EDGE PRESERVING LAND COVER CLASSIFICATION REFINEMENT USING MEAN SHIFT SEGMENTATION

Torsten Büschenfeld and Jörn Ostermann

Institut für Informationsverarbeitung
Leibniz Universität Hannover
Appelstraße 9A, 30167 Hannover, DE
bfeld@tnt.uni-hannover.de, ostermann@tnt.uni-hannover.de
http://www.tnt.uni-hannover.de/

KEY WORDS: Classification, Imagery, Land Cover, Satellite, Segmentation

ABSTRACT:

Recently, a lot of classification methods within GIS applications rely on object-based image analysis. The image is partitioned by a low-level segmentation method, and classification is processed on each segment individually. Often, the resulting homogeneous segments do not allow for a correct classification. Hence, a semantic processing considering neighbouring segments is needed to correctly classify the segments. Therefore, we use a pixel-wise classification in different scales which ensures independent consideration of context beyond segment boundaries. However, local context introduces a smoothing to the classification process and is restricted to a small local neighbourhood. To reduce smoothing effects and keep spatial coherence, we propose a method that incorporates a mean shift segmentation on the input image with tendency towards over-segmentation. Each unique segment is analysed for enclosed classes from pixel-wise classification. A weighted majority vote decides for the resulting class which is assigned to the whole segment. Our experiments show that the refinement improves the classification results in all test sets. An average improvement of 4.51 % is achieved with improvements from 66.75. % to 71.40 % and 81.58 % to 83.97 %. This includes a better representation of image edges, since high frequency contours are restored. We also show that small context size adds noise to classification results which is shown to be significantly reduced by our approach.

1 INTRODUCTION

Today, the increasing amount of image data originating from sensors like satellites is employed for several applications in GIS systems. Evaluation, however, often demands more manpower than available. Hence, (semi-)automatic systems based on computer vision and machine learning algorithms are of great interest with respect to these applications (Förstner, 2009). With regard to this, methods involving pixel-wise and object-wise classification as well as segmentation are proposed for land cover classification (Weis et al., 2005),(Helmholz et al., 2010), (Vogt et al., 2010). A comprehensive review of (Mountrakis et al., 2011) shows that a lot of research has been done in the area of support vector machine classification (SVM),(Vapnik, 1998), recently.
However, simultaneous classification of diverse homogeneous regions like grassland, fine grained textures such as forests, and large structures like those in industrial areas still challenge modern methods. Due to highly varying scales, the extent of local context for feature extraction is crucial. Since large structures require a large context while smaller context improves accuracy and classification of smaller structures, parameter selection is a good compromise at most. Context in high-dimensional feature space accounts for good classification results, while losing spatial coherence. On the other hand, low-level segmentation methods like mean shift segmentation (Comaniciu et al., 2002) preserve discontinuities (i. e. edges) and spatial coherence, while lacking textural information for classification without a priori knowledge. In recent literature, these opposing properties are approached in two steps. The image is partitioned by a low-level segmentation method. Then, each segment is classified individually by a high-level classification method. (Lin, 2008) states that mean shift segmentation is generally well suited for classification of land cover data, but he does not deal with the crucial aspect of classification in detail. Since segments from low-level segmentation do not exhibit a lot of textural information in most cases, applications are limited to certain scenarios or a priori knowledge. In (Liu et al., 2010) classification is limited to urban areas, which show many homogeneous rooftops. The texture-based classification of the segments leads to low-level classes like 'grey roof' and 'white roof', only. If a more general representation is targeted, a semantic (post-)processing considering neighbouring segments is needed. (Yang and Förstner, 2011) present such a system in a well defined scenario for building facade classification. They employ a hierarchical segmentation where context and semantics are jointly incorporated by a Conditional Random Field model. (Moser and Serpico, 2009) focus on urban areas in satellite imagery. A Canny edge map is integrated to their Markov Random Field classification in order to preserve edges.
In this paper, we propose a method that employs high-level pixel-wise classification and low-level segmentation in parallel. A pixel-wise SVM classification ensures context consideration beyond segment borders in different scale levels. To compensate lowered accuracy due to smoothing from larger context, the classification result is refined utilising a discontinuity preserving segmentation like mean shift, which is applied on the input image data. For each of its segments, the dominant class label is determined and applied to all enclosed pixels. Hence, edges are preserved while keeping spatial coherence.
The next section first explains the base system for SVM classification and feature extraction. Followed by a description of the mean shift segmentation, the section concludes with the proposed joining scheme based on weighted majority voting. Finally, experimental results are discussed, which show the benefit of the proposed refinement.

