# Fusion of Multitemporal Satellite Images and GIS Data for Land-use Classification

Anne H. Schistad Solberg[1], Anil K. Jain[2] and Torfinn Taxt[3]

[1] Norwegian Computing Center, P.O. Box 114 Blindern, N-0314 Oslo 3, NORWAY
[2] Michigan State University, East Lansing, Michigan 48824
[3] University of Bergen, 5000 Bergen, NORWAY

**Abstract**

*In this paper, we present a new method for fusion of satellite images and GIS ground cover data for land-use classification. The method is suited for classification of images captured at different dates, by incorporating knowledge about the class-dependent probabilities of changes with respect to the pattern classes. The method is tested by fusing Landsat TM images, ERS-1 SAR images, and GIS ground cover data for land-use classification. The test results show that significant improvements in the classification accuracy are possible by fusing GIS data with spectral imagery.*

## 1 Introduction

The availability of remotely sensed data of the same scene from different sensors is increasing. In addition to spaceborne spectral data, aerial photographs, forest maps, ground cover maps, and topographic information such as elevation and slope might be available to the analyst. The two main sources of information are (i) multispectral images from different satellite sensors, and (ii) spatial data from Geographic Information Systems (GIS). It is desirable to use most of these data in the image analysis to maximize the amount of useful information extracted. For an image segmentation or classification task, fusion of data from different sensors can be utilized to reduce the classification error obtained by single-source classification. Conventional multivariate classification methods require that the multisource data be described by a common spectral model. Such a model cannot be easily established for combining different data types, e.g., spectral data and categoric data from a GIS system. Another problem with the conventional multivariate approach is that the different sources might not be equally reliable. The development of an appropriate model for the classification of data from multiple sources is, thereby, an important problem for the processing of remotely sensed data.

Among the different approaches to data fusion in the remote sensing literature are statistical methods [1], and Demster-Shafer theory [6]. These previous studies assume that no changes with respect to the pattern classes have occurred between the acquisition of the different images. A challenge for further research is to include the concepts of time and class changes in the fusion models. In [12], we presented a model for fusion of multisensor satellite images in which the temporal aspect was included by allowing changes in the pattern classes and considering the class-dependent probabilities of changes. In this paper, we extend the model by incorporating GIS ground cover data. The performance of the model is evaluated by fusing Landsat TM images, ERS-1 SAR images, and GIS ground cover data for land-use classification. The fusion model gives significant improvements in the classification error rates compared to the conventional single-source classification results, or to a fusion model for combining spectral data only.

The remainder of this paper is organized as follows: In section 2, the basic elements in the fusion model are presented. Section 3 describes the interaction model between different sources. Experimental results are presented in section 4, and discussion and conclusions are given in section 5.

## 2 The basic fusion model

We present a brief description of the fusion model used in our study.

## 2.1 Bayesian data fusion

In the general multisource or sensor fusion case, we have a set of measurements or data, from $n$ sensors $X_s(i,j)$, $s \in \{1,n\}$ for each pixel $(i,j)$. The goal is to assign each pixel into one of the information classes $\omega_1, ..., \omega_C$. We might have some a priori information about the likelihood of the different classes, represented as $P(\omega_1), ..., P(\omega_C)$, the apriori probabilities of the classes $\omega_c$, $c \in [1,C]$. The relationship between the measurements and the a priori information is represented by

$$P(\omega_c \mid X_1(i,j), ..., X_n(i,j)) = \frac{P(X_1(i,j), ..., X_n(i,j) \mid \omega_c) P(\omega_c)}{P(X_1(i,j), ..., X_n(i,j))}, \quad (1)$$

where $c \in \{1,2,...,C\}$, $P(\omega_c \mid X_1(i,j), ..., X_n(i,j))$ is the a posteriori probability that $\omega_c$ is the correct class,
given the observed data $X_1(i,j), ..., X_n(i,j)$. The image formation model, $P(X_1(i,j), ..., X_n(i,j) \mid \omega_c)$ is the conditional probability that $X_1(i,j), ..., X_n(i,j)$ is the observed data , given that $\omega_c$ is the correct class. Each pixel is assigned to the class $c$ which maximizes $P(\omega_c \mid X_1(i,j), ..., X_n(i,j))$.

