) + K_B(x_i^B, x_j^B) \quad (2)$$

$$K_C(x_i^C, x_j^C) = \mu K_A(x_i^A, x_j^A) + (1 - \mu) K_B(x_i^B, x_j^B) \quad (3)$$

$$K_C(x_i^C, x_j^C) = K_A + K_B + K_{AB}(x_i^A, x_j^B) + K_{BA}(x_i^B, x_j^A) \quad (4)$$

Eq.2 is called direct summation kernel, the most simple form of kernel-composition. Eq.3 is called weighted summation kernel. Its main advantage is, that the weighting parameter  $\mu \in (0, 1)$  allows to regulate the relevance of the two data sources  $A$  and  $B$  for the classification problem. Eq.4 is called the cross-information kernel. It consist of four single kernels while the last two  $K_{AB}$  and  $K_{BA}$  allow incorporating the mutual information between the data sources  $A$  and  $B$  (e.g. differences between the value of both data sources yielded for a particular data point). Based on these basic composition approaches (Camps-Valls et al., 2008), (Camps-Valls et al., 2006a) extend the kernel-composition framework to the field of multitemporal classification and change detection. The key idea is to use images from the same landscape but from different points in time as input data for kernel-composition and SVM classification. Given two points in time  $t_0$  and  $t_1$  two kernels  $K_{t_0}$  and  $K_{t_1}$  are built. These kernels only incorporate the spectral information given at each point of time. Then, a new kernel can be build using one of the Eqs.2 to 4. For instance,  $K_{Change}(x_i^C, x_j^C) = K_{t_0}(x_i^{t_0}, x_j^{t_0}) + K_{t_1}(x_i^{t_1}, x_j^{t_1})$  represents a direct summation kernel which incorporates the information about the change of the spectral responses of pixels implicitly. Although the basic composite kernels can be used for multitemporal classification as well, the authors developed specialized kernels in order to combine traditional change detection techniques with kernel-composition. For instance, the image difference kernel is introduced in Eq.5.

$$\begin{aligned} K_C(x_i^C, x_j^C) = & K_{t_1}(x_i^{t_1}, x_j^{t_1}) + K_{t_0}(x_i^{t_0}, x_j^{t_0}) - \dots \\ & \dots K_{t_1, t_0}(x_i^{t_1}, x_j^{t_0}) - K_{t_0, t_1}(x_i^{t_0}, x_j^{t_1}) \quad (5) \end{aligned}$$

Note that Eq.5 is a particular case of the cross-information kernel (Eq.4) that performs the change detection technique of image differencing in the RKHS.

## 3 RELATED WORK

Within this section, an overview on relevant contribution from the field of change detection and kernel-composition will be given. Since kernel-composition has been introduced only in 2006, it has been dedicated far less research than change detection in general.

### 3.1 Change detection and multitemporal classification

Herein, a short outline on important reviews and state-of-the-art papers in change detection is presented. A very comprehensive introduction into multitemporal classification is given by (Gillanders et al., 2008). (Singh, 1989) and (Coppin et al., 2004) present reviews with emphasis on signal processing. (Wang and Xu, 2010) give a comparison on change detection methods emphasising particular aspects of different applications. (Holmgren and Thuresson, 1998) and (Wulder et al., 2006) present reviews

focused on applications of change detection of satellite images in forestry. A comparable contribution for landscape monitoring is given by (Kennedy et al., 2009). (Almutairi and Warner, 2010) present important considerations on accuracy assessment and the influence of accuracy to the final change detection result. (Van Oort, 2007) presents an insightful contribution on the importance of the error matrix of multitemporal classification. Some state-of-the-art papers are given e.g. by (Bruzzone and Serpico, 1997) and (Bruzzone et al., 2004) which present an iterative approach for change detection. (Coops et al., 2010) apply Landsat time series for assessing forest fragmentation. (Dianat and Kasaei, 2010) use a polynomial regression technique for change detection which considers neighborhoods. Application schemes based on SVMs are presented by (Nemmour and Chibani, 2006), (He and Laptev, 2009) and (Bovolo et al., 2008). (Mota et al., 2007) and (Feitosa et al., 2009) present fuzzy approaches based on modeling the class transitional probabilities.