2 EDGE-PRESERVING CLASSIFICATION REFINEMENT

The system for edge-preserving classification refinement as proposed in this paper consist of three parts, which are shown in

Figure 1. To keep coherence in feature space as well as spatial coherence, a pixel-wise SVM classification and a mean shift segmentation are parallelly computed on the input image bands. Both results are joined by a weighted majority vote. In the follow-

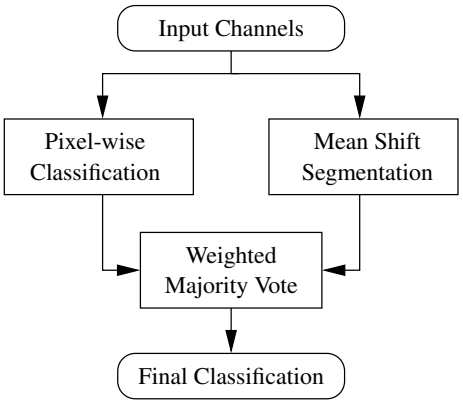


Figure 1: Proposed system for edge preserving classification refinement. Results from pixel-wise classification and mean shift segmentation are joined by a weighted majority voting.

ing subsections, we will first describe the pixel-wise classification followed by the mean shift segmentation. The section concludes with the final joining of results from both approaches.

2.1 Pixel-wise Classification

Our base system is a flexible framework for classification of satellite imagery. It consists of two main modules, the feature extraction and the SVM classification. The feature extraction pipeline is depicted in Figure 2. Aiming for a pixel-wise classification implies features to be extracted for each pixel location. Hence, for an arbitrary number of image bands, local features are extracted within an $N \times N$ neighbourhood.

Due to widely varying structure sizes, it is indispensable to incorporate the according local context information of likewise varying size. This is done by considering different scales, which is a well known technique that is approached in several different ways. Depending on the application, the neighbourhood's size can be increased, different bands in frequency domain can be analysed (as done with SIFT features (Lowe, 1999)) or resolution pyramids are build. We use a low-pass cascade of Gaussian filters which leads to a less complex model than solely increasing neighbourhood size of feature extraction. (Lindeberg, 1994) comprehensively evaluates scale-space, proving Gaussian filter kernels to be best suited in the general case. Although resolution pyramids are well suited for efficient calculation of scale invariant features, we do not sub-sample lower scales. The reason are inaccuracies of lower levels that propagate to the final classification, resulting in blocking artefacts.

The neighbourhood influence for different scale levels can be estimated as follows: Let σ_1 and σ_2 be the standard deviation of two Gaussian filters G_1 and G_2, respectively. The resulting standard deviation σ_r when convolving $G_1 * G_2$ is

$$\sigma_r = \sqrt{\sigma_1^2 + \sigma_2^2}. \tag{1}$$

With N_g being the number of Gaussian scale levels and σ defined as $\sigma_1 = \sigma_2$ the resulting standard deviation becomes

$$\sigma_r = \sigma\sqrt{N_g}. \tag{2}$$

For each additional scale level, the neighbourhood N is increased by a factor f_n. Therefore, the major neighbourhood influence N_i

