

# Technical Report

Department of Computer Science  
and Engineering  
University of Minnesota  
4-192 Keller Hall  
200 Union Street SE  
Minneapolis, MN 55455-0159 USA

TR 13-003

Supplement for "Change Detection from Temporal Sequences of Class  
Labels: Application to Land Cover Change Mapping"

Varun Mithal, Ankush Khandelwal, Shyam Boriah, Karsten  
Steinhaeuser, and Vipin Kumar

January 25, 2013



# Supplement for “Change Detection from Temporal Sequences of Class Labels: Application to Land Cover Change Mapping”

## S-1 Methodological Details

---

**Algorithm 1** Calculate change score using dynamic programming

---

**Require:**  $c^i, k, M, T, t1, t2, c1, c2$

```

 $n \leftarrow \text{length}(c^i)$ 
 $\alpha_{num}^{i,1} \leftarrow \frac{1}{k}$ 
 $\alpha_{den}^{i,1} \leftarrow \frac{1}{k}$ 
for  $t = 2$  to  $n + 1$  do
  for  $j = 1$  to  $k$  do
     $\alpha_{den}^{j,t} \leftarrow \sum_{i=1}^k \alpha_{den}^{i,t-1} t_{i,j}$ 
     $\alpha_{den}^{j,t} \leftarrow \alpha_{den}^{j,t} m_{j,c_i^i}$ 
    if  $t = t1$  then
       $\alpha_{num}^{j,t} \leftarrow \alpha_{num}^{c1,t-1} t_{i,j}$ 
    else
      if  $t = t2$  then
         $\alpha_{num}^{j,t} \leftarrow \alpha_{num}^{c2,t-1} t_{i,j}$ 
      else
         $\alpha_{num}^{j,t} \leftarrow \sum_{i=1}^k \alpha_{num}^{i,t-1} t_{i,j}$ 
      end if
    end if
     $\alpha_{num}^{j,t} \leftarrow \alpha_{num}^{j,t} m_{j,c_i^i}$ 
  end for
end for
 $\xi_{t1,t2}(c1, c2) \leftarrow \frac{\sum_j \alpha_{num}^{j,n}}{\sum_j \alpha_{den}^{j,n}}$ 

```

---

## S-2 Change Detection Experiments

Here, we provide the enlarged figure for the case study on change detection as Figure 4.

## S-3 Mixed Pixel Experiments

Here, we provide the enlarged figures for the case study on mixed class modeling as Figure 5.

## S-4 Experiments with Synthetic Data

We have shown using case studies on real data that the proposed framework uses the temporal context to improve classification accuracy of land cover classification products, imputes missing labels, significantly improves the precision of the change detection query, identify mixed class pixels from a label sequence generated using a classifier trained only on pure classes. Lack of available ground truth restricts any quantitative evaluation on the real data. In this section, we provide quantitative results for a synthetic data set for the aforementioned tasks. In addition, we show that the *change score* discussed in Section 3 can be used to meaningfully rank

the change detection results.

**S-4.1 Data Set Generation** Here we describe the procedure used for generating synthetic temporal label sequence data for evaluating the proposed framework. Input parameters are: the number of classes  $k$ , confusion matrix  $M$ , transition matrix  $T$ , prior probability  $P$ , number of pixels  $N$ , and the number of time steps  $n$ .

Step 1. Select a class for each pixel from  $k$  classes at random according to the prior probability  $P$ . The entire sequence is given that label. We calculate the prior probability using the probability of each label in the real data set.

Step 2. To introduce transitions, we first select a time of change randomly between 1 to  $n$ . Next, we use a transition probability matrix  $T$  to select the other class. If this class is same as class selected in Step 1, then no change occurs in the pixel. This step gives a data set with  $N$  pixels and  $n$  time steps with initial class in proportion to  $P$  and transitions in proportion to  $T$ . This gives the ground truth data set.

Step 3. In this step, we add misclassifications by reassigning entries in the data for each class according to the row in confusion matrix  $M$  corresponding to that class. This step gives us a noisy data set with misclassification according to  $M$ .  $M$  used to generate synthetic data is estimated using the confusions seen in the one-month apart change experiment.

Step 4. Next, we add missing values in the data. We first select the number of missing values to be added to a pixel using the same distribution as in real data and then add missing label to those many randomly chosen time steps.

The steps 1 to 4 are used to generate DS1.

