CASCADE ACTIVE LEARNING FOR EVOLUTION PATTERN EXTRACTION FROM SAR IMAGE TIME SERIES

Shiyong Cui, Mihai Datcu

Remote Sensing Technology Institute (IMF)
German Aerospace Center (DLR)
shiyong.cui, mihai.datcu@dlr.de

ABSTRACT

In this paper, a cascade active learning approach relying on a coarse-to-fine strategy for evolution pattern indexing is developed, which allows fast indexing and hidden spatial and temporal pattern discovery in multi-temporal SAR images. In this approach, a hierarchical multi-level image representation is adopted and each level is associated with a specific patch size. SVM active learning is applied at each level to obtain reliable samples and reduce the manual effort in labeling the images. When moving to a new level, all the negative patches are neglected and the learning at the new level focuses only on the positive patches. In this way, the computation burden in annotating large data set could be remarkably reduced while keeping the accuracy. Through temporal pattern retrieval, the cascade active learning has been compared with a baseline SVM active learning operating only at the last level in terms of both accuracy and time complexity. We have demonstrated that cascade active learning can not only achieves better accuracy but also reduce remarkably the computation time.

***Index Terms*—** Synthetic aperture radar (SAR), SAR Image Time Series, Multi-temporal SAR, Cascade Active Learning, Evolution Pattern.

1. INTRODUCTION

In recent years, because of the short revisit times of high resolution satellites, like TerraSAR-X and TanDEM-X, SAR Image Time Series (SITS) can be constructed easily. SITS exhibits the dynamic property of the scene and provides detail information in the time span, which is a quite new topic and has a promising potential in a earth observation. However, most works in the literature focus on bi-temporal image analysis, among which SAR change detection [1] is a typical application since last several decades. Nevertheless, what evolution patterns can be extracted from SITS and how to represent and extract these patterns are still unanswered. Recently, a method based on frequent sequential pattern mining is proposed in [2] to extract groups of pixels sharing the same temporal evolution. Unfortunately, it is a pixel level method, which becomes insufficient for high resolution SAR images as contextual information is needed. Thus, developing methods working at a higher patch level for exploring SAR image time series in a large scale is of significant importance for practical applications. In this paper, a cascade active learning system relying on a coarse-to-fine strategy for evolution pattern extraction from SAR image time series is presented.

2. CASCADE ACTIVE LEARNING

2.1. Overview

The cascade active learning methods [3] [4] operating at patch level are developed to manage the increasing volume of SAR image time series. The motivation of the approach is to disregard as many as irrelevant patches at coarse level and and focus the learning and computation on the relevant patches at detail level. The overview of the cascade active learning is shown in Fig.1. It mainly consists of three components, which are feature extraction, cascade active learning, and visualization. The learning part is at the heart of the system, which includes SVM active learning and multiple instance learning. Visualization is an important part for multi-temporal SAR images, which involves conventionally dimension reduction techniques. However, any dimension reduction method would distort the image content and lose certain information. To keep all the information visitable to the oracle and highlight the content variation, a simple yet efficient color animation is used.

Starting from the first level as shown in Fig.1, all patches are classified as positive and negative. However positive patches probably contain some other classes, as shown in the classification results at the first level (the two patches between water and beach). These positive patches are classified further at the next level to eliminate the irrelevant parts. Nevertheless, training samples are not available at the finer level, which should be inferred from the samples at the previous level. As each training sample is considered as a bag containing sub-patches as instances, the optimal positive sub-patches can be learned by multiple instance learning. Using the learned optimal sub-patches as training samples, all the sub-patches of the positive patches at the previous level are classified again,

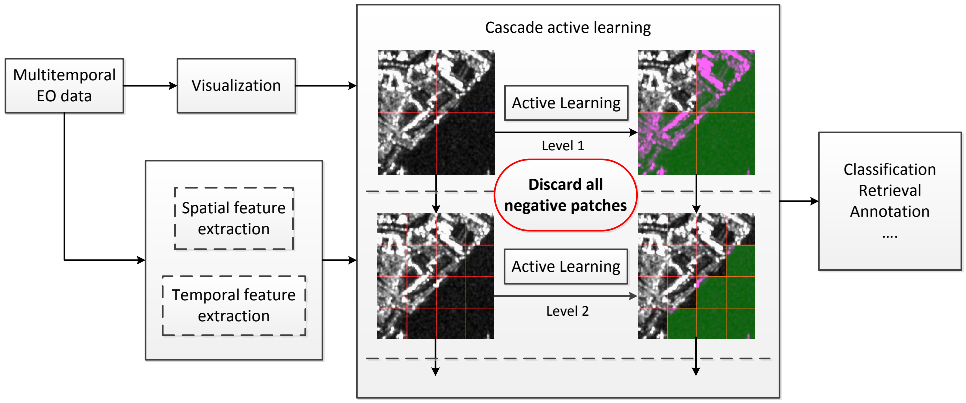


Fig. 1. Overview of cascade active learning for evolution pattern extraction from SAR image time series.

resulting in a better annotation.

