

SELECTING APPROPRIATE FEATURES FOR DETECTING BUILDINGS AND BUILDING PARTS

Martin Drauschke, Wolfgang Förstner

University of Bonn, Department of Photogrammetry
Nussallee 15, 53115 Bonn, Germany
martin.drauschke@uni-bonn.de, wf@ipb.uni-bonn.de

KEY WORDS: Building Detection, Classification, Feature Selection, Stable Regions, Adaboost

ABSTRACT:

The paper addresses the problem of feature selection during classification of image regions within the context of interpreting images showing highly structured objects such as buildings. We present a feature selection scheme that is connected with the classification framework Adaboost, cf. (Schapire and Singer, 1999). We constricted our weak learners on threshold classification on a single feature. Our experiments showed that the classification with Adaboost is based on relatively small subsets of features. Thus, we are able to find sets of appropriate features. We present our results on manually annotated and automatically segmented regions from facade images of the eTRIMS data base, where our focus were the object classes *facade*, *roof*, *windows* and *window panes*.

1 INTRODUCTION

The paper addresses the problem of feature selection during classification of image regions within the context of interpreting images showing highly structured objects such as buildings. Our region classification is meant to be used as an initial interpretation which can be inspected by a high-level system, cf. (Hartz and Neumann, 2007), leading to new hypotheses to be verified by new image evidence.

The diversity of buildings and their environment is too rich for the determination of appropriate region features by only a few features, such as the color channels themselves, gradient images or texture measures. This would lead to unsatisfying results during classification of building parts. If we, as a remedy, significantly expand the dimension of the feature vector, we need to select appropriate features for efficiency reasons.

This paper addresses the problem of feature selection. We present a feature selection method based on Adaboost and investigate its performance with respect to classification of image regions which show buildings or building parts. In the following we first relate our work to previously proposed methods and use it to motivate our approach. Then we introduce our technique and describe our experiments with which we want to show the capability of our approach.

2 RELATED WORKS

Building extraction is an active research area in photogrammetry. Concerning building extraction from *aerial images*, the review in (Mayer, 1999) presents well discussed approaches. Many of these methods work on extracted image edges to identify buildings, e. g. (Nevatia et al., 1997), or combine detected image edges and detected homogeneous image regions in a hierarchical aggregation process, e. g.

(Fischer et al., 1998). Regrettably, these approaches often fail as soon as the building structures become too complex. The interpretation of *terrestrial images* of buildings is not developed as far as the interpretation of aerial images. Some approaches only investigate the grouping of repetitive textures and structures, cf. (Wang et al., 2002) and (Tuytelaars et al., 2003).

Building detection is also a very active research area in photogrammetry and computer vision. In recent approaches, graphical models are often used for integrating further information about the content of the whole scene, cf. (Kumar and Hebert, 2003) and (Verbeek and Triggs, 2007). In another paradigm, the bag of words, objects are detected by the evaluation of histograms of basic image features from a dictionary, cf. (Sivic et al., 2005). Unfortunately, both approaches have not been tested with high resolution building images. Furthermore, the bag of words approaches have not applied to multifarious categories as building or mammals. Epshtein and Ullman (2005) and Lifschitz (2005) propose a hierarchical interpretation scheme and show first results on very small and strongly smoothed building images. Thus, their concept is also not applicable on detailed building scenes. In photogrammetry, the interpretation of building facades is currently studied by Reznik and Mayer (2007), where they make use of image sequences.

Feature Selection is used in data mining to extract useful and comprehensible information from data, cf. (Liu and Motoda, 1998). In our case, we want to select the most valuable features from a set of candidates to keep the classification efficient and reliable. One classical approach, the principal component analysis, reduces the dimension of the feature space by projecting all features, cf. (Bishop, 2006). Thus, the obtained new features are not a subset of all candidate features, but combinations of the original features.

The exhaustive evaluation of all 2^D combinations of D features takes too much time. Therefore, we may choose be-

tween random or greedily deterministic approaches to find an appropriate subset of features to perform satisfying classifications.