Step 5. We add 500,000 sequences to data DS1, with labels from class 2 and 3, randomly chosen using a Bernoulli trial with probability 0.5 for each class. These represent the mixed class pixels and are added to DS1 to generate the synthetic data set DS2.

**S-4.2 Classification Performance** For data set DS1, when we compared between  $Z^1$  and  $C^o$ , the classification errors (aggregated over all time steps) reduced from 10% to 1%. However, we found that improvement in classification is not uniform at all time steps; and only a small improvement is seen at the two ends of the label sequence due to lack of sufficient temporal context. For  $Z^2$  we found that the classification errors reduced to 4%. The reason for higher classification error in  $Z^2$  compared to  $Z^1$  is that mixed class has some confusion

with the pure vegetation and urban class on DS1. Our approach is also able to accurately impute 90% of the introduced missing labels with their correct labels in the ground truth data. This observation corroborates with the qualitative results on the real data, where the images of new classification outputs ( $Z^1$  and  $Z^2$ ) appeared to be correctly imputing the missing data stripes.

Classification on DS2 is challenging for both  $C^o$  and  $Z^1$  which do not have a notion of mixed class. For  $Z^2$ , we found that 89% of the 0.5 million mixed pixels, were assigned a state sequence with all the labels as mixed class. In principle, no pixel in DS1 should get impacted by addition of a latent mixed class but the new, mixed class has some confusion with the pure vegetation and urban class on DS1. Quantitatively, we found that 3.1% of the pixels in DS1 have at least one of the labels in their sequence assigned to mixed class. The fact that 89% of mixed pixels are identified and only 3% of pixels in DS1 get confused with mixed label suggest that the model for mixed class has a high recall and precision.

**S-4.3 Change Detection** For this experiment, we consider the same query that we studied in the real data experiment, i.e., find pixels that transitioned from vegetation to urban between time step 3 and time step 17. Table 2 shows the precision and recall for the three classification products  $C^o$ ,  $Z^1$  and  $Z^2$  on DS1 and DS2. On the synthetic data set DS1, the precision for  $Z^1$  (50%) is higher than  $C^o$  (14%) as expected since the new classification  $Z^1$  is more accurate. These precision values are also in agreement with the estimates using the analysis  $\frac{p-2\epsilon p}{p+2\epsilon}$ . Real data resembles DS2 (that has the mixed class pixels in the data) and therefore we repeat the above experiment on DS2. The precision of  $C^o$  and  $Z^1$  on DS2 is much lower compared to DS1 because mixed class sequences tend to have a higher disparity between the class labels at any pair of time steps. The model without mixed class is able to correct these errors only to some extent and the precision improves from 8% to 18%. Subsetting by the confidence score further improves the precision from 18% to 24% on DS2. This is because most of the mixed pixels, which are responsible for these spurious changes, get a low confidence score and are therefore removed by the high confidence score filter.

Next, we evaluated  $Z^2$  on DS1 and DS2. We find that there is a decrease in recall on using  $Z^2$ . This decrease is expected because if there is confusion around the time of transition, the mixed model labels the subsequence around that transition as mixed. The precision of  $Z^2$  similar to  $Z^1$  for DS1, but is significantly higher on data set DS2. This improvement shows that the new model captures the notion of mixed class, instead of forcing them to be in either vegetation or

urban class which may lead to possibility of spurious changes. Thus, we see a trade-off between using  $Z^1$  or  $Z^2$  based on our preference for higher precision or recall. In a separate experiment (not discussed here for brevity), we found that the decrease in recall of  $Z^2$  compared to  $Z^1$  depends on the degree of confusion in DS1 and *increases* as this confusion *increases*. Similarly, the improvement in precision of  $Z^2$  compared to  $Z^1$  depends on the fraction of mixed pixels in the data set and *increases* as this fraction *increases*. Typically, we would prefer  $Z^2$  over  $Z^1$  on a large, real data set because the loss in recall is only moderate ( $\approx 10\%$ ) if the confusion in DS1 is between 5% to 10%, while the precision can be extremely low for  $Z^1$  if the fraction of mixed pixels in the data is large.