### 3.2 Kernel-composition

Kernel-composition as a method of data fusion has originally been proposed by (Camps-Valls et al., 2006b). An example for multispectral data fusion is given by (Camps-Valls and Bruzzone, 2009). Kernel-composition is widely applied for spectral-spatial classification fusing hyperspectral data with wavelets (Tan and Du, 2011), (Velasco-Forero and Manian, 2009), spatial correlation data (Chen et al., 2011), morphological profiles (Tan and Du, 2010), (Tuia and Camps-Valls, 2009), empirical mode decomposition (Demir and Erturk, 2010), median filters (Marconcini et al., 2009) or self-complementary filters (Fauvel et al., 2008). An approach fusing hyperspectral and laserscanning data is given by (Braun et al., 2011). Another application domain is semi-supervised learning, where kernel-composition is used to incorporate the spectral information of unlabeled pixels (Tuia and Camps-Valls, 2011), (Tuia and Camps-Valls, 2009), (Marconcini et al., 2009), (Camps-Valls et al., 2007), (Bandos et al., 2006). (Camps-Valls et al., 2008), (Camps-Valls et al., 2006a) provide the framework for multitemporal classification used herein. The framework has so far been evaluated e.g. by (Bovolo et al., 2010) in a support vector domain description approach on Landsat images and by (Volpi et al., 2011) in an unsupervised approach on multitemporal VHR data. To the knowledge of the authors, an evaluation on a multitemporal scenario with medium resolution data based on supervised classification has not been published so far.

## 4 RESULTS

Within this section, results for three change detection approaches will be presented. The main objective is to monitor the growth of a limestone pit close to the village Mauer (near Heidelberg) in the Upper Rhine Graben, Germany, Latitude: 49° 19' 50" N, Longitude: 8° 48' 1" E (cf. Fig. 1). Two 122 × 135 pixel (≈14.8km<sup>2</sup>) subsets of Landsat ETM+ images are used. The first is from 02-11-2001 (Fig.2(a)), the second is from 17-4-2005 (Fig.2(b)). Fortunately, the village is located in the very center of both images, so the failure of the Landsat ETM+ scan line corrector does not affect the work at all. Between the two points in time, the limestone pit has considerably grown. Although more landuse classes are assigned for classification in the first place (e.g. meadows, forests, settlements), only two classes of interest will be considered in the final result: *LS-Pit present in 2001* (yellow) and *LS-Pit new between 2001 and 2005* (red). Complete groundtruth has been made available by a digitization in the field and is shown in Fig.2(c). Except the limestone pit, all other landuse classes will not be considered and set to black. At first, change detection based on a post-classification approach will be employed.

Secondly, features will be stacked to represent the change of pixels intensities between the two points in time. Lastly, kernel-composition approaches based on (Camps-Valls et al., 2008), (Camps-Valls et al., 2006a) and (Camps-Valls et al., 2006b) will be employed. These approaches also incorporate changes in pixels intensities. All change detection approaches are based on image classification. Each classification was done using an SVM with a Gaussian RBF kernel<sup>1</sup>. Kernel parameters were tuned using a 5-fold grid search in the ranges  $\gamma \in [2^{-15}, 2^5]$  and  $C \in [2^{-5}, 2^{15}]$ . The LibSVM 3.11 library was utilized (Chang et al., 2001).

### 4.1 Post-classification approach

Two classifications are performed using the same landuse classes in each dataset. The limestone pit is represented by a single class. Since classification of the two points in time is performed separately, it is not possible to assign the class *LS-Pit new between 2001 and 2005* in the 2005 dataset. No features are available which indicate whether or not a pixel has belonged to the limestone pit in 2001 when classifying the 2005 image. From there, one class *LS-Pit* has been assigned in both datasets and the two classes of interest have been determined by overlaying the results afterwards. The overall accuracy on the two classes is 86.7%. A visual result after setting other classes (like meadows, forest, settlements) to black is given in Fig.2(d). Note the high amount of pixels falsely assigned to the class *LS-Pit new between 2001 and 2005*. Since the entire image scene is made up of limestone, many places of bare soils have been confused with the limestone pit.

