CLUSTERING OF DETECTED CHANGES IN SATELLITE IMAGERY USING FUZZY C-MEANS ALGORITHM

O. Sjahputera, G.S. Scott, M.K. Klaric, B.C. Claywell, J.M. Keller, N.J. Hudson, C.H. Davis

Center for Geospatial Intelligence
Department of Electrical & Computer Engineering
University of Missouri-Columbia

1. INTRODUCTION

We introduced a change-detection and exploitation system for high resolution commercial satellite imagery called GeoCDX [1]. In this paper, two multitemporal and multiband images of the same scene are ingested and co-registered. Fourteen spectral, linear, and textural features are calculated for each pixel [2][3]. Feature differences are found and aggregated to produce the pixel confidence. The pixel confidence values are filtered using the stack filter [1] producing the change score. The scene is divided into a tile grid; each tile covering 256x256 m area. The sum of all pixel change scores in a tile is defined as the tile change score. The tile ranking based on change score is generated as the main product of GeoCDX. Using the relevance feedback (RF) tool in GeoCDX, users can re-rank the tiles based on similarity to user-selected tile samples. Although RF is a useful tool, it still requires users to find the sample tiles from the tile ranking. Sometimes the first hundreds of tiles in the ranking are taken by tiles containing high-degree yet uninteresting changes such as agricultural, cloud cover, and seasonal changes. Thus, finding sample tiles of specific type of change can be a time-consuming process. Another alternative is to give users groups of tiles where tiles within a group share similar change characteristics. This allows users to zero in quickly on the type of change of interest. Here, we propose to utilize clustering algorithms to find these tile groups.

Clustering algorithms have been utilized in a number of prior studies in image change detection. Celik [4] employed *c*-means clustering and principal component analysis (PCA) to perform change detection on multi-temporal satellite imagery. Gosh, *et al.* [5] found that change detection on multi-temporal satellite imagery using fuzzy *c*-means (FCM) and Gustafson-Kessel clustering algorithms produced better results than those obtained using Markov Random Field and other neural network based algorithms. Carlotto [6] proposed cluster-based anomaly detection (CBAD) based on the Gaussian-mixture model to detect man-made objects in multitemporal multiband imagery where change pixels are found by detecting significant deviations from the distribution of a cluster containing mostly background pixels. Gaussian-mixture is an EM-style algorithm. Another EM-style algorithm called the hierarchical modal clustering (HMAC) was also used in a multi-band change detection approach proposed in [7]. All of these studies utilized some clustering algorithms to identify change pixels. In our case, the change confidence value for each pixel has been calculated using the stack filter, and these confidence values are aggregated to produce a tile change score. So we do not utilize the clustering algorithm to detect change. Here, propose to find tile clusters where each cluster contains tiles with similar change characteristics.

2. FEATURE VECTOR AND DISTANCE METRIC

For each pixel in the imagery, we extract fourteen features comprising of six spectral features (PAN, R, G, B, NIR, and NDVI), four linear features (pixel length, pixel width, length azimuth angle, and angle/width angle) [2], and four texture-based features calculated using Shannon's entropy and skewness measure [3]. Therefore, there is a feature vector in $\mathbb{R}^{14}$ associated with each pixel. Let K be the number of feature (K=14). Since the clustering will operate on tiles, we need to define a suitable feature vector for these tiles. Let I_1 and I_2 be the two multitemporal images where I_1 is older than I_2. Let T be the number of tiles found in I_1 and I_2 coverage area. Let h_{tk}^i be the histogram of feature k from all pixels in tile t from image i, where $k \in \{1,\cdots,K\}$, $t \in \{1,\cdots,T\}$ and $i \in \{1,2\}$ for I_1 and I_2, respectively. The histogram bins in h_{tk} are normalized such that the sum of distribution density over the range of feature values represented by each bin is equal (16 bins are used). The distribution density is derived from the entire coverage area of both imageries. Moreover, h_{tk}^i is normalized to 1 by area. We now define the intersection histogram $\ell_{tk} = h_{tk}^1 \cap h_{tk}^2$ for tile t that captures the unchanged profile of feature k in tile t of I_1 and I_2. The intersection operator is implemented using the minimum. In addition to its features, a pixel in tile t also has a confidence value. This is the raw confidence used by the stack filter to calculate the pixel-wise change score. Similarly to h_{tk}^i, we also generate the histogram of pixel confidence in tile t defined as hc_t. We now can define the feature vector for tile t as $\Phi_t = [hc_t \quad \ell_{t1} \quad \cdots \quad \ell_{tk} \quad \cdots \quad \ell_{tK}]$. The distance between tiles t' and t'' is given by $D^2(t',t'') = d(hc_{t'},hc_{t''})^2 + \sum_{k=1}^{K} d(\ell_{t'k},\ell_{t''k})^2$ with the histogram dissimilarity given by $d(x,y) = \sum_{b=1}^{B} |x(b)-y(b)| \big/ \sum_{b=1}^{B} (x(b)+y(b))$ if x or y is a non-empty histogram, otherwise d=0, B is the number of histogram bins (B=16). The distance D^2 is simply a Euclidean distance whose vector components are the histogram dissimilarity measures. Here, d allows us to calculate a distance metric involving a variety of features having different ranges of value because d reduces the histogram comparison on each feature to a dissimilarity value bounded in [0, 1]. The minimum dissimilarity d=0 is achieved if both histograms are identical, whereas the maximum d=1 is achieved if none of the histogram bins overlap.