## 2.2 Elements of the fusion model

A schematic view of the fusion model is given in Figure 1. The fusion model consists of the following basic elements:

- *Input data.* The input images are assumed to be geocoded and co-registered. GIS data are assumed to be converted to raster format and co-registered with the images.

- *Image formation model.* A sensor-specific image formation model contains information about the underlying noise characteristics of the sensor. For remote sensing applications, the image formation model typically describes the class-conditional probability density function of the image data.

  We will model the Landsat TM images using the multivariate normal distribution. The six original non-thermal TM bands with a resolution of $30m \times 30m$ are the input data. TM images are not severely degraded by noise.

  To specify an image formation model for SAR images, we can either simply specify a probability distribution function for the speckle noise, or we can use a more complex model based on specific knowledge about the SAR signal processing operations and the resulting speckle autocorrelation function. Various probability distribution functions and more complex models derived from speckle statistics and autocorrelation have been developed [3, 4, 7, 8, 10]. Previous studies [11] have demonstrated that classification based on textural features in addition to the average backscatter can improve the classification error rates significantly. Following Frankot and Chellappa [4], we will model the SAR image using a multiplicative autoregressive random field (MAR). The parameters of the model have been used as texture descriptors [11].

  Let the observed SAR image $X(i,j)$ be represented by a white-noise-driven multiplicative system. Define $Y(i,j) = \ln X(i,j)$. Then we assume that $Y(i,j)$ follows a Gaussian autoregressive (AR) model

$$Y(i,j) = \sum_{r \in N} \theta_r (Y(i+r, j+s) - \mu_y) + u(i,j), \quad (2)$$

  where $N$ is the neighborhood system, and $\mu_y$ is the mean value of the stationary random process $Y$. The noise process $u(i,j)$ is uncorrelated white noise with variance $\sigma^2$[4]. We use the least squares estimates for the parameters $\theta$, $\sigma^2$, and $\mu_y$ [4]. The parameter estimates are computed in local windows of size 9x9 and with $N = \{(0,-1),(-1,-1),(-1,0)\}$. The estimated parameters will be used as feature vectors in the pixel classification.

- *Spatial context.* Contextual information from neighboring pixels normally improves the classification results compared to a pixel-by-pixel classification. In remote sensing applications, contextual information is often modelled using Markov Random Fields in a local spatial neighborhood. The choice of contextual model will influence the complexity of the classification algorithm. With computational aspects in mind, we have y chosen a non-iterative contextual model: Haslett's model [5]. The classification rule is based on a stochastic model for the behavior of the classes in the scene and the behavior of the feature vectors, given the underlying classes. For the detailed model, see [5, 12].

- *A priori probabilities.* In the Bayesian formulation of the classification problem, a priori prob-