### 4.2 Stacked-features approach

In order to provide implicit information on the changes of pixels intensities, a stacked-features approach was performed. The data matrices of both 8 channel Landsat datasets were concatenated to build a 16 channel data matrix. Within this feature space, the two classes *LS-Pit present in 2001* and *LS-Pit new between 2001 and 2005* can be distinguished in a single classification step. While the class *LS-Pit present in 2001* shows grey color in both images, *LS-Pit new between 2001 and 2005* would change from e.g. green to grey. This indicates that a change from e.g. meadows to limestone pit has taken place between the two points in time. From there, only one classification needs to be performed. The overall accuracy on the two classes is 87.5%. A visual result after setting other classes to black is given in Fig.2(e). Note that the amount of false positive pixels is considerably reduced. False positives are now assigned to the class *LS-Pit present in 2001*. Much less pixels are found in *LS-Pit new between 2001 and 2005*. Since the latter class is characterized by a change in color from green to grey, it is less confused with bare soils which would stay grey in both points in time.

### 4.3 Kernel-composition approach

The last approach performed is similar to the stacked-features approach. In order to incorporate information on e.g. color changes, both datasets are combined to a new dataset. However, it is aimed to perform the fusion not in the feature space, but in the RKHS. Therefore, a composed kernel matrix was build, e.g. by  $K_{change}(x_i^C, x_j^C) = K_{2001}(x_i^{01}, x_j^{01}) + K_{2005}(x_i^{05}, x_j^{05})$ . The following kernel-composition approaches were followed: direct summation (Eq.2), weighted summation (Eq.3), cross-information (Eq.4) and image differencing (Eq.5). The overall accuracy values for each approach can be seen in Tab.1. The best overall accuracy yielded

<sup>1</sup> $K(x_i, x_j) = \exp(-\gamma(\|x_i - x_j\|)^2)$ , implemented in LibSVM 3.11

