COARSE TO FINE PATCHES-BASED MULTITEMPORAL ANALYSIS OF VERY HIGH RESOLUTION SATELLITE IMAGES

Shiyong Cui, Mihai Datcu

Remote Sensing Technology Institute (IMF),German Aerospace Center (DLR)

ABSTRACT

In this paper, a patch based method for multi-temporal analysis of high resolution image is proposed. Conventionally, multi-temporal analysis performed at pixel level suffer from several restrictions, e.g., registration, bi-temporal analysis. To overcome these restrictions, two methods for multi-temporal analysis are proposed at patch level. One is for change detection in time series data by classifying all pairs of patches along time axis in the whole sequence into two classes. Features used for classification are similarity measures based on local statistical models and histogram of local patterns. The other aims at evolution analysis in long image time series. To characterize the evolution patterns, spatio-temporal local pattern features are extracted from time series data. ν-support vector machine (ν-SVM) is applied to classify different kinds of evolution at patch level. Performance is evaluated based on our database produced by iterative classification.

Index Terms— Multi-temporal analysis, image time series, local pattern histogram, $\mathcal{G}$ model, Support vector machine (SVM), Synthetic aperture radar (SAR).

1. INTRODUCTION

In recent years, there has been a rapid development of very high resolution satellites, such as the TerraSAR-X and TanDEM-X German satellites and the five RapidEye satellites launched in 2008. A large amount of multi-temporal data containing valuable information can now be acquired, which enables rapid response disaster and emergency monitoring. However, most available methods for multi-temporal analysis are not able to fully explore the high resolution data because they are introduced mainly for medium resolution images and most are conventionally performed on pixel level [1]. Therefore, multi-temporal analysis of very high resolution images is still a challenging problem in the community of remote sensing. In this paper, we focus on patches-based multi-temporal analysis [2] of high resolution SAR images including change detection and evolution of dynamic processes. The goal of the proposed method is to obtain the categories of evolution at patch level and annotate them.

2. FEATURE EXTRACTION

2.1. Image registration

As an important preprocessing step for multi-temporal analysis, image registration with sub-pixel accuracy is needed for the classical pixel-based multi-temporal image analysis. However, in the case of very high resolution images, co-registration is very difficult. Therefore, to overcome these drawbacks, a patch-based image registration instead of global registration is originally proposed. In this way, the impact of mis-registration on the accuracy of multi-temporal analysis can be reduced. Therefore, we assume affine transform between two corresponding patches. An FFT- based method [3] was applied to recover translation, rotation and scale parameters. As in this paper, TerraSAR-X image is used, in most cases we can assume there is only translation along azimuth and range. To determine the translation parameter along range and azimuth, phase correlation or normalized correlation coefficients can be applied which will be further improved by manual investigation. It is worth to note that patch-based registration aims at optimization of feature extraction rather than global registration accuracy.

2.2. Statistical features

In the literature, many statistical models have been proposed for SAR image modeling., e.g., $\mathcal{K}$ distribution, $\mathcal{F}$ distribution and $\mathcal{G}$ distribution. The parameters of these models are usually needed to be estimated and used as features for further processing. Among these models, $\mathcal{G}$ model has been proved be able to model extremely heterogeneous region, especially for urban area. However, the estimation of the model can be only addressed by maximum likelihood or method of moment (MOM) [4]. MOM can not estimate the number of looks, therefore the accuracy of data fitting is decreasing especially for urban areas. An estimation method based on second kind statistics is proposed. The $\mathcal{G}$ model of the SAR intensity is defined as

$$P(x) = \frac{n^n \Gamma(n-\alpha)}{\gamma^\alpha \Gamma(n)\Gamma(-\alpha)} \frac{x^{n-1}}{(\gamma+nx)^{n-\alpha}} \quad (1)$$

where x is the SAR intensity value, n is the number of looks, γ is scale parameter and α is shape parameter. We pro-

pose to use second kind statistics [5] for estimation, which is based on Mellin transform. The first second-kind characteristic function is defined as the Mellin transform of the $\mathcal{G}$ model and given as.

$$\phi(x) = \left(\frac{\gamma}{n}\right)^{x-1} \frac{\Gamma(n+x-1)\Gamma(-\alpha-x-1)}{\Gamma(n)\Gamma(-\alpha)} \quad (2)$$

The second second-kind characteristic function is given by the logarithm of the first second-kind characteristic function, defined as

$$\xi(x) = (s-1)\ln(\frac{\gamma}{n}) + \ln\Gamma(n+x-1)$$
$$+ \ln\Gamma(-\alpha-x+1) - \ln\Gamma(n) - \ln\Gamma(-\alpha) \quad (3)$$