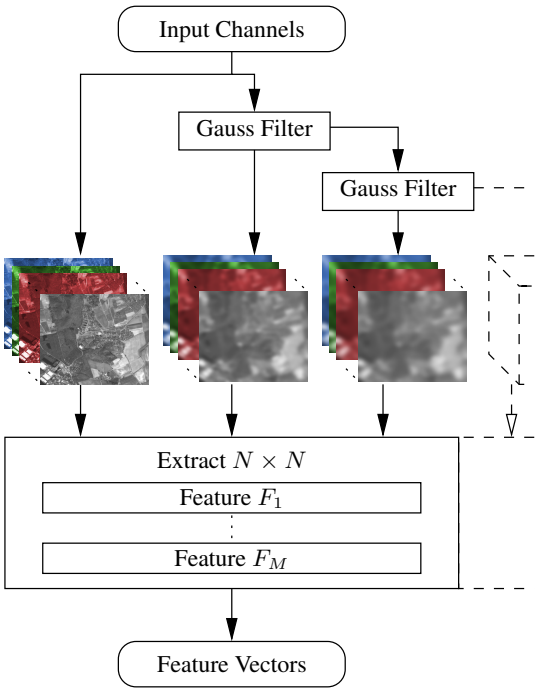


Figure 2: All image bands are analysed in different scales. For each pixel location, features $1 - M$ are extracted within an $N \times N$ neighbourhood which finally compose the feature vector.

for feature extraction can be approximated by

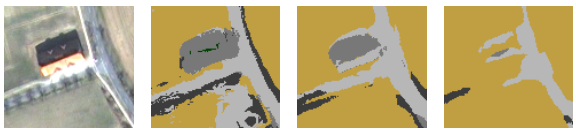
$$N_i = f_n^{N_g - 1} N + \sigma\sqrt{N_g}, \tag{3}$$

assuming the 1-Sigma interval.

Features like statistical or textural features are extracted from all scales of all image bands and compose the feature vector for each pixel location that is passed to SVM classification. The SVM is based on the implementation of (Chang and Lin, 2001) using the standard Gaussian RBF kernel.

Classification quality highly depends on local context. Adding lower scales, i. e. increasing context size, improves classification of larger structures and reduces noise. However, additional smoothing is introduced to classification results. This effect is exemplarily depicted in Figure 3.

It shows a small extract of the RGB bands from the IKONOS in-



(a) IKONOS (b) 1 level (c) 2 levels (d) 3 levels

Figure 3: Classification results (b)–(d) of IKONOS image (a) using a different number of scales. Ochre: cropland/grassland; green: forest; from light to dark grey: large, medium, small building structure.

put image (Figure 3(a)). From Figures 3(b) to 3(d) an increasing number of scale levels was used to classify the scene. Referring to Equation 3 with the 1-Sigma interval, areas with approximately 11 m (b), 28 m (c), and 43 m (d) edge length account for feature extraction of each pixel location.

Class labels encoded in colour values point out the smoothing characteristic. With a small local context classification is noisy, while it becomes more homogeneous with larger context. Con-

sequently, small regions, e. g. the house in the center, even show green forest labels for one scale level, but almost completely vanish in that they are smoothed out by surrounding regions.

2.2 Mean Shift Segmentation

Parallel to SVM classification, a discontinuity preserving mean shift segmentation is applied on the input image bands. Here, we chose the mean image of all input bands. This corresponds to a grey value of an RGB image, yet considering additional channels like near infra red.
The mean shift segmentation analyses each pixel within a joint spatial-range domain spanned by a spatial neighbourhood (spatial range s) and an intensity range (range r). The analysis window is shifted towards the mean of all enclosed feature points. This process is repeated until convergence and defines the final intensity value for the initially analysed pixel.
The spatial range s is set to match the spatial radius of the most detailed level of feature extraction ($N \times N$ neighbourhood in the original input image data). Intensity range r depends on the distribution of intensity values. To cope with different image characteristics, we adjust range r of the mean shift segmentation with respect to the image's histogram. Referring to pixel intensities in the interval $[0, I_{max}]$, 5 % of potential outliers next to both interval boundaries are disregarded. The adjusted range r' results in

$$r' = r\frac{I_{95} - I_5}{I_{max}} \tag{4}$$

with I_5 and I_{95} being the fifth percentile and the 95^{th} percentile, respectively.