Data set	$C^o$	$Z^1$	$Z^2$
DS1	14%/63%	50%/93%	51%/84%
DS2	8%/63%	18%/93%	44%/84%

**Table 2:** Precision/Recall on synthetic data sets DS1 and DS2 for  $C^o$ ,  $Z^1$  and  $Z^2$ .

Finally, we demonstrate how the *change score* can help in providing better response to change queries. The set of changed pixels are ranked using this score and Figure 6 and 7 show the corresponding precision and recall curves. The *green* curves correspond to a random ranking of pixels, while the *red* correspond to the ranking using the proposed change score. We see that by ranking these pixel using the designed change score we can see the true changes with high precision before we start encountering the false positives as we improve our coverage of changes. This is useful for domain end-users who use the results of change detection for conducting ground surveys. These surveys are expensive and require tremendous manual effort, which can now be guided by use of change score. Figure 6 and 7 also show the precision and recall curves by ranking using confidence score (blue). We see that though confidence score does better than random, but change score is significantly better as it is specific to our query while confidence score measures the overall fit to the generative model.

Data set	$C^o$	$Z^1$	$Z^2$
DS1	15%/68%	53%/94%	56%/90%
DS2	11%/68%	24%/94%	46%/90%

**Table 3:** Precision/Recall on synthetic high confidence subsets of DS1 and DS2 for  $C^o$ ,  $Z^1$  and  $Z^2$ .



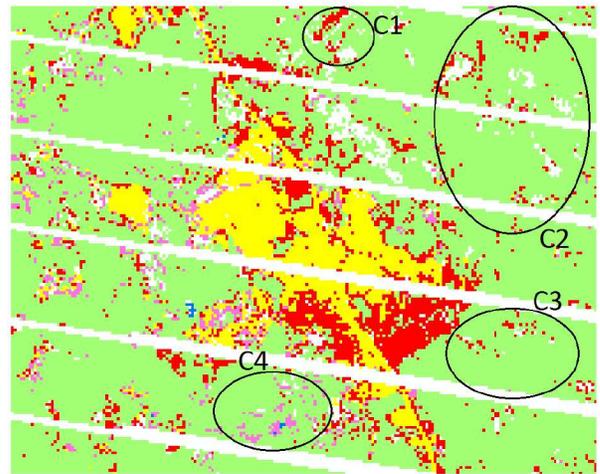
(a) Validation image for August 2003.



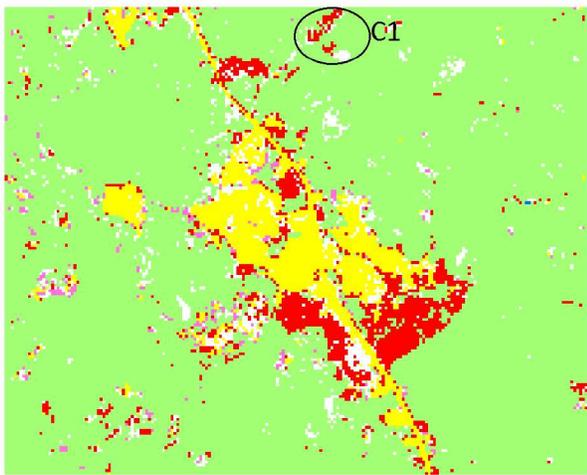
(b) Validation image for August 2011.



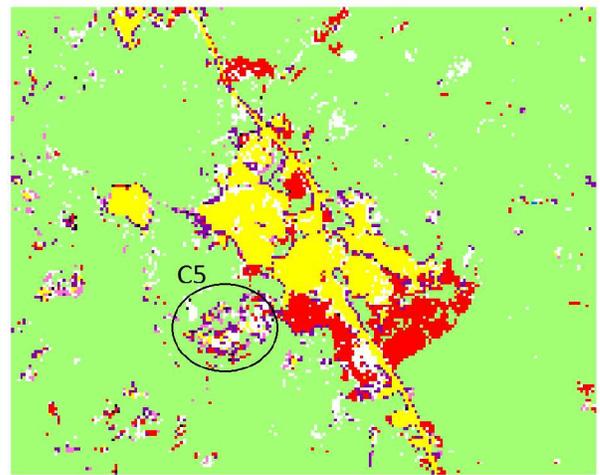
(c) Confidence score map, black implies low confidence.



(d) Change map using  $C^0$ .