Second kind log-cumulants of $\mathcal{G}$ model is given as follows by the derivative of the second second-kind characteristic functions.

$$\begin{aligned} K_1 &= \ln(\gamma/n) + \Psi(n) - \Psi(-\alpha) \\ K_i &= \Psi(i-1,n) + (-1)^i \Psi(i-1,-\alpha) \end{aligned} \quad (4)$$

where $\Psi(x)$ is the Digamma function and $\Psi(i-1,x)$ is the $i-1$ order Polygamma function. If assume the sample log-cumulants $\hat{K}_i$ are equal to second kind log-cumulants K_i, the following equations can be derived

$$\begin{aligned} \ln(\hat{\gamma}/\hat{n}) + \Psi(\hat{n}) - \Psi(-(\hat{\alpha})) &= \hat{K}_1 \\ \Psi(1,\hat{n}) + \Psi(1,-\hat{\alpha}) &= \hat{K}_2 \\ \Psi(2,\hat{n}) - \Psi(2,-\hat{\alpha}) &= \hat{K}_3 \end{aligned} \quad (5)$$

These nonlinear equations can be solved for parameters by numerical methods, such as Newton-Raphson Method and Trust region algorithms. It is worth to note that it may not converge rarely if the initial value of the parameters are not appropriate. In this case, the parameters can be derived by interpolation using the neighboring values. These estimated parameters $\hat{\gamma}$, $\hat{n}$ and $\hat{\alpha}$ can be used as feature vector for image content characterization.

2.3. Local pattern features

Although statistical model is important in characterizing scatters in SAR images, they are not able to describe the structures presented in high resolution SAR images. A promising texture feature particularly for SAR image based on local patterns [6] was proposed for classification. It was shown that it outperforms other texture features, e.t., features based gray level co-occurrence matrix, gabor filter and Gaussian markov random filed. It first thresholds the images with a series of thresholds, which results in three kinds of patterns. Then the histograms of the local pattern size are computed and concatenated. This is only applicable for images rather than long image time series. In this paper, we extend this feature for
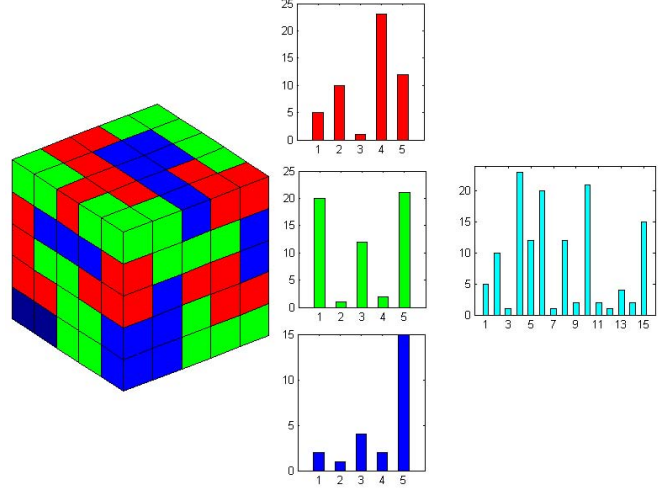

Fig. 1. Extended local pattern feature extraction for image time series.

time series data by computing the histogram of local patterns in spatio-temporal space. The procedure is demonstrated in Fig.1, which supposes the patch size is 5×5 and images at five different times are available, resulting a $5 \times 5 \times 5$ spatio-temporal space. A series of thresholds t_i are selected to threshold the patches by the interval $g_c \pm t_i$ (g_c is the value of the center pixel). Each threshold will result in three kinds of local patterns, representing local volumes in patio-temporal space. Three kinds of patterns in spatio-temporal space are indicated by colors (red, green, blue), corresponding to bright, homogeneous and dark targets in SAR image time series. The histogram of the three kind local patterns H_r, H_g and H_b in spatio-temporal space are computed with increasing bin size. The three histogram are concatenated as a feature vector $H = (H_r, H_g, H_b)$.

3. CHANGE DETECTION AND EVOLUTION ANALYSIS

3.1. Change detection

Conventional change detection suffers from registration and can not provide high level change information as low level features do not have semantic meaning, Changes at patch level is more interesting rather than changes at pixel level. Therefore it is proposed to perform change detection at patch level by classifying all pairs neighboring patches along time axis, which can detect the time and the position of changes in image time series. In this work, similarity metric based on statistical features and local pattern feature is applied as features for change detection. The symmetric Kullback-leibler divergence, defined as follows, is used for detecting statistical
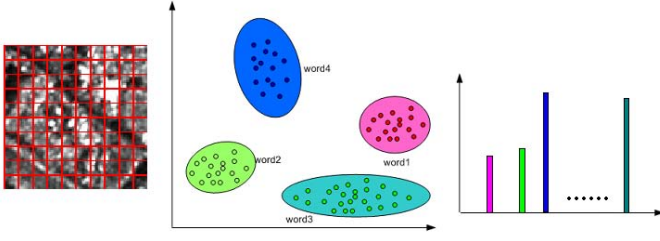
Fig. 2. Visual words extraction.