2.3 Weighted Majority Voting

In order to use the capabilities of pixel-wise classification by well known classifiers like SVMs concerning feature space coherence, robustness, and generalisation while taking advantage of spatial coherence brought by segmentation methods, we introduce a scheme to join results from both, SVM classification and mean shift segmentation.
Figure 4 exemplarily depicts the mean shift segmented image (b) of the scene in (a). Since mean shift segmentation basically quantises pixel intensities, resulting segments do not necessarily represent a unique label. Therefore, a connected-components analysis is employed before further processing. The resulting segments serve as basis for the final refinement. For each segment, class labels from pixel-wise SVM classification (c) are determined. In the following, a weighted majority voting is presented that assigns the dominant class label to all enclosed pixels of a segment, resulting in (d).

(a) (b) (c) (d)

Figure 4: Rooftop next to grassland. (a) IKONOS (b) Mean shift segmentation (c) Classification result (d) Refined classification result. Colours for (c) and (d): ochre: cropland/grassland; from light to dark grey: large, medium, small building structure

In the preceding sections, pixel-wise classification in different scales was shown to cause reduced accuracy at image edges. Thus, for each position p, the class label is weighted with respect to its distance d from segment borders of mean shift segmentation with weight w_p and factor f to control distance influence. Additionally, probability estimates for SVMs (Lin et al., 2007) allow for calculation of classification confidence. For each pixel location p, the *first-best-to-second-best* ratio determines the certainty c_p of the classification result. This leads to

$$w_p = c_p \ln(f d_p). \tag{5}$$

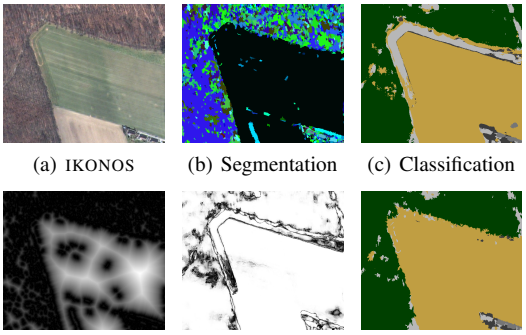
Within each segment S, the weights w_p with label l_p at position p are summed up for each unique classification label l:

$$W(l) = \sum_{\substack{p \in S \\ l_p = l}} w_p. \tag{6}$$

Finally, the resulting label l_r for a segment S is given by

$$l_r(S) = \arg\max_i (W(l = i)). \tag{7}$$

Intermediate steps of the weighted majority voting are depicted in Figure 5. The input image (a) is segmented (b) and classified (c). Distance shown in (d) and certainty c (e) define the weights w as stated in Equation 5 which results in the final classification (f).



(a) IKONOS (b) Segmentation (c) Classification

(d) Certainty (e) Distance (f) Final result

Figure 5: Intermediate steps of the weighted majority voting. Colours for (c) and (f): ochre: cropland/grassland; green: forest; from light to dark grey: large, medium, small building structure. Mean shift segmentation (b) was colour adjusted for better visibility of segments.

As can be seen, the misclassified grey region in (c) at the transition from forest to cropland is almost completely corrected in the refined result (f) due to its low weight that contributes to the segment.
In a worst case scenario a large image segment will be almost arbitrarily relabelled if the underlying classification result consists of two or more classes with a similar amount of spatial coverage. Thus, parameters tending towards over-segmentation are chosen to avoid extensive merging of distinct regions. The smaller the segments are, the lower their influence is in refinement. In exchange, results become more robust. This adds convenience to parameter adjustment.

3 RESULTS

For our tests, we use ortho-rectified images from the IKONOS satellite. Four spectral bands – red (R), green (G), blue (B), and near infra red (NIR) – offering 1 m spatial resolution and 8-bit radiometric resolution are available. The scenes cover areas from Hildesheim/Germany and Weiterstadt/Germany. Validation sets originate from both areas and were manually labelled with pixel accuracy.