3. CLUSTERING ALGORITHMS AND INITIAL RESULTS

We need a clustering algorithm that produces a soft partition matrix that allows a tile non-zero membership in more than one clusters since a tile may contain more than one type of change. Currently we use the fuzzy c-means (FCM) clustering algorithm [8] which observes the cluster membership constraint of $\sum_{c=1}^{C} u_{tc} = 1$. FCM requires users the number of cluster C and the fuzzifier m>1 to be set. As m approaches 1, FCM is becoming more crisp in its membership calculation, thus behaving more like the traditional hard c-means algorithm. On the other hand, as $m \rightarrow \infty$ FCM cluster membership degenerates into 1/C. To initialize FCM, we first run the agglomerative hierarchical clustering (AHC) algorithm with a user specified C as the stopping criteria. The cluster prototype for each cluster produced by AHC is generated by taking the mean of Φ_t from tiles in the cluster. These cluster prototypes are used to initialize the FCM. Currently, the clustering process is still supervised in that a user determines the initial value for C and m. The user then determines whether these values need to be adjusted by visually evaluating the final clusters and the membership values of the tiles in the clusters produced by FCM. If tiles containing very different types of change are found in the same cluster and both have rather high u_{tc}, then we increase C expecting AHC to find two different clusters for the two types of change. We also found that the common m=2 tends to drive $u_{tc} \rightarrow$1/C. The high dimensionality and the fact that a tile may contain different types of change might have created significant overlaps between clusters. Therefore, we reduce the fuzzifier to 1.2. We use an IKONOS (taken 04/30/2000) and a QUICKBIRD (taken 06/28/2006) imagery of Columbia, MO, USA as our test data. The pair covers an area of 159 km^2 and contains 2538 tiles. We perform the clustering on the 615 tiles having change score ≥ 1.0 using C=15. Columbia, MO, is a small urban area surrounded by agricultural fields with a large number of property developments in the area within the 6 years time period. We show two of these clusters in Fig. 1 and 2. Each column shows tile k from I_1, I_2, and the pixel change score mapped into a color table (brighter color means higher intensity of change). These tiles are ordered based on their cluster membership. As shown in Fig. 1, the tiles contain changes made on previously developed areas. GeoCDX detects not only new construction, but also removed structures such as shown in the 5th ranked tile. Fig. 2 shows a cluster of tiles of changes made on previously undeveloped rural/agricultural areas. Both clusters contain changes in the forms of newly constructed buildings, but the second cluster also contains new road constructions which would affect the linear features while building construction or landscaping mostly affect spectral and textural features. We are currently moving towards unsupervised clustering where C is determined programmatically either by optimizing the cluster validity measures for a specific range of C, or by using unsupervised clustering algorithms currently under evaluation. The current clustering process is done offline, but the clustering results are available online through the GeoCDX user interface.

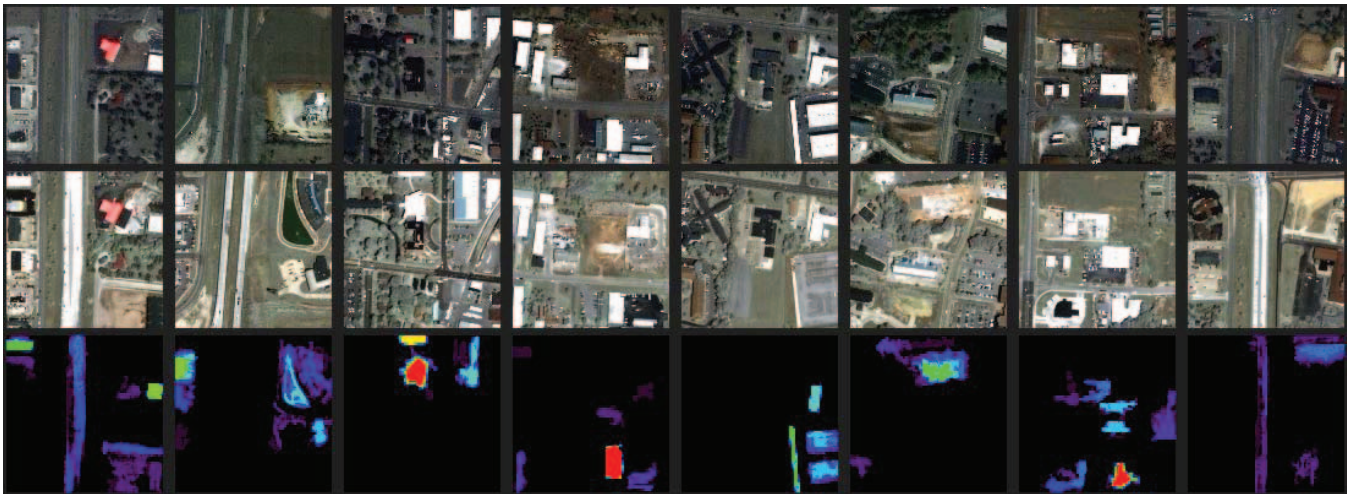
Fig. 1 Tile cluster containing changes found in developed areas.