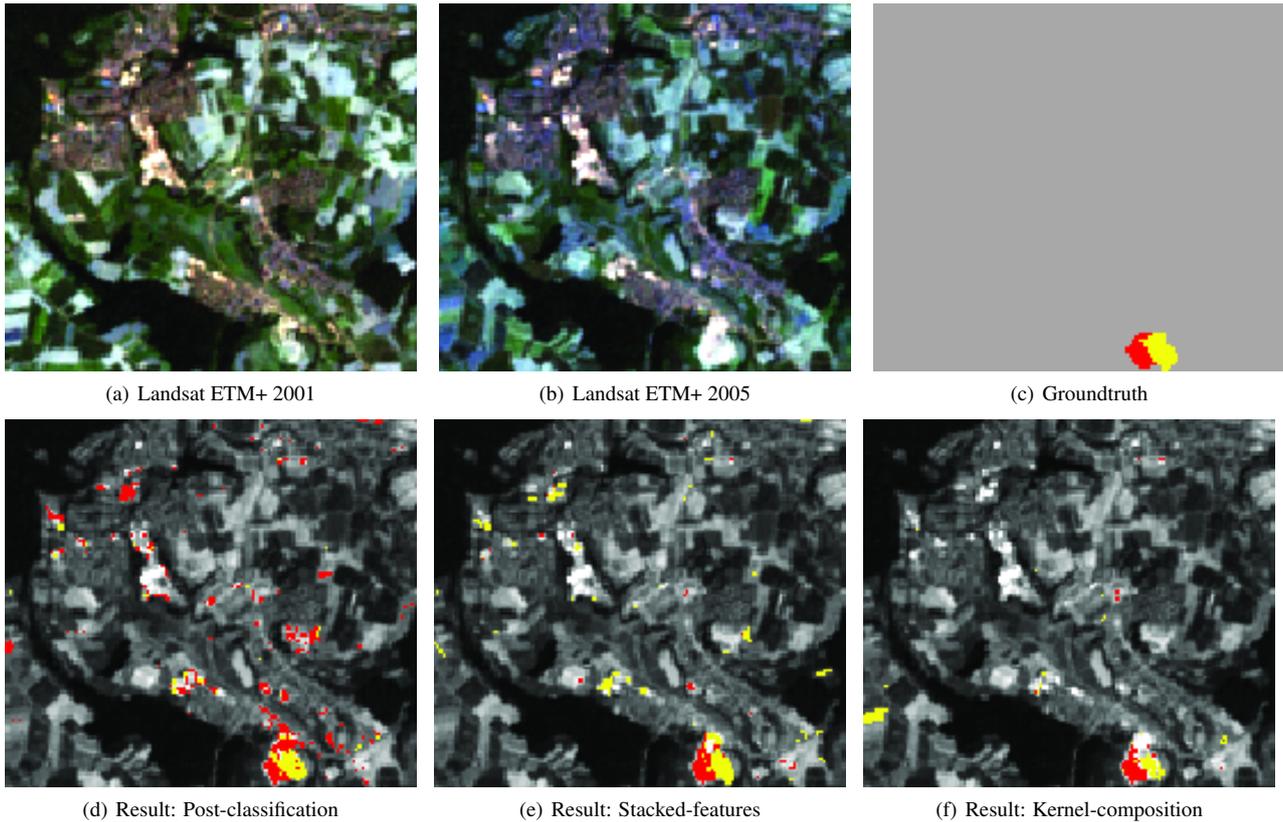


Figure 2: Landsat ETM+ image scenes, groundtruth and results

Table 1: Approaches and overall accuracies

Approach	Composition	Overall accuracy
Post-classification	–	86.7%
Stacked-features	–	87.5%
<b>Kernel-Composition</b>	<b>Direct. Sum.</b>	<b>88.8%</b>
Kernel-Composition	Weighted. Sum.	88.1%
Kernel-Composition	Cross-Info.	87.9%
Kernel-Composition	Image Diff.	88.6%

by the direct summation approach on the two classes is 88.8%. As can be seen, all kernel-composition approaches yield slightly higher accuracy values than the other approaches. Although specifically designed for this task, the image difference kernel does not yield the highest accuracy value. However, the performance difference to the direct summation kernel is only 0.2 percent points – a value which should not be over-interpreted. It should be noted that more simple kernels yield better results than more complex ones. A result which is in agreement with the findings of the inventors of the framework (Camps-Valls et al., 2006b), (Camps-Valls et al., 2008), (Tuia and Camps-Valls, 2009). A visual result after setting other classes to black is given in Fig.2(f). There are much less false positives than in the other approaches. The only exception is a large barren field where open Loess soil mixed with limestone rocks is found (south-east corner of the image). The spectral characteristics of this field are very similar to the limestone pit thus making the classifier susceptible for confusion with the limestone pit.

#### 4.4 McNemar’s Test

The advantages in overall accuracy of e.g. kernel-composition over post-classification may seem only a slight gain. Therefore,

they were tested for significance using McNemar’s test (Foody, 2004). McNemar’s test is based on  $\chi^2$  statistics and can be employed to test the significance of differences between two nominal labellings. The advantage is considered as significant if the resulting test value  $|z| \geq 1.96$ . Testing the advantage of kernel-composition over post-classification yielded  $|z| \approx 13.95$  indicating a significant advantage. A test of kernel-composition against the stacked-features approach yielded  $|z| \approx 6.50$  which also is significant. The stacked-features approach yielded  $|z| \approx 10.16$  over post-classification. Thus, all advantages described are significant.

## 5 DISCUSSION

We present a comparison of change detection based on kernel-composition with two traditional methods, post-classification change detection and a stacked-features approach. As Fig.2(d) reveals, post-classification change detection produces many false alerts. In fact, the limestone pit is only located in the very south east. The overall accuracy yielded is 86.7%. Stacking features before classification (i.e. combining the two 8 channel datasets to a new 16 channel dataset) and assigning two limestone pit classes (*LS-Pit present in 2001* and *LS-Pit new between 2001 and 2005*) produces better results. However, there are still too many false alerts Fig.2(e). The overall accuracy yielded is 87.5%. Approaches based on kernel-composition produce the most accurate results. However, there are slight differences when using kernel-composition that depend on the type of composition. In agreement with other authors, more simple composition approaches tend to produce better results than more complex ones. In our case, the direct summation and the image difference kernel produced the best results. The highest overall accuracy yielded is 88.8% by a direct summation kernel. There are almost no false alerts and the change between 2001 and 2005 becomes most