2.2. SVM-Based Active Learning

In practice, it is usually costly to obtain the labeled data due to limited financial and human resources. To cope with the shortage of samples, active learning has been applied for large data set. In active learning, it is allowed to query the important instances in the feature space to maximize the accuracy and shorten the time of learning. Based on the notation of version space [5], the instance selected in each iteration should be able to split the current version space into two equal parts as much as possible. However, it is not practical to explicitly compute the sizes of the two parts given a new unlabeled instance which separates the version space into two parts. If we assume that the version space is symmetric, the normal vector w^* of the optimal decision surface is often roughly in the center of the version space. With this assumption, the instance that correspond to the closest hyperplane to the optimal w^* in the version space should be able to separate the version space into two equal parts. Thus, the unlabeled instance from the pool that is the closet one to the current decision surface should be queried as the most informative one.

2.3. Multiple Instance Learning (MIL)

Another issue in the system is how to infer the training samples from the ones that are selected or learned at the previous level. This is approached by SVM multiple instance learning. Different from supervised learning, data points in MIL are grouped into bags and only the bag labels are available. Formally, input data points $x_1, ..., x_n$ are grouped into some bags $B_1, ..., B_m$, each associated with a bag label $Y_I = \{-1, 1\}, I \subseteq \{1, ..., m\}$. The assumption of MIL is that there is at least only one positive instances in a positive bag and all instances are negative in negative bags. In our case, positive and negative bags are respectively positive and negative patches with contained sub-patches as instances. This assumption can be interpreted as two linear constraints for optimization.

$$\sum_{i \in I} \frac{y_i + 1}{2} \geq 1, \quad s.t. \quad Y_I = 1 \tag{1}$$

$$y_i = -1, \quad s.t. \quad Y_I = -1. \tag{2}$$

Based on these two assumptions, the objective is to identify the witness instance of each positive bag. The concept of margin maximization in SVM is extended from instance margin to bag margin, which results in MI-SVM. The margin of a bag with respect to a hyperplane is defined as

$$D_I = Y_I \max_{i \in I} (\boldsymbol{w}^T x + b) \tag{3}$$

In other words, the margin of a bag is the maximum distance between the hyperplane and all of its instances. Based on the notation of bag margin, MI-SVM is formulated as

$$\begin{aligned} \min_{\boldsymbol{w}, b, \boldsymbol{\xi}} \quad & \frac{1}{2} \|\boldsymbol{w}\|^2 + C \sum_{I=1}^{m} \xi_I \\ \text{s.t.} \quad & Y_I \max_{i \in I} (\boldsymbol{w}^T x_i + b) \geq 1 - \xi_I, \quad \xi_I \geq 0 \end{aligned} \tag{4}$$

For a negative bag, the constraints can be decomposed for each instance as $-\boldsymbol{w}^T x_i - b \geq 1 - \xi_I$ with $i \in I$ and $Y_i = -1$. For a positive bag, a selector variable $S(I) \in I$ is defined to denote the index of the witness instance in the positive bag. All instances which are not selected are discarded, thus has no influence on the learning. Therefore, the constraint for a positive bag can be rewrite as $\boldsymbol{w}^T x_{S(I)} + b \geq 1 - \xi_I$ with $Y_i = 1$.

Nevertheless, the optimization in Eq. (4) is an mixed integer programming problem, which can not be solved efficiently by the methods in the literature. Therefore, we adopt the heuristic optimization proposed by S. Andrews in [6], which is summarized in 1. Starting form the mean instances of the positive bags as positive instances, an alternate optimization between the update of the classifier and the identification of

Data: The training bags $D = \{B_1, ..., B_m\}$ and bag labels $L = \{l_i, ..., l_m\}1$
Result: The bag-level classifier C-SVM
begin
  $P = \{X_I^0 = \frac{1}{|I|}\sum_{i \in I}^{x_i} |Y_I = 1\}$;
  $N = \{x_i | Y_I = -1, i \in I\}$;
  repeat
    $P^k = P$;
    $S = P^k \cup N$;
    Train a classifier C-SVM using samples S ;
    $P = \phi$;
    forall the *positive bags* B_I **do**
      $y_i = \mathrm{sgn}\,(f(x_i)), \quad i \in I$;
      $X_I^k = \max_{x_i \in B_I} f(x_i)$;
      Add X_I^k to P ;
    end
  until $P^k = P$;
end

Algorithm 1: MI-SVM

the witness is performed. In each iteration, the most positive ones in the positive bags are selected as the witnesses.