changes between two neighboring patches X_{t_i} and $X_{t_{i+1}}$

$$K(X_{t_i}, X_{t_{i+1}}) = KL(P_{X_{t_i}}(x), P_{X_{t_{i+1}}}(x)) \\ + KL(P_{X_{t_{i+1}}}(x), P_{X_{t_i}}(x)) \qquad (6)$$

where $P_{X_{t_i}}(x)$ and $P_{X_{t_{i+1}}}(x)$ are the $\mathcal{G}$ models estimated in section 2.2. As statistical features are low level features, high level features, such as visual words representing the content of the local patch, should be incorporated for change detection. Therefore, each patch is divided into 100×100 tiles. Multi-level local pattern features in 2D image presented in section 2.3 are extracted from each sub-tiles. The features are further quantized into visual words by using Gaussian mixture model (GMM), which has been shown superior to K-means clustering especially in the case of Gaussian distribution [7]. Each patch is represented by a unique feature vector which is the occurrence frequency histogram of visual words H_{t_i}. In this work, we assume the vocabulary size is 20. This is demonstrated in Fig.2. The Euclid distance between two frequency histograms of visual words is used to assess the similarity of two neighboring patches X_{t_i} and $X_{t_{i+1}}$.

$$D(X_{t_i}, X_{t_{i+1}}) = \|H_{t_i} - H_{t_{i+1}}\|^2 \qquad (7)$$

3.2. Evolution classification

Multi-level local pattern features in spatio-temporal space are used as temporal features and evaluated for evolution analysis. Each volume in spatio-temporal space is divided into tiles as shown in Fig.1. Multi-level local pattern features are extracted from each volume. All features extracted from each patch are quantized into visual words using GMM model. Therefore, each patch volume is represent by a histogram $H_{t_i}^x$ indicating the frequency of visual words. Support vector classification is applied to separate different kinds of evolution. Histogram intersection kernel [8] as a promising kernel function for image classification, defined as follows, is used for evolution classification.

$$k(H_{t_i}^x, H_{t_j}^y) = \sum_{l=1}^n \min(H_{t_i}^x(l), H_{t_j}^y(l)) \qquad (8)$$

where $H_{t_i}^x(l)$ and $H_{t_j}^y(l)$ are the components of the histograms.

3.3. ν-support vector machine

Given a set of training samples $\mathbf{x}_i \in R^n, i = 1, \ldots, l$ and the corresponding label vector $\mathbf{y} \in R^n$ such that $y_i \in \{1, -1\}$, it is achieved by solving the following quadratic programming problem in dual form

$$\min_\alpha \quad \frac{1}{2}\alpha^T Q \alpha$$
$$\text{s.t.} \quad \begin{cases} \mathbf{y}^T \alpha = 0, \\ \mathbf{e}^T \alpha = \nu l, \\ 0 \leq \alpha_i \leq 1/l, \ i = 1, \ldots, l. \end{cases} \qquad (9)$$

where $\mathbf{e}$ is the unit vector of length l, ν is the parameter controlling the number of support vectors and training errors, Q is the kernel matrix with $Q_{ij} = y_i y_j K(\mathbf{x}_i, \mathbf{x}_j)$. The kernel function used for change detection is a combination of two individual kernel given as

$$k(X, Y) = \mu k_1(X, Y) + (1 - \mu)k_2(X, Y) \qquad (10)$$

where μ is a constant and $k_1(X, Y)$ and $k_2(X, Y)$ are defined as

$$k_1(X, Y) = e^{-\gamma\sqrt{K(X_{t_i}, X_{t_{i+1}})K(Y_{t_i}, Y_{t_{i+1}})}}$$
$$k_2(X, Y) = e^{-\gamma\sqrt{D(X_{t_i}, X_{t_{i+1}})D(Y_{t_i}, Y_{t_{i+1}})}} \qquad (11)$$

4. EXPERIMENTS

A dataset consisting of 12 TerraSAR-X images covering Vâlcea county in Romania is used to creat the test database. These images are acquired every 11 days since angst 5, 2010 with incidence angle around 35.93^o and average height about 384 km. The images have a ground resolution of 2.9m. Each image is cut into 34×41 patches with patch size 100×100 pixels. Image alignment is performed by normalized correlation coefficients as there is only translation along azimuth and range direction. In most cases, the average translation is 2 pixels. The reference data for change detection and evolution classification based on this database is produced respectively by iterative classification using support vector machine. Two experiments were carried out respectively for change detection and evolution classification. The change detection result is shown in table 2. The overall accuracy is 95.42%.