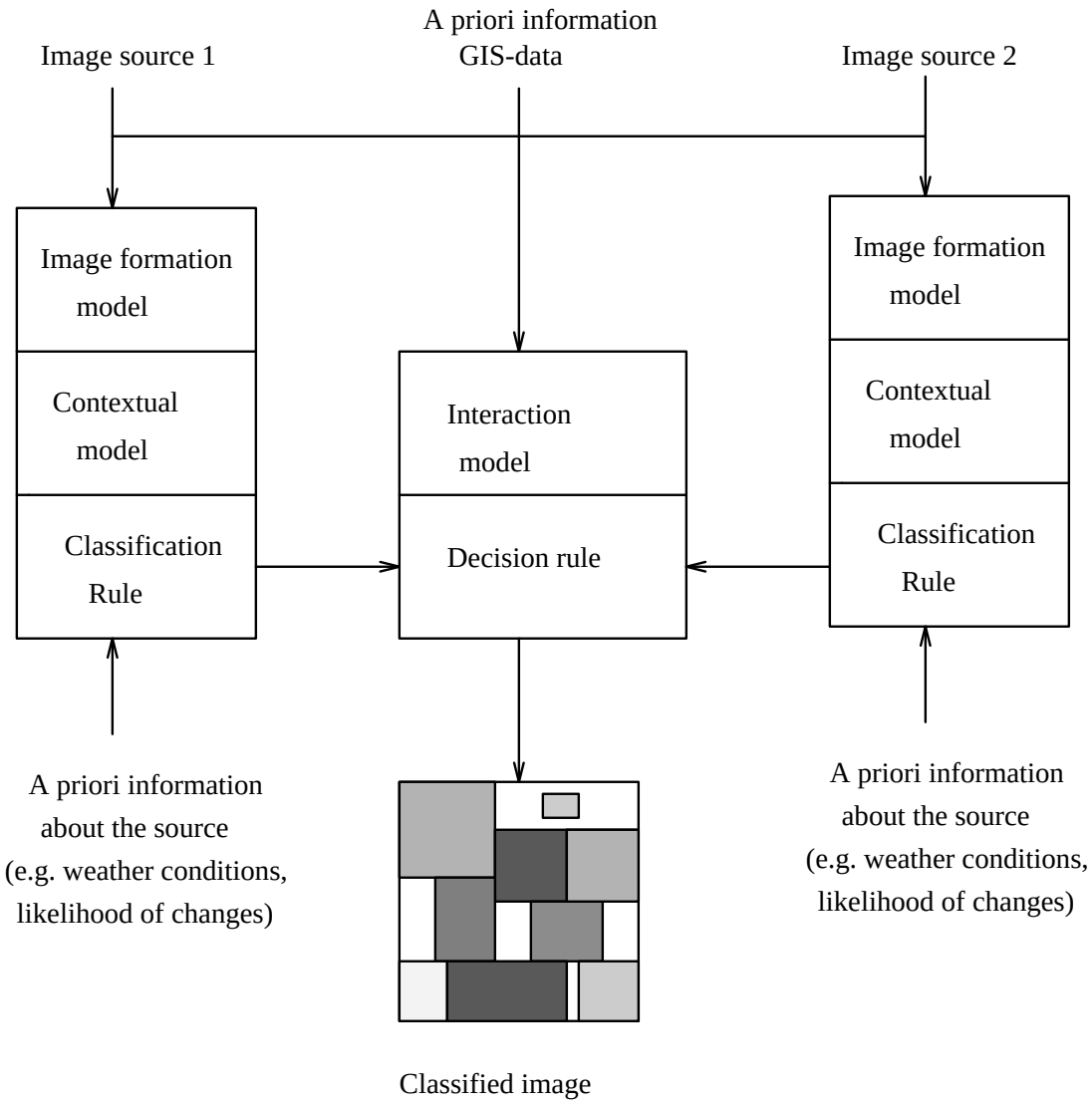


Figure 1: *A schematic view of the fusion model.*

abilities for each class can be specified and used in the classification process.

- *Temporal information.* Temporal information in the form of probabilities of changes with respect to the pattern classes between the acquisition of the different sources is included in the fusion model.
- *The interaction model.* The essential part of a fusion model is the interaction between the modules for the different sources. We will discuss the interaction model in the next section.

## 3 The interaction model

First, we will present a simple model for fusing Landsat TM, ERS-1 SAR images and GIS ground cover data. Then, we will extend this model to incorporate the temporal aspect. Let the Landsat image consist of $MN$ pixels or feature vectors $\{X_L(1,1), ..., X_L(M,N)\}$ and let the corresponding feature vectors in the ERS-1 SAR image be denoted as $\{X_S(1,1), ..., X_S(M,N)\}$. The classification of pixel $(i,j)$ in the SAR image is denoted $c_{S,ij}$ and in the Landsat image as $c_{L,ij}$. Let $P_S(X_S(i,j) \mid c_{S,ij} = c)$ and $P_L(X_L(i,j) \mid c_{L,ij} = c)$ denote the image formation model for SAR and Landsat, respectively. The probabilities $P_S(X_S(i,j) \mid c_{S,ij} = c)$ and $P_L(X_L(i,j) \mid c_{L,ij} = c)$ are computed using Haslett's contextual rule in the "cross" neighborhood of $X_n(i,j)$. For the GIS data, each pixel is assigned to the land-use category $c_{G,ij}$.

Let us assume that the measurements from each source are independent. We can then write down the expression for the a posteriori probabilities for each class at each pixel given the two spectral data sources:

$$P(c_{ij} = c \mid X_S(i,j), X_L(i,j)) = P_S(c_{S,ij} = c \mid X_S(i,j)) P_L(c_{L,ij} = c \mid X_L(i,j)) \quad (3)$$

The validity of the independence assumption between the measurements from each sensor is difficult to establish. In the case of radar images (SAR) and optical images (LANDSAT), the two sensors measure quite different characteristics in different wavelengths. If we do not assume independence, then we must specify the joint distributions of the measurements from the different sources. This is a very difficult task. The robustness of our fusion scheme will be determined based on its performance even when this independence assumption is not true.