clearly visible. The advantages between the approaches may seem only moderate. However, in the major part of the pit, a separation between the two limestone pit classes seem to be quite simple and therefore, all approaches yield good results in a large part. Furthermore, it has to be kept in mind that a major source of error comes from confusion between the limestone pit and natural outcrops of limestone or open chalky soils. Confusion between these landcover types narrow the differences between the change detection approaches. It should be noted though, that false positives based on this source of error are much less for kernel-composition approaches and more concentrated to single spots. According to McNemar's test, the advantage in overall accuracy of the kernel-composition approach over the other approaches are significant. The reason for the advantage of the kernel-composition approach and the stacked-features approach over post-classification change detection is straightforward. While post-classification change detection does not include any information on the change in pixels intensities between the two points of time, both kernel-composition and stacked-features do incorporate this information implicitly. However, the advantage of kernel-composition over the stacked-features approach is remarkable. Both approaches include information on the change in pixels intensities. However, kernel-composition appears to be a better suited technique to exploit this information. It is assumed that the main advantage lies in the fact, that the kernel-composition represents this information in the RKHS, while the stacked-features approach represents it in the original feature space. Since SVMs operate in the RKHS when finding their optimal solution, kernel-composition and SVM seem to be a more suitable combination for representing this implicit information.

## 6 CONCLUSIONS

Kernel based change detection is a conceptually elegant and useful method for change detection and multi temporal classification. Standard techniques like image differencing can be executed in RKHS, thus benefiting from the advantages of kernel based SVM classification. Changes in landuse for the given dataset from Upper Rhine Graben can be visualized and furthermore quantified with high precision. In future work, the approach will be tested on more complex change detection problems.

## REFERENCES

- Almutairi, A. and Warner, T., 2010. Change detection accuracy and image properties: A study using simulated data. *Remote Sensing* 2(6), pp. 1508–1529.
- Amari, S. and Wu, S., 1999. Improving support vector machine classifiers by modifying kernel functions. *Neural Networks* 12(6), pp. 783–789.
- Bandos, T., Zhou, D. and Camps-Valls, G., 2006. Semi-supervised hyperspectral image classification with graphs. In: *Proceedings of the IEEE Int. Geoscience and Remote Sensing Symposium, 2006. IGARSS'06*.
- Boser, B., Guyon, I. and Vapnik, V., 1992. A training algorithm for optimal margin classifiers. In: *Proceedings of the fifth annual workshop on Computational learning theory, ACM*, pp. 144–152.
- Bovolo, F., Bruzzone, L. and Marconcini, M., 2008. A novel approach to unsupervised change detection based on a semisupervised svm and a similarity measure. *IEEE Transactions on Geoscience and Remote Sensing* 46(7), pp. 2070–2082.
- Bovolo, F., Camps-Valls, G. and Bruzzone, L., 2010. A support vector domain method for change detection in multitemporal images. *Pattern Recognition Letters* 31(10), pp. 1148–1154.
- Braun, A., Weidner, U., Jutzi, B. and Hinz, S., 2011. Integrating external knowledge into svm classification - fusing hyperspectral and laserscanning data by kernel composition. Vol. Heipke C, Jacobsen K, Rottensteiner F, Miller S, S02rgel U (Eds) *High-resolution earth imaging for geospatial information. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 38 (Part 4 / W19) (on CD).
- Bruzzone, L. and Serpico, S., 1997. An iterative technique for the detection of land-cover transitions in multitemporal remote-sensing images. *IEEE Transactions on Geoscience and Remote Sensing* 35(4), pp. 858–867.
- Bruzzone, L., Cossu, R. and Vernazza, G., 2004. Detection of land-cover transitions by combining multivariate classifiers. *Pattern Recognition Letters* 25(13), pp. 1491–1500.
- Burges, C., 1998. A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery* 2(2), pp. 121–167.
- Camps-Valls, G. and Bruzzone, L., 2009. *Kernel methods for remote sensing data analysis*. Wiley Online Library.
- Camps-Valls, G., Bandos Marsheva, T. and Zhou, D., 2007. Semi-supervised graph-based hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing* 45(10), pp. 3044–3054.
- Camps-Valls, G., Gomez-Chova, L., Muñoz-Mari, J., Alonso, L., Calpe-Maravilla, J. and Moreno, J., 2006a. Multitemporal image classification and change detection with kernels. In: *SPIE International Symposium Remote Sensing XII, Vol. 6365*, p. 63650H.
- Camps-Valls, G., Gómez-Chova, L., Muñoz-Marí, J., Rojo-Álvarez, J. and Martínez-Ramón, M., 2008. Kernel-based framework for multitemporal and multisource remote sensing data classification and change detection. *IEEE Transactions on Geoscience and Remote Sensing* 46(6), pp. 1822–1835.
- Camps-Valls, G., Gomez-Chova, L., Muñoz-Marí, J., Vila-Francés, J. and Calpe-Maravilla, J., 2006b. Composite kernels for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters* 3(1), pp. 93–97.
- Chang, C., Lin, C. et al., 2001. Libsvm: a library for support vector machines.
- Chen, Y., Nasrabadi, N. and Tran, T., 2011. Sparse representation for target detection in hyperspectral imagery. *IEEE Journal of Selected Topics in Signal Processing* 5(3), pp. 629–640.
- Coops, N., Gillanders, S., Wulder, M., Gergel, S., Nelson, T. and Goodwin, N., 2010. Assessing changes in forest fragmentation following infestation using time series landsat imagery. *Forest Ecology and Management* 259(12), pp. 2355–2365.
- Coppin, P., Jonckheere, I., Nackaerts, K., Muys, B. and Lambin, E., 2004. Review article digital change detection methods in ecosystem monitoring: a review. *International journal of remote sensing* 25(9), pp. 1565–1596.
- Cortes, C. and Vapnik, V., 1995. Support-vector networks. *Machine learning* 20(3), pp. 273–297.
- Demir, B. and Erturk, S., 2010. Empirical mode decomposition of hyperspectral images for support vector machine classification. *IEEE Transactions on Geoscience and Remote Sensing* 48(11), pp. 4071–4084.
- Dianat, R. and Kasaei, S., 2010. Change detection in optical remote sensing images using difference-based methods and spatial information. *IEEE Geoscience and Remote Sensing Letters* 7(1), pp. 215–219.