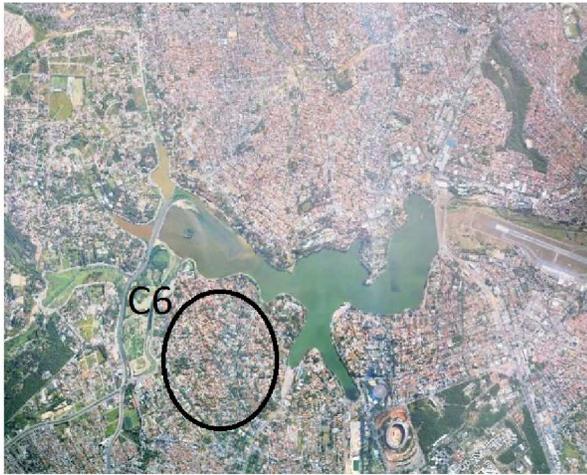


(e) Change map using multi-temporal context-based  $Z^1$  classification.

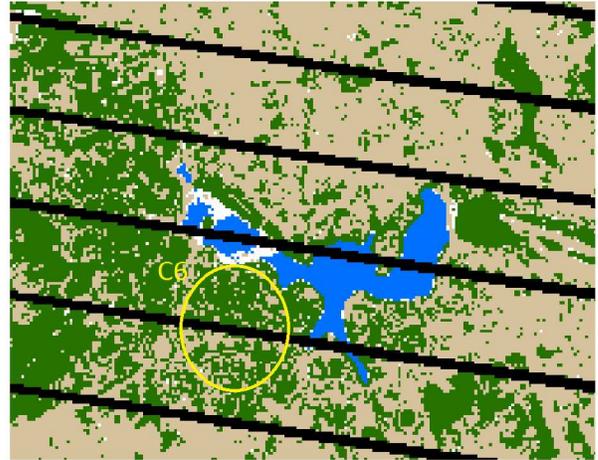


(f) Change map using multi-temporal context-based  $Z^1$  classification.

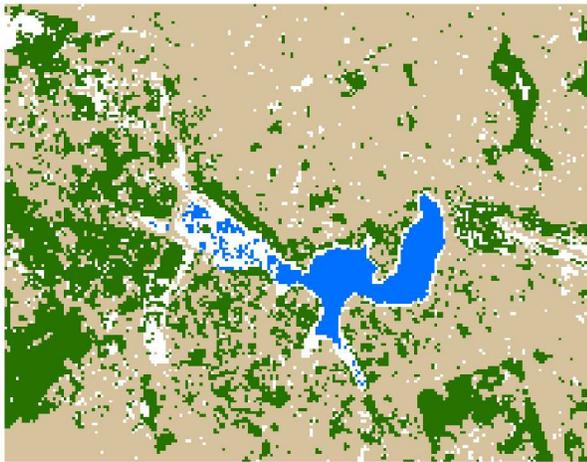
**Figure 4:** These figures show the change detection for a region of study between 2003 and 2011. (This is a larger version of Figure 2.)



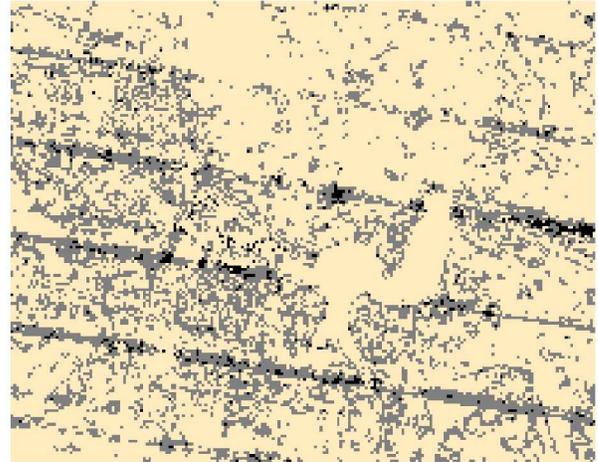
(a) Validation image for August 2008.



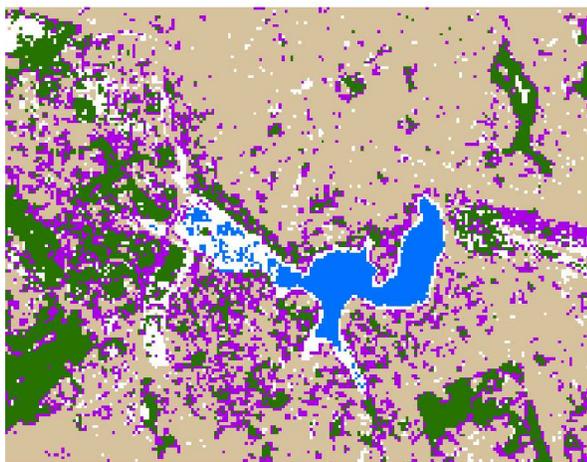
(b) Classification map using .5DLRclass for August 2008.