The simplest fusion algorithm would consist of assigning each pixel to the class that maximizes Eq. (3). Next, we introduce the GIS data and assign reliability factors to each data source. The modified version of Eq. (3) becomes in logarithmic form:

$$\begin{aligned} \log P(c_{ij} = c \mid X_S(i,j), X_L(i,j)) = \\ \alpha_S \log P_S(c_{S,ij} = c \mid X_S(i,j)) \\ +\alpha_L \log P_L(c_{L,ij} = c \mid X_L(i,j)) \\ +\alpha_G \delta(c, c_{G,ij}), \end{aligned} \quad (4)$$

where $\alpha_i$, $0 < \alpha_i < 1$, $i \in \{S, L, G\}$ are the reliability factors associated with the different sources. $S$, $L$, and $G$ represents the sources SAR, Landsat and GIS, respectively. Let $\delta(k,l) = 1$ if $(k = l)$, and zero otherwise. With this simple fusion model, consistency between the sources is encouraged. We use the overall classification accuracy as the value of the reliability factors for Landsat TM and SAR ($\alpha_S$ and $\alpha_L$), respectively. We will later refer to Eq. (4) as the simple fusion model.

To incorporate possible changes with respect to the pattern classes in the model, we need to consider possibly different classes for the different sources. This requires constraints to control the behavior of the fusion algorithm to create a consensus interpretation of the scene. Assume that our classification modules for SAR images and Landsat images assign a pixel $(i,j)$ to classes $c$ and $l$, respectively, and that the GIS category for the pixel is $g$. Let us introduce penalty terms in the likelihood function whenever the different sources do not agree on their choice of classes. Extending the basic model in Eq. (3) to incorporate the constraints the modified logarithmic likelihood function becomes:

$$\begin{aligned} P(c,l) = \alpha_S \log P_S(c_{S,ij} = c \mid X_S(i,j)) \\ +\alpha_L \log P_L(c_{L,ij} = l \mid X_L(i,j)) \\ -\beta V(c_{S,ij}, c_{L,ij}, c_{G,ij}), \end{aligned} \quad (5)$$

where $V(c_{S,ij}, c_{L,ij}, c_{G,ij})$ denotes the penalty associated with individual pairs of sources and their classification results, and $\beta$ is a constant. In the following, we will refer to Eq. (5) as the extended fusion model. Experiments have shown that $\beta = 0.5$ is a reasonable choice. The parameters $\alpha_S$ and $\alpha_L$ are set equal to the overall classification accuracy based on SAR and Landsat alone, respectively. The parameter $\alpha_G$ must be specified by the user. We consider the following form for the penalty function:

$$V(c_{S,ij}, c_{L,ij}, c_{G,ij}) = \gamma_{SL}(1 - \delta(c,l))$$

$$+\gamma_{SG}(1-\delta(c,g))+\gamma_{LG}(1-\delta(l,g)) \quad (6)$$

where

$$\delta(c,l)=\begin{cases}1 & \text{if } c=l\\ 0 & \text{otherwise}\end{cases} \quad (7)$$

No penalty is assigned when the sources assign a pixel to the same class ($c=l$). If $c\neq l$, then the size of the penalty is determined by $\gamma_{cl}$. We relate the $\gamma_{cl}$ parameters to the a priori probability of an actual class transition from class $c$ to class $l$. Let $\phi(c,l)$ denote the a priori probability for an actual transition from class $l$ to class $c$ during the time between the acquisition of the two sources. Now, let

$$\gamma_{SL}=(1-\phi(c,l))(\theta_{S,ij}+\theta_{L,ij}), \quad (8)$$
$$\gamma_{SG}=(1-\phi(c,g))(\theta_{S,ij}+\alpha_G),$$
$$\gamma_{LG}=(1-\phi(l,g))(\theta_{L,ij}+\alpha_G),$$

where $\theta_{n,ij}$ represents a local reliability factor for pixel $(i,j)$ in the image. The specification of the $\theta_{n,ij}$ parameters will be explained shortly. The effect of this is that if the probability of an actual change from class $l$ to class $c$ between the acquisition of two different sources is high, then the associated penalty when the source-specific classifiers choose the two different classes $c$ and $l$ is low. If the likelihood is low, then the penalty is higher.