- Fauvel, M., Chanussot, J. and Benediktsson, J., 2008. Adaptive pixel neighborhood definition for the classification of hyperspectral images with support vector machines and composite kernel. In: 15th IEEE International Conference on Image Processing, ICIP 2008., IEEE, pp. 1884–1887.
- Feitosa, R., Costa, G., Mota, G., Pakzad, K. and Costa, M., 2009. Cascade multitemporal classification based on fuzzy markov chains. ISPRS Journal of Photogrammetry and Remote Sensing 64(2), pp. 159–170.
- Foody, G., 2004. Thematic map comparison: evaluating the statistical significance of differences in classification accuracy. Photogrammetric Engineering and Remote Sensing 70(5), pp. 627–634.
- Gillanders, S., Coops, N., Wulder, M., Gergel, S. and Nelson, T., 2008. Multitemporal remote sensing of landscape dynamics and pattern change: describing natural and anthropogenic trends. Progress in Physical Geography 32(5), pp. 503–528.
- He, L. and Laptev, I., 2009. Robust change detection in dense urban areas via svm classifier. In: Joint Urban Remote Sensing Event, Shanghai.
- Holmgren, P. and Thuresson, T., 1998. Satellite remote sensing for forestry planning—a review. Scandinavian Journal of Forest Research 13(1-4), pp. 90–110.
- Honeine, P. and Richard, C., 2010. The angular kernel in machine learning for hyperspectral data classification. In: 2nd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), 2010, IEEE, pp. 1–4.
- Ivanciuc, O., 2007. Applications of support vector machines in chemistry. Reviews in computational chemistry pp. 291–400.
- Kennedy, R., Townsend, P., Gross, J., Cohen, W., Bolstad, P., Wang, Y. and Adams, P., 2009. Remote sensing change detection tools for natural resource managers: Understanding concepts and tradeoffs in the design of landscape monitoring projects. Remote Sensing of Environment 113(7), pp. 1382–1396.
- Marconcini, M., Camps-Valls, G. and Bruzzone, L., 2009. A composite semisupervised svm for classification of hyperspectral images. IEEE Geoscience and Remote Sensing Letters 6(2), pp. 234–238.
- Mercer, J., 1909. Functions of positive and negative type, and their connection with the theory of integral equations. Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character 209, pp. 415–446.
- Mercier, G. and Lennon, M., 2003. Support vector machines for hyperspectral image classification with spectral-based kernels. In: Proceedings of the IEEE Int. Geoscience and Remote Sensing Symposium, 2003. IGARSS'03, Vol. 1, IEEE, pp. 288–290.
- Mota, G., Feitosa, R., Coutinho, H., Liedtke, C., Muller, S., Pakzad, K. and Meirelles, M., 2007. Multitemporal fuzzy classification model based on class transition possibilities. ISPRS journal of photogrammetry and remote sensing 62(3), pp. 186–200.
- Mountrakis, G., Im, J. and Ogole, C., 2010. Support vector machines in remote sensing: A review. ISPRS Journal of Photogrammetry and Remote Sensing.
- Nemmour, H. and Chibani, Y., 2006. Multiple support vector machines for land cover change detection: An application for mapping urban extensions. ISPRS journal of photogrammetry and remote sensing 61(2), pp. 125–133.