Alternatively, Guyon and Elisseeff (2003) propose a feature selection framework which is based on the ranking of features. Then, the classification step itself is done by only one classifier. In our experiments, the features have low correlation coefficients with the class target and they are highly correlated with each other. Furthermore, the training samples do not form compact clusters in feature space. Then, it is a hard task to find a single classifier that is able to separate the classes. Thus, we prefer a feature selection scheme which combines several classifiers.

One of such schemes is the framework of adaptive boosting (Adaboost), cf. (Schapire, 1990). The first weak classifier (or best hypothesis) is learnt on equally weighted training samples (x_n, y_n) . Then, the influence of all misclassified samples is increased by adjusting the weights of the feature vectors before training the second weak classifier. So, the second weak classifier will focus especially on the previously misclassified samples. Again, the weights are adjusted once more depending on the classification result of the second weak classifier before training the third weak classifier, and so on. Finally, we obtain the resulting classifier by a majority vote of all weak classifiers. The discriminative power of this resulting classifier is much higher than the discriminative power of each weak classifier, cf. (Rätsch et al., 2001).

This majority voting is also done when using random forests, which are based on random decisions when constructing decision trees. In (Ho, 1998), a random selection of a predefined small number of features is used for constructing decision trees. This procedure is equivalent to projecting the feature vectors into a space with much lower dimensionality, but the projections differ from one decision tree to the others. The decision trees in (Breiman, 2001) randomly choose a feature from the whole feature set and takes it for determining the best domain split. Both methods only work well, if the number of random decision trees is large, especially if there are only features which are nearly uncorrelated with the classes. In the experiments by Breiman (2001), decision trees were used five times more than weak learners in Adaboost. Therefore, we focus our work on Adaboost and its variants.

3 FEATURE SELECTION WITH ADABOOST

The basic algorithm of Adaboost as taken from (Schapire and Singer, 1999) is shown in alg. 1. Input are the number T of iterations and N samples (x_n, y_n) with binary targets, i. e. $y_n \in \{+1, -1\}$. In each iteration, the best weak classifier is determined with respect to the samples weights. Each weak classifier h_t is a function $h_t : x_n \mapsto \{+1, -1\}$. After Adaboost has terminated after T iterations, the resulting strong classifier H can be depicted as weighted majority voting over the responses of all weak classifiers:

$$H(x_n) = \text{sign} \left(\sum_t \alpha_t h_t(x_n) \right). \quad (1)$$

The α_t are predictive values and depend on the weak classifiers success rates.

Algorithm 1 Adaboost algorithm

```

1: function ADABOOST( $T, (x_1, y_1), \dots, (x_N, y_N)$ )
2:    $W_1^n = \frac{1}{N}$ 
3:   for  $t = 1, \dots, T$  do
4:     Determine best weak hypothesis  $h_t$  using  $W_t$ 
5:     Determine  $\alpha_t$ 
6:     Determine distribution  $W_{t+1}$ 
7:   end for
8:   return  $H$  with  $H(x) = \text{sign} \left( \sum_t \alpha_t h_t(x) \right)$ .
```

If the weak classifiers are designed in a way that they use only a limited set of features, e. g. 1 feature only, then we are able to derive the relevance of features from the relevance of the weak classifiers. In the case, where only classifiers on single features are used, the strong classifier H consists of T weak classifiers h_1, \dots, h_T , and we obtain a list of maximally T features that are involved in classification. If $T < D$, then these maximally T features are the most appropriate features for classification.

Another strategy for obtaining the best feature subset is presented in (Drauschke and Förstner, 2008). In case $T \approx D$ or $T > D$, we cannot assume to find the best subset by only taking those features that have been used by the first weak classifiers. The influence of a feature depends on the absolute value of the predictive values α_t . Since these α_t do not monotonously decrease with increasing t , we have to evaluate the features after the iterative process has terminated.