Let the local reliability factors $\theta_{c,ij}$ indicate the form of the probability function for the a posteriori probabilities for class $c$ at pixel $(i,j)$ for source $n$. Let

$$\theta_{c,ij}=(P_c(max)-P_c(c_{n,ij})) \quad (9)$$

where $P_c(max)$ is the maximum a posteriori probability among the different classes. $\theta_{c,ij}$ is now a measure of the uncertainty or doubt associated with the a posteriori probabilities.

## 4 Experimental results

The performance of the fusion model was investigated on a data set consisting of one Landsat TM image from Oct. 20, 1991, four ERS-1 SAR images from Aug. 27, Oct. 17, Oct. 20, and Nov. 19, 1991, and ground cover data from a topographic map in the series M711 of Kjeller, Norway. A five-class classification problem is considered, with the following classes: water, urban areas, forests, and two classes of agricultural areas: plowed and unplowed. Between the acquisition of the different images, a large portion of the agricultural fields was tilled, and ground control measurements for specific fields are available for training and testing the classifier's ability to detect changes.

The ERS-1 images are 3-look images processed at Tromsø Satellite Station. The SAR images have been resampled to a pixel size of $30m\times 30m$ and co-registered with the Landsat image. GIS data corresponding to the land-use categories water, urban areas, agricultural areas, and forests have been digitized and co-registered with the other images. The GIS data originated from two map revisions each covering only part of the area. The original maps were revised in 1976 and 1988, respectively. This means that the GIS data for part of the scene are inaccurate with respect to changes in area use after 1976. During this period, some areas have changed from forest/agricultural areas to urban areas.

As a reference for evaluating the performance of the multisource classifier, we will use the single-source classification error rates. These are given in *Table 1*. The large variations in the SAR classification error rates are due to the different weather conditions at the time the images were acquired. The typical overall classification accuracy is 95% for Landsat and 65-70% for SAR.

*Table 2* shows the error rate for the various approaches to fusion of the different sources. On an average, fusion of the Landsat TM image with the SAR image reduced the error rate from 33% based on SAR alone to 7%. Inclusion of GIS data in the model further reduced the error rate by 3.1% on an average. Using the simple fusion model, the average error rate was 7.2%. By using the extended fusion model, this error rate was reduced to 4.6%. To further illustrate the effect of the extended model, we have used test regions where the ground cover has changed from forest/agricultural to urban areas after the production of the topographic map. The error rates for these regions are shown in *Table 3*. By using the extended fusion model, the error rate was reduced from $60.0\%\pm 3.2\%$ to $36.4\%\pm 7.4\%$.

## 5 Discussion and conclusions

In this paper, we have presented a method for fusion of Landsat TM, ERS-1 SAR images, and GIS ground cover data for land-use classification. The method can be used to fuse images captured at different dates, by allowing changes in the pattern classes. By combining spectral imagery with GIS ground cover data, a more accurate interpretation of the scene can be obtained. The expected improvements in the classifica-

tion accuracy due to inclusion of data from additional sources depend on the general discrimination ability of the source. Even by fusing SAR images containing a relatively low discrimination information, and using partly inaccurate GIS ground cover data, the fusion resulted in significant improvements in the classification accuracy. Our results are in accordance with other experiments of including GIS data in the classification process [1, 9].

The method should be further evaluated on a larger data set. Future work will include methods for estimation of the model parameters.

## References


[1] J. A. Benediktsson and P. H. Swain, "A Method of Statistical Multisource Classification with a Mechanism to Weight the Influence of the Data Sources", *IEEE Symp. Geosc. Rem. Sens. (IGARSS)*, Vancouver, Canada, pp. 517-520, July 1989.

[2] J. A. Benediktsson and P. H. Swain, "Consensus Theoretic Classification Methods", *IEEE Trans. on Sys. Man and Cyber.*, vol. 22, pp. 688-704, 1992.

[3] H. Derin, P. A. Kelly, G. Vezina and S. G. Labitt, "Modeling and segmentation of speckled images using complex data", *IEEE Trans. Geosc. Remote Sensing,* vol. 28, pp. 76-87, 1990.

[4] R. T. Frankot and R. Chellappa, "Lognormal random-field models and their applications to radar image synthesis", *IEEE Trans. Geosc. Remote Sensing,* vol. 25, pp. 195-206, 1987.