- Platt, J., 1998. Sequential minimal optimization: A fast algorithm for training support vector machines. Advances in Kernel Methods Support Vector Learning 208(MSR-TR-98-14), pp. 1–21.
- Plaza, A., Benediktsson, J., Boardman, J., Brazile, J., Bruzzone, L., Camps-Valls, G., Chanussot, J., Fauvel, M., Gamba, P., Gualtieri, A. et al., 2009. Recent advances in techniques for hyperspectral image processing. Remote Sensing of Environment 113, pp. S110–S122.
- Schölkopf, B. and Smola, A., 2002. Learning with kernels: Support vector machines, regularization, optimization, and beyond. The MIT Press.
- Schölkopf, B., Sung, K., Burges, C., Girosi, F., Niyogi, P., Poggio, T. and Vapnik, V., 1997. Comparing support vector machines with gaussian kernels to radial basis function classifiers. IEEE Transactions on Signal Processing 45(11), pp. 2758–2765.
- Serra, P., Pons, X. and Saurí, D., 2003. Post-classification change detection with data from different sensors: some accuracy considerations. International Journal of Remote Sensing 24(16), pp. 3311–3340.
- Shawe-Taylor, J. and Cristianini, N., 2004. Kernel methods for pattern analysis. Cambridge Univ Pr.
- Singh, A., 1989. Review article digital change detection techniques using remotely-sensed data. International Journal of Remote Sensing 10(6), pp. 989–1003.
- Tan, K. and Du, P., 2010. Classification of hyperspectral image based on morphological profiles and multi-kernel svm. In: 2nd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), 2010, IEEE, pp. 1–4.
- Tan, K. and Du, P., 2011. Combined multi-kernel support vector machine and wavelet analysis for hyperspectral remote sensing image classification. Chinese Optics Letters 9(1), pp. 011003.
- Tuia, D. and Camps-Valls, G., 2009. Semisupervised remote sensing image classification with cluster kernels. IEEE Geoscience and Remote Sensing Letters 6(2), pp. 224–228.
- Tuia, D. and Camps-Valls, G., 2011. Urban image classification with semisupervised multiscale cluster kernels. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 4(1), pp. 65–74.
- Van Oort, P., 2007. Interpreting the change detection error matrix. Remote sensing of environment 108(1), pp. 1–8.
- Velasco-Forero, S. and Manian, V., 2009. Improving hyperspectral image classification using spatial preprocessing. IEEE Geoscience and Remote Sensing Letters 6(2), pp. 297–301.
- Volpi, M., Tuia, D., Camps-Valls, G. and Kanevski, M., 2011. Unsupervised change detection in the feature space using kernels. In: Geoscience and Remote Sensing Symposium (IGARSS), 2011 IEEE International, IEEE, pp. 106–109.
- Wang, F. and Xu, Y., 2010. Comparison of remote sensing change detection techniques for assessing hurricane damage to forests. Environmental monitoring and assessment 162(1), pp. 311–326.
- Wulder, M., Dymond, C., White, J., Leckie, D. and Carroll, A., 2006. Surveying mountain pine beetle damage of forests: A review of remote sensing opportunities. Forest Ecology and Management 221(1-3), pp. 27–41.
- Zhang, T., 2001. An introduction to support vector machines and other kernel-based learning methods. AI Magazine 22(2), pp. 103.