(c) Classification map from  $Z^1$  for August 2008.



(d) Confidence score map, black implies low confidence.

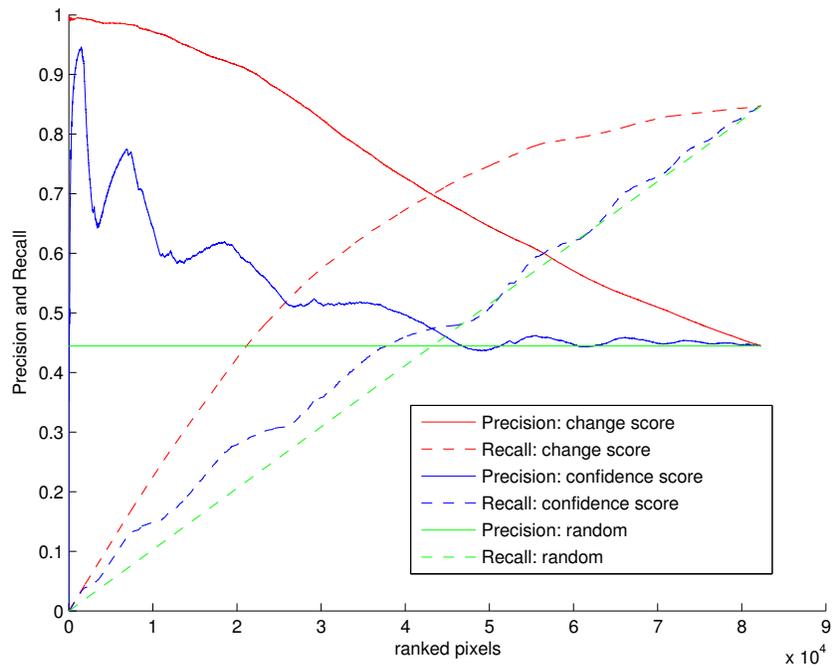


(e) Classification map  $Z^2$  after including mixed class in model for August 2008.

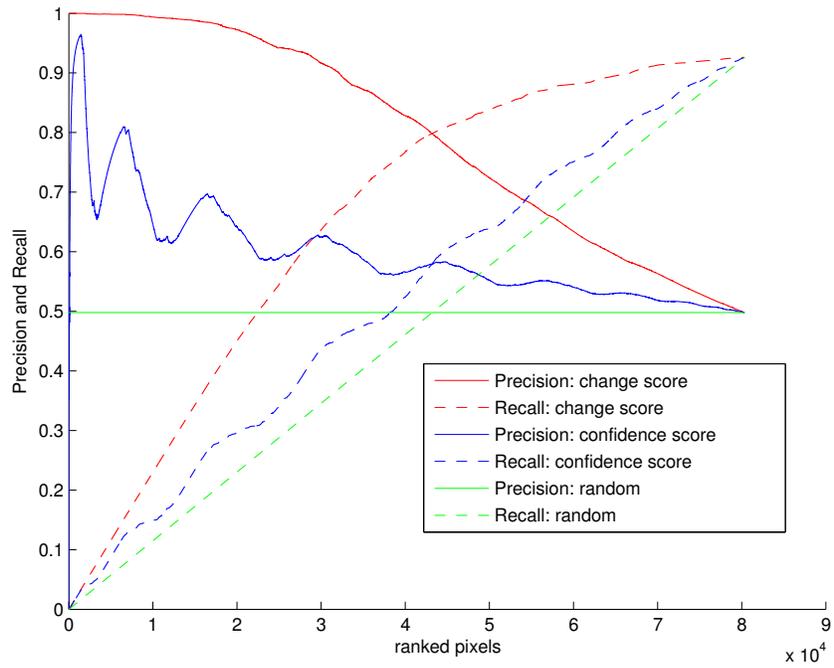


(f) Zoomed-in image for the area in C6 in August 2008.

**Figure 5:** These figures show the classification maps for a region of study in Belo Horizonte between 2003 and 2011. (This is a larger version of Figure 3.)



**Figure 6:** Precision and Recall curves for changed pixels discovered in DS2 using classification output  $Z^2$ . The three ranking mechanisms: Change score (red), Confidence score (blue) and random (green).



**Figure 7:** Precision and Recall curves for changed discovered in DS1 using  $Z^1$ .