Table 2. Confusion matrix of change detection

Pred. \ Ref.	No Change	Change
No Change	98.26%	15.94%
Change	1.74%	84.06%

Table 1. Confusion matrix of evolution classification

Ref. / Pred.	Forest	Building	Grassland	Water	Agriculture
Forest	90.33%	6.52%	7.69%	25.00%	8.99%
Building	2.42%	86.96%	2.14%	0.00%	0.72%
Grassland	3.82%	2.17%	81.20%	0.00%	3.96%
Water	0.25%	3.26%	2.14%	75.00%	0.00%
Agriculture	3.18%	1.09%	6.84%	0.00%	86.33%
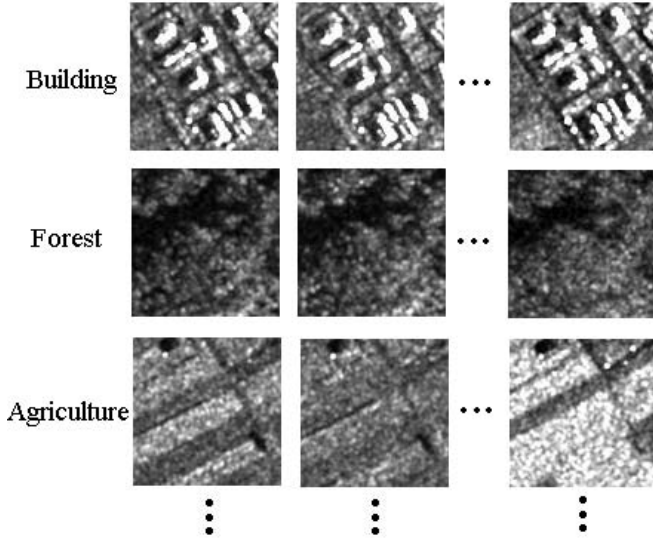
Fig. 3. Example of different evolution patterns from left to right.

In principle, each physical class should has one evolution classes and the number of evolution classes can be estimated by complex technique. For the sake of simplicity, we assume there are five evolution classes in our database, e.g., building, forest,agriculture,grassland and water as shown Fig.3. Classification result of evolution patterns is shown in table 1. The overall accuracy is 87.73%. The experimental results show promising ability of patch based multi-temporal analysis.

5. CONCLUSION

In this paper, a patch-based multi-temporal analysis approach is proposed and evaluated using TerraSAR-X images. Our method is different from pixel based analysis and overcomes the drawbacks of conventional methods by applying a patch-based registration and analysis. Patch-based registration aims at optimization of feature extraction. Statistical feature based on $\mathcal{G}$ model and patch-level feature, e.g., visual words based on local pattern features, are integrated for change detection and evolution pattern classification. Experimental results show promising ability of patch-based multi-temporal anal-

ysis. As a preliminary experiment, ν-SVM was applied for change detection and evolution classification. Multiple kernel learning will be investigated as a next step for performance improvement.

6. REFERENCES

[1] 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. 1171–1182, 2000.

[2] M. Molinier, J. Laaksonen, and T. Hame, "Detecting man-made structures and changes in satellite imagery with a content-based information retrieval system built on self-organizing maps," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 4, pp. 861–874, 2007.

[3] B. S. Reddy and B. N. Chatterji, "An fft-based technique for translation, rotation, and scale-invariant image registration," *IEEE Trans. Image Process.*, vol. 5, no. 8, pp. 1266–1271, 1996.

[4] A. C. Frery, H.-J. Muller, C. C. F. Yanasse, and S. J. S. Sant'Anna, "A model for extremely heterogeneous clutter," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 3, pp. 648–659, 1997.

[5] L. Bombrun and J.-M. Beaulieu, "Fisher distribution for texture modeling of polarimetric sar data," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 3, pp. 512–516, 2008.

[6] Dengxin Dai, Wen Yang, and Hong Sun, "Multilevel local pattern histogram for sar image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 2, pp. 225–229, 2011.

[7] R. R. Vatsavai, A. Cheriyadat, and S. Gleason, "Unsupervised semantic labeling framework for identification of complex facilities in high-resolution remote sensing images," in *Proc. IEEE Int Data Mining Workshops (ICDMW) Conf*, 2010, pp. 273–280.

[8] F. Odone, A. Barla, and A. Verri, "Building kernels from binary strings for image matching," *IEEE Trans. Image Process.*, vol. 14, no. 2, pp. 169–180, 2005.