For training, a collection of different sets was used to show robustness in results. This includes sets of manually selected samples as well as automatically extracted sets from the German ATKIS[1]. Training sets are listed in Table 1.

Depending on the processed scene, the sets of four (LCC4) to six (LCC6) land cover classes were trained:

- small building structure (LCC6)
- small and medium building structure (LCC4)
- medium building structure (LCC6)
- large building structure (LCC4, LCC6)
- cropland/grassland (LCC4, LCC6)
- trees/bushes (LCC4, LCC6)
- water (LCC6)

Parameters for the classification process were chosen with respect to robustness for the processed scenes. To cover all significant structures without losing too much detail, the neighbourhood size for feature extraction was set to $N = 11$, corresponding to an area of $11 \times 11\,\mathrm{m}^2$. One additional Gaussian filtered scale level was considered with $\sigma^2 = 10$. The neighbourhood size for this level was slightly increased by factor $f_n = 1.25$. This results in an area of $13.75 \times 13.75\,\mathrm{m}^2$. According to Equation 3, major influence N_i for feature extraction originates from areas with approximately 20 m edge length.

For our scenes, two basic features – median and variance – are used for classification. Additional features like the common Haralick's textural features (Haralick et al., 1973) did not improve the classification result concerning our test scenarios. All results were evaluated by pixel-wise comparison to the manually referenced scenes. The results for training sets T01 – T11 are given in Table 2. The average relative improvement for different voting strategies is listed Table 3. The results clearly show an improvement for all sets with an average improvement of 4.51 %.

As can be seen, the major contribution for refinement comes from certainty information (ms_c). The distance (ms_dist) does not significantly increase the classification result if used without certainty information. However, in combination (ms_dist_c) there is an additional improvement. This emphasises the importance of certainty information. Weighted by distance information, classification results near the center of a segment are more reliable only as long as their certainty is high.

3.1 Further Discussion

Figure 6 shows a part of an IKONOS scene from Hildesheim/Germany and results from classification and refinement. Figure 6(b) demonstrates the characteristics of the refinement in (d) based on classification in (c). It shows errors that were removed (green) as well as new errors that occur due to refinement (red). In large, homogeneous regions like those in the center of the image, the approach clearly benefits from mean shift segmentation in that spatial coherence is kept and segment boundaries align to image edges. In fine structured regions, however, there is no significant difference in the number of new errors and removed errors. For instance, this is noticeable in the top right corner of the image where a lot of small building structure is located.

The noise reduction capabilities of the mean shift refinement are shown in (e) and (f). Based on set T06, neighbourhood size was set to 9 m × 9 m without any lower scale incorporated. The noise can be significantly reduced by the proposed refinement as seen in Figure 6(f). This is backed up by overall classification correctness, too, increasing from 75.39 % to 78.18 % for this configuration.

[1] Amtlich topographisch-kartographisches Informationssystem (Authoritative Topographic Cartographic Information System)

Set	Scene	Nr. of Classes	Sample Selection
T01 – T03	Weiterstadt	4	manual
T04 – T05	Weiterstadt	4	GIS
T06 – T07	Hildesheim	6	manual
T08 – T11	Hildesheim	4	manual

Table 1: Training data sets from Weiterstadt/Germany and Hildesheim/Germany. Sample selection for two scenes was automatically done using a GIS.