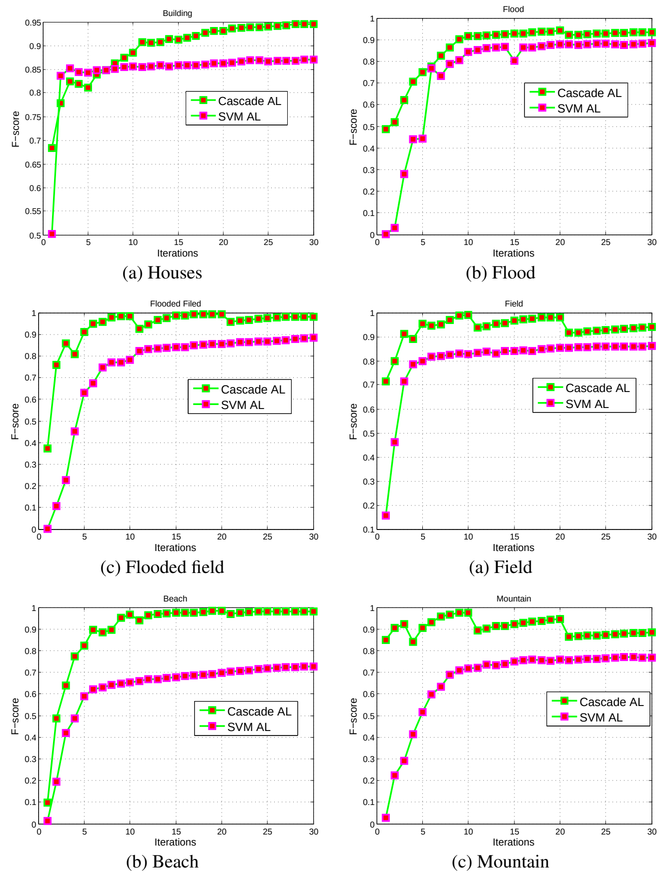
3. DATA SET AND EVALUATION

3.1. Data set

Before and after the earthquake in Japan in 2011, two sets of 17 Stripmap TerraSAR-X images every 11 days covering Sendai with HH polarization are acquired with respectively descending (9 images) and ascending (8 images) orbits. There is a small part of overlap between descending and ascending orbit. The temporal intervals of these two data sets are around 11 days. The images in each data set have exactly the same resolution and number of looks in both range and azimuth direction. The average incidence angles and the height are quite close. Due to the earthquake, it triggered an devastating tsunami, which destroyed a lot of constructions close to the Sendai airport and gives different temporal pattern. Therefore, these two data sets are quite good candidates of SAR image time series. The feature used in this system is the Bag-of-Spatial-Temporal-Words (BoSTW), which is a generalization of normal Bag-of-Words (BoW) feature [7] to multi-temporal SAR images

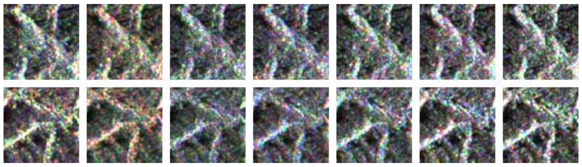
3.2. Experiment

We have evaluated the method and the system through SAR ITS temporal pattern retrieval. Among the six classes, flood, flooded field, and field are dynamic temporal patterns and the other three classes, etc., houses, beach, and mountain, are stable classes. The evolution patterns for two examples patches of houses and flood are shown in Fig.2. SVM active learning performed only on the last level is selected as a baseline for comparison with the cascade active learning. Sample patches with size 200×200 of each class are selected manually to initialize the system. Totally, 10 samples are selected and in each round of active learning 20 patches that are close to the decision boundary are manually annotated. 10 iterations are performed at each level, thus, totally 3×10 iterations are performed. For the sake of fairness, 30 iterations of the baseline SVM active learn is also performed on the last level with patch size 50×50. The same number 20 of patches are manually labeled and added to the training data. In each iteration, F-score is computed. Meanwhile, the time consumed by the training and classification is recorded. The F-scores for the six classes are shown in Fig.3. It can be seen clearly that the cascade active learning performs always better than the baseline SVM active learning. More important is that the computation burden has been remarkably reduced as can be seen from Fig.4. Obviously, the computation time of SVM active learning increases linearly with respect to iterations. In contrast, the computation of cascade active learning does not increase as the level goes down for medium classes. For large classes, like flood and houses in the ascending data set,

