Improving Spatial and Temporal Consistency in Environmental Classification

Jonas Callmer, Fredrik Lindsten, and Fredrik Heintz
{callmer,lindsten}@isy.liu.se,frehe@ida.liu.se

Summary

We present an approach to improving image stream classification by improving temporal and spatial consistency. Temporal consistency is improved by generating a prior from previous classifications and spatial consistency is improved by reasoning about the local classification using Markov Logic Networks.

Introduction

To segment and classify images in a video stream into semantically meaningful regions is important for many applications. It could for example be used to improve map making and navigation capabilities of a UAV. There exists many good methods for classifying images using local features. However, classifying images taking global features such as temporal and spatial consistency into account is still an open problem. To address this problem, we suggest a three step classification procedure:

1. Each image is segmented into uniform regions and each region is classified independently.
2. Spatial consistency is improved through logical reasoning.
3. Temporal consistency is improved by taking previous classifications into consideration.

Local Classification

The segmentation results in a set of small uniform regions called *superpixels*. These are classified locally without concern of neighboring regions by a neural network as either grass, road or house. The local classification is based on texture and color cues. However, the local features do not always contain enough information for the classification to be correct.

Spatial Consistency

Better spatial consistency is achieved through reasoning about the local classification using an induced Markov Logic Network (MLN). MLN combines first-order logic and probabilistic graphical models in a single representation where formulas have a weight which can be interpreted as probabilities.

From a locally classified image a MLN theory is derived. For each region the probability distribution over the classes and whether the region is connected or surrounded is encoded:

$-1.637541\, cls(\mathsf{R6}, \mathsf{G})$ (corresponds to 19.4%)
$-0.488074\, cls(\mathsf{R6}, \mathsf{R})$ (corresponds to 61.4%)
$-1.651641\, cls(\mathsf{R6}, \mathsf{H})$ (corresponds to 19.2%)
$\forall c\, con(\mathsf{R6}, c) \leftrightarrow cls(\mathsf{R1}, c) \vee cls(\mathsf{R8}, c)$.
$\forall c\, sur(\mathsf{R6}, c) \leftrightarrow cls(\mathsf{R1}, c) \wedge cls(\mathsf{R8}, c) \wedge cls(\mathsf{R11}, c)$.

The theory also contains a number of rules such as "a house region is either large or connected to a large house region" and "a small surrounded region has the same class as the surrounding region".

$\forall r\, cls(r, \mathsf{H}) \rightarrow \neg\, small(r) \vee con(r, \mathsf{H})$.
$\forall r, c\, sur(r, c) \wedge small(r) \rightarrow cls(r, c)$.

Temporal Consistency

Temporal consistency is achieved by using the previously classified image as a prior when classifying the current image. Point feature matching is used to calculate the transformation between the two images. With this information a class prior is calculated for each pixel in the new image which is merged with the local class estimate. This removes local, temporary errors introduced by the local classification.

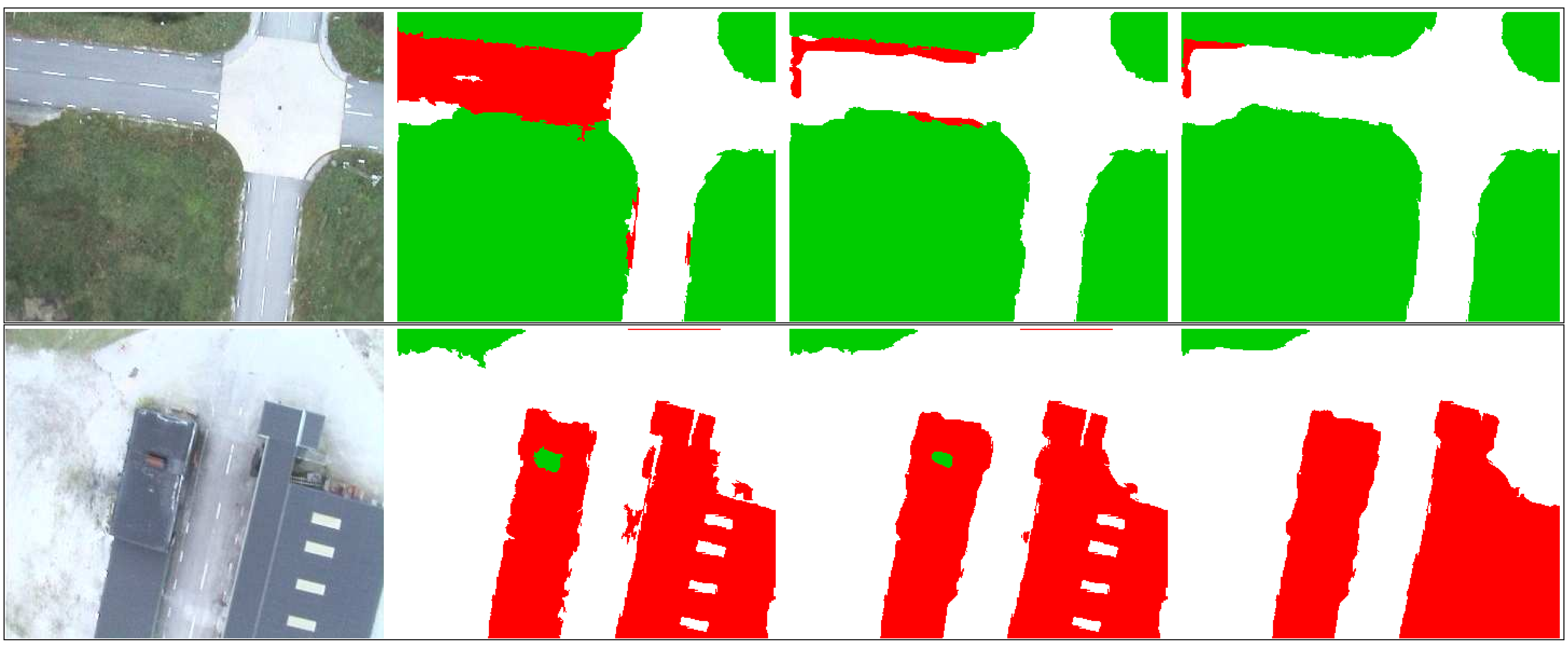

Fig. 1 From left to right; on-board image, locally classified (LC) image, LC with prior, LC with prior and MLN.

Supported by grants from the SSF Strategic Research Center MOVIII and the VR Linnaeus Center CADICS.