Set	SVM [%]	ms [%]	ms_dist [%]	ms_c [%]	ms_dist_c [%]
T01	**66.75**	67.82	68.01	70.57	**71.40**
T02	**71.33**	73.08	73.02	73.74	**73.53**
T03	**71.19**	73.76	73.66	73.83	**73.89**
T04	**61.63**	63.55	63.71	63.90	**64.35**
T05	**61.90**	65.02	64.60	65.43	**65.61**
T06	**77.24**	80.14	79.96	80.45	**80.37**
T07	**67.25**	70.86	70.73	71.22	**70.99**
T08	**75.82**	78.56	78.55	78.55	**78.94**
T09	**81.58**	83.82	83.99	83.82	**83.97**
T10	**65.10**	67.00	67.10	67.44	**68.43**
T11	**74.33**	77.37	77.69	77.25	**77.53**
Avg	**70.37**	72.82	72.82	73.29	**73.55**

Table 2: Classification results from SVM of validation set for training data sets T01 – T11, see Table 1 for reference. Refined results for voting strategies ms (no weighting, $w_p = 1$), ms_c (only certainty, $w_p = c_p$), ms_dist (only distance, $w_p = ln(fd_p)$), ms_dist_c (full weighting, $w_p = c_p ln(fd_p)$)

ms [%]	ms_dist [%]	ms_c [%]	ms_dist_c [%]
3.47	3.48	4.15	4.51

Table 3: Average relative improvement for voting strategies given in Table 2.

4 CONCLUSIONS

The approach proposed in this paper combines pixel-wise SVM classification with mean shift segmentation to improve the overall classification result. To join the results from both, classification and segmentation, we introduced a weighted majority voting that takes distance and classification confidence into account. We have shown that our approach aligns class labels to image edges to cope with the smoothing, which typically originates from classification of image content with varying scales. Results were presented showing that refinement leads to higher detection rates with 4.51 % relative improvement on average. Additionally, it was evaluated that a smaller neighbourhood in feature extraction reduces smoothed classification results, but also adds noise, which is shown significantly reduced by our approach.

REFERENCES

Chang, C. C. and Lin, C. J., 2001. LIBSVM: a library for support vector machines.

Comaniciu, D., Meer, P. and Member, S., 2002. Mean shift: A robust approach toward feature space analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 24, pp. 603–619.

Förstner, W., 2009. Computer vision and remote sensing - lessons learned. In: Photogrammetric Week 2009, pp. 241–249.

Haralick, R. M., Shanmugam, K. and Dinstein, I., 1973. Textural features for image classification. IEEE Transactions on Systems, Man, and Cybernetics 3(6), pp. 610–621.

Helmholz, P., Becker, C., Breitkopf, U., Büschenfeld, T., Busch, A., Grünreich, D., Heipke, C., Müller, S., Ostermann, J., Pahl, M., Vogt, K. and Ziems, M., 2010. Semiautomatic quality control of topographic reference datasets. In: ISPRS Commission 4 Symposium.

Lin, H., 2008. Method of image segmentation on high-resolution image and classification for land covers. In: Natural Computation, 2008. ICNC '08. Fourth International Conference on, Vol. 5, pp. 563 –566.

Lin, H.-T., Lin, C.-J. and Weng, R. C., 2007. A note on platt's probabilistic outputs for support vector machines. Machine Learning 68(3), pp. 267–276.

Lindeberg, T., 1994. Scale-space theory: A basic tool for analysing structures at different scales. Journal of Applied Statistics pp. 224–270.

Liu, Y., Cai, W., Li, M., Hu, W. and Wang, Y., 2010. Multi-scale urban land cover extraction based on object oriented analysis. In: Geoinformatics, 2010 18th International Conference on, pp. 1–5.

Lowe, D., 1999. Object recognition from local scale-invariant features. In: Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on, Vol. 2, pp. 1150 – 1157 vol.2.

Moser, G. and Serpico, S., 2009. Edge-preserving classification of high-resolution remote-sensing images by markovian data fusion. In: Geoscience and Remote Sensing Symposium,2009 IEEE International,IGARSS 2009, Vol. 4, pp. IV–765 –IV–768.

Mountrakis, G., Im, J. and Ogole, C., 2011. Support vector machines in remote sensing: A review. ISPRS Journal of Photogrammetry and Remote Sensing 66(3), pp. 247–259.