[5] J. Haslett, "Maximum likelihood discriminant analysis on the plane using a Markovian model of spatial context", *Pattern Recognition,* vol. 18, pp. 287-296, 1985.

[6] T. Lee, J. A. Richards and P. H. Swain, "Probabilistic and evidential approaches for multisource data analysis", *IEEE Trans. Geosc. Remote Sensing*, vol. 25, pp. 283-293, 1987.

[7] Q. Lin and J. P. Allebach, "Combating speckle in SAR images: vector filtering and sequential classification based on a multiplicative noise model", *IEEE Trans. Geosc. Remote Sensing,* vol. 28, pp. 647-653, 1990.

[8] A. L. Maffett and C. C. Wackerman, "The modified Beta density function as a model for synthetic aperture radar clutter statistics", *IEEE Trans. Geosc. Remote Sensing,* vol. 29, pp. 277-283, 1991.

[9] J. Middelkoop and L. L. F. Janssen, "Implementation of temporal relationships in knowledge based classification of satellite images", *Photogrammetric Engineering & Remote Sensing*, vol. 57, pp. 937-945, 1991.

[10] E. Rignot and R. Chellappa, "Segmentation of synthetic-aperture-radar complex data", *J. Opt. Soc. Am. A,* vol. 8, pp. 1499-1509, 1991.

[11] A. H. Schistad and A. K. Jain, "Texture analysis in the presence of speckle noise", *IEEE Symp. Geosc. Rem. Sens. (IGARSS)*, Houston, Texas, pp. 884-886, May 1992.

[12] A. Schistad Solberg, A. K. Jain, and T. Taxt, "Multisource classification of remotely sensed data: fusion of Landsat TM and SAR images", Submitted to *IEEE Trans. Geosc. Rem. Sens.*



**Acknowledgement**

This work was supported by a grant from the Royal Norwegian Council for Scientific and Industrial Research.

| Landsat, Oct. 20 | SAR, Aug. 27 | SAR, Oct. 17 | SAR, Oct. 20 | SAR, Nov. 19 |
|---|---|---|---|---|
| 5.2% | 33.4% | 32.5% | 35.2% | 21.7% |

Table 1: *Classification error rates for the single-source classifiers.*

| Image fused | Fusion method 1 | Fusion method 2 |
|---|---|---|
| TM and SAR Oct.20 | $6.3 \pm 0.3$ % | - |
| TM, SAR Oct.20, and GIS | $2.9 \pm 0.2$% | - |
| TM and SAR Aug.27 | $11.4 \pm 0.5$% | $7.5 \pm 0.4$% |
| TM, SAR Aug.27, and GIS | $7.3 \pm 0.2$% | $5.1 \pm 0.3$% |
| TM and SAR Oct.17 | $9.1 \pm 0.5$% | $8.1 \pm 0.4$% |
| TM, SAR Oct.17, and GIS | $6.3 \pm 0.4$% | $4.8 \pm 0.3$% |
| TM and SAR Nov.19 | $11.5 \pm 0.3$% | $6.3 \pm 0.2$% |
| TM, SAR Nov.19, and GIS | $8.1 \pm 0.3$% | $3.9 \pm 0.2$% |

Table 2: *Performance of fusion models.*
Average classification error rates (in %) for fusion of the Landsat TM image from Oct. 20, 1991 with the SAR images from Aug. 27, Oct. 17, Oct. 20, and Nov. 19, 1991, and the GIS data. The error rates are averages over 10 experiments. The fusion methods are the following: Method 1 = Simple fusion with reliability factors, Method 2 = Extended fusion model with reliability factors and transition probabilities. Transition probabilities used are 0.4 from the class unplowed to the class plowed for the spectral data, and from forests/agricultural areas to urban areas for the GIS data. Reliability factors are $\alpha_S = 0.7$, $\alpha_L = 0.95$, and $\alpha_G = 0.7$.

| Date | Error rate for simple fusion | Error rate for extended fusion |
|---|---|---|
| Aug.27 | 61.0 % | 29.7% |
| Oct.17 | 58.7% | 30.6% |
| Oct.20 | 63.7% | 40.2% |
| Nov.19 | 56.3% | 44.9% |

Table 3: *Performance for detection of changes.*
Classification error rates for the simple and the extended fusion model for test regions where the ground cover has changed from forests/agricultural to urban areas. Transitions probabilities of 0.4 from the categories forests/agricultural to urban areas are used.