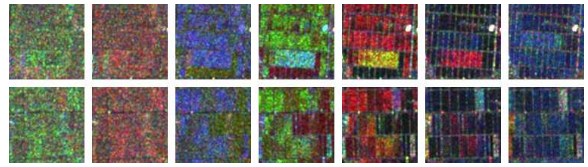


(a) Houses (b) Flood
(c) Flooded field (a) Field
(b) Beach (c) Mountain

Fig. 3. F-score of the six classes

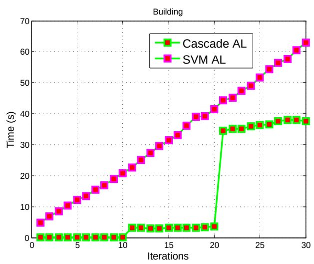


(a) Houses

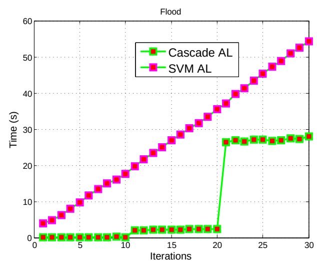


(b) Flood

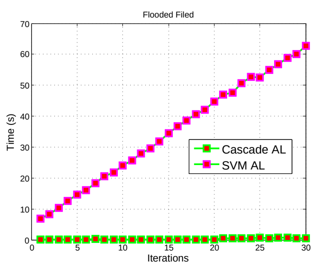
Fig. 2. Temporal evolution of example patches (a) patches in each row show the temporal evolution in mountains from level to right; (b) patches in each row show the temporal evolution in flooded field from level to right.



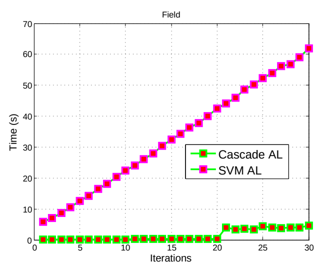
(a) Houses



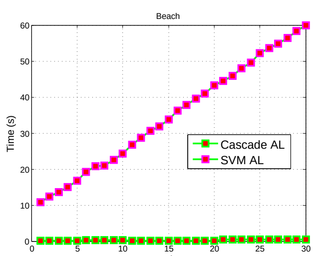
(b) Flood



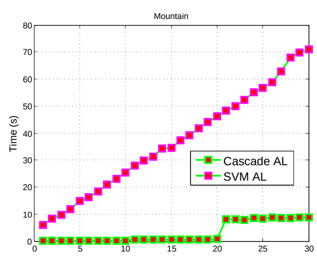
(c) Flooded field



(a) Field



(b) Beach



(c) Mountain

Fig. 4. Computation time for the six classes.

the computation of cascade active learning would increase because there are much more patches that are needed to classify when moving to a new level.

4. CONCLUSION

In this paper, a cascade active learning method has been developed and evaluated for evolution pattern indexing from SAR image time series. With this system, we can quickly find the target class. We compared this cascade active learning approach with a baseline SVM active learning performed only at the last level through temporal pattern retrieval using a real data base consisting of TerraSAR-X image time series. It has been demonstrated that cascade active learning can not only improve the accuracy but also reduce significantly the the computation burden for large data set.

5. REFERENCES

[1] J. Inglada and G. Mercier, "A new statistical similarity measure for change detection in multitemporal sar images and its extension to multiscale change analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 5, pp. 1432–1445, 2007.

[2] A. Julea, N. Meger, P. Bolon, C. Rigotti, M.-P. Doin, C. Lasserre, E. Trouve, and V. N. Lazarescu, "Unsupervised spatiotemporal mining of satellite image time series using grouped frequent sequential patterns," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 4, pp. 1417–1430, 2011.

[3] P. Blanchart, M. Ferecatu, and M. Datcu, "Cascaded active learning for object retrieval using multiscale coarse to fine analysis," in *Proc. 18th IEEE Int Image Processing (ICIP) Conf*, 2011, pp. 2793–2796.

[4] S. Cui, M. Datcu, and P. Blanchart, "Cascade active learning for SAR image annotation," in *Proc. IEEE Int. Geoscience and Remote Sensing Symp. (IGARSS)*, 2012, pp. 2000–2003.

[5] Simon Tong, *Active learning: theory and applications*, Ph.D. thesis, 2001.

[6] Stuart Andrews, Ioannis Tsochantaridis, and Thomas Hofmann, "Support vector machines for multiple-instance learning," in *Advances in Neural Information Processing Systems 15*, 2003, vol. 15, pp. 561–568.

[7] S. Cui, C. O. Dumitru, and M. Datcu, "Ratio-detector-based feature extraction for very high resolution SAR image patch indexing," *IEEE Geosci. Remote Sens. Lett.*, to be published, Early Access.