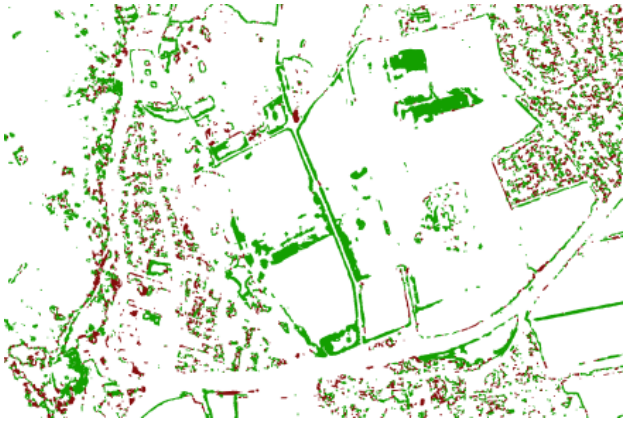
Vapnik, V. N., 1998. Statistical learning theory. 1 edn, Wiley.

Vogt, K., Scheuermann, B., Becker, C., Büschenfeld, T., Rosenhahn, B. and Ostermann, J., 2010. Automated extraction of plantations from ikonos satellite imagery using a level set based segmentation method. In: ISPRS Technical Commission VII Symposium, Vol. 38, pp. 275–280.

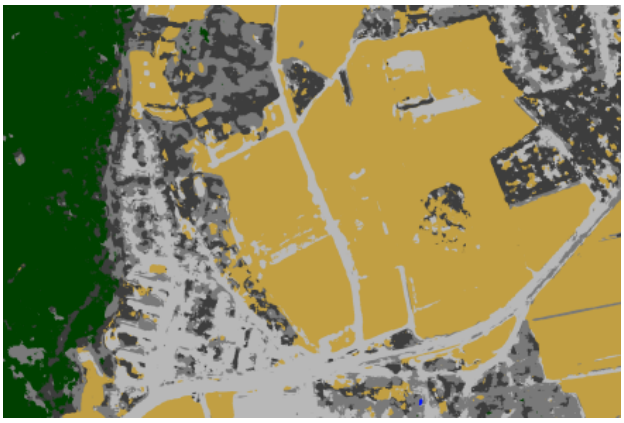
Weis, M., Müller, S., Liedtke, C.-E. and Pahl, M., 2005. A framework for gis and imagery data fusion in support of cartographic updating. Information Fusion 6(4), pp. 311–317.

Yang, M. Y. and Förstner, W., 2011. Regionwise classification of building facade images. In: Photogrammetric Image Analysis (PIA2011), LNCS 6952, Springer, pp. 209 – 220.

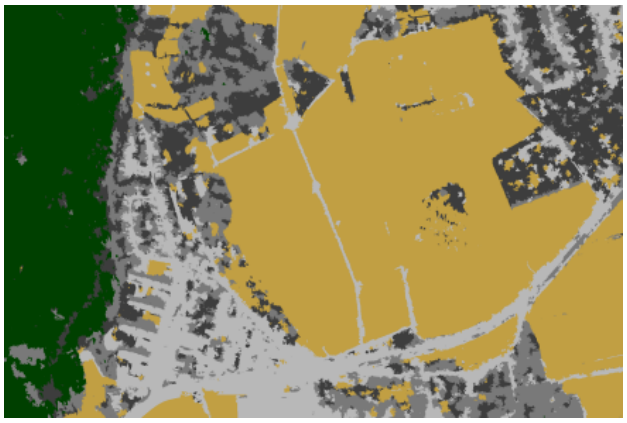
(a) IKONOS image, Hildesheim/Germany



(b) Errors from (c) to (d), green: removed, red: new



(c) Classification result, large context



(d) Mean shift refined classification result, large context

(e) Classification result, large context

(f) Mean shift refined classification result, small context

Figure 6: Scene from Hildesheim/Germany and results. Colours for (c) to (f): ochre: cropland/grassland; green: forest; from light to dark grey: large, medium, small building structure; blue: water.