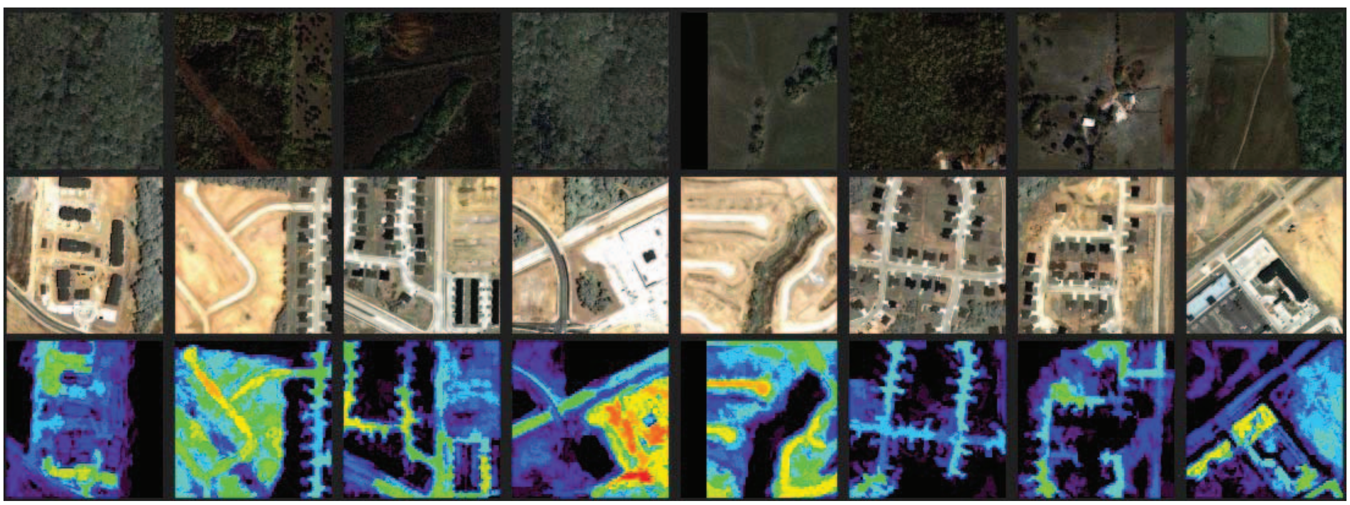
Fig. 2 Tile cluster containing changes found in newly developed areas.

4. REFERENCES

[1] O. Sjahputera, C.H. Davis, B.C. Claywell, N.J.Hudson, J.M. Keller, M.G. Vincent, Y. Li, M.N. Klaric, C.-R. Shyu, "GeoCDX: An Automated Change Detection & Exploitation System for High Resolution Satellite Imagery," *IEEE Int'l Geoscience and Remote Sensing Symposium (IGARSS)*, Boston, MA, vol. V, pp. 467-470, 7-11 July 2008.

[2] A.K. Shackelford, C.H. Davis, "A Hierarchical Fuzzy Classification Approach for High-Resolution Multispectral Data Over Urban Areas," *IEEE Trans. On Geoscience and Remote Sensing,* vol. 41, no. 9, September 2003.

[3] B. Claywell, C.H. Davis, C.R. Shyu, "Fusion of Spectral and Spatial Information for Automated Changed Detection in High Resolution Satellite Imagery," *IEEE Int'l Geoscience and Remote Sensing Symposium (IGARSS),* p.2510-2513, 2006.

[4] S. Ghosh, N.S. Mishra, A. Ghosh, "Unsupervised Change Detection of Remotely Sensed Images using Fuzzy Clustering," *Int'l. Conf. on Advanced in Pattern Recognition (ICAPR),* pp. 385-388, 4-6 Feb. 2009.

[5] T. Celik, "Unsupercised Change Detection in Satellite Images Using Principal Component Analysis and k-Means Clustering," *IEEE Letters On Geoscience and Remote Sensing,* vol. 6, no. 4, pp. 772-776, Oct. 2009.

[6] M.J. Carlotto, "A Cluster-Based Approach for Detecting Man-made Objects and Changes in Imagery," *IEEE Trans. On Geoscience and Remote Sensing,* vol. 6, no. 1, pp. 189-202, Feb. 2005.

[7] K. Griffis, and M. Bystrom, "Modeling and Clustering Techniques for Multi-Band Change Detection," *IEEE Int'l Geoscience and Remote Sensing Symposium (IGARSS)*, Boston, MA, vol. IV, pp. 89-92, 7-11 July 2008.

[8] J.C. Bezdek, *Pattern Recognition With Fuzzy Objective Function Algorithms,* New York, Plenum, 1981.