4 DATA

We use terrestrial images of the eTRIMS data base for testing our feature selection and classification. This data base contains several hundred buildings from several major European cities or their suburbs. We selected 82 images from Berlin, Bonn and Munich, Germany, and from Prague, Czech Republic.

Typically, buildings have dominant vanishing lines in horizontal and vertical directions, respectively. If a pair of lines in the image is selected for each of both directions, we are able to determine the vanishing points and the homography for rectifying the image, cf. (McGlone, 2004), p. 775. So far, we work on rectified images which where the image rectification has been calculated after manual selection of these two pairs of lines. We show such a rectified image in fig. 1.

Our class ontology contains several object classes, and our goal is to find instances of these classes in images. At this stage, we limit our experiments to a subset of classes. On the one hand, we are interested in building detection, and therefore, we want to classify image regions as *facade*, *roof*, *sky* or *vegetation*. On the other hand, we are interested in detection of building parts, and therefore, we want to classify image regions as *windows* or *window panes*. Objects of both classes appear in all images, and regarding their cardinality, these are the most dominant structures in all images of the eTRIMS data base.



Figure 1: Rectified facade image and an detail from eTRIMS data base with manually annotated regions, e. g. window panes, vegetation.

We did our first experiments on manually annotated regions from rectified facade images, see fig. 1. These tests can show us the relevance of the features with respect to an optimal image segmentation.

In the second experiments, we used automatically segmented image regions. We obtain these image regions from the analysis of the image's scale-space with S discrete layers, which we have already used in (Drauschke et al., 2006). We adopted some parameters, i. e. we only consider the original image and additional 41 layers in scale-space with scales between $\sigma = 0.5$ and $\sigma = 8$. Then, we automatically trace the regions through the scale-space structure to derive a region's hierarchy, similar to the approach of Bangham et al. (1999). For complexity reason, we reduce the number of regions by selecting only stable regions in scale-space structure. Distinct region borders are often good evidences for stable regions. Thus, most stable regions correspond to man-made structures or are caused by shadows. The process of determining stable regions is explicitly described in (Drauschke, 2008). Fig. 2 shows all detected stable regions in scales $\sigma = 2$.



Figure 2: Segmented stable regions at scales $\sigma = 2$.

We derive 164 features from each manually annotated and each automatically segmented image region. Thus, our samples x_n are 164-dimensional feature vectors. These features are roughly described in tab. 1. Color features are determined with respect to two color channels RGB and HSV. Last, we derive the features f_{157} to f_{164} from the

vectorized region's border, which is a simplification of the region's boundary by using the algorithm of Douglas and Peucker (1973).

Table 1: List of derived features from image regions.

f_1	area
f_2	circumference
f_3	form factor
f_4	vertical elongation of bounding box
f_5	horizontal elongation of bounding box
f_6	ratio $f_4 : (f_4 \cdot f_5)$
f_7-f_{12}	mean color value in original image regarding the six channels
$f_{13}-f_{18}$	variance of color values in original image regarding the six channels
$f_{19}-f_{108}$	normalized histogram entries of gradients magnitude, 15 bins per channel
$f_{109}-f_{156}$	normalized histogram entries of gradients orientation, 8 bins per channel
f_{157}	portion of lengths of parallel edges in vectorized region's boundary
f_{158}	portion of number of parallel edges
f_{159}	portion of lengths of boundary edges which are parallel to region's major axis
f_{160}	portion of number of boundary edges which are parallel to region's major axis
f_{161}	portion of lengths of boundary edges which are parallel to region's minor axis
f_{162}	portion of number of boundary edges which are parallel to region's minor axis
f_{163}	portion of number of orthogonal angles in vectorized region's boundary
f_{164}	portion of lengths of boundary edges which are adjacent to orthogonal angles

The targets y_n are obtained differently. For manually annotated regions, we additionally select the appropriate class. Otherwise, the automatically segmented regions inherit the class target from the best fitting manually annotated region. The best fitting annotated region A_{i^*} is determined by

$$i^* = \operatorname{argmax}_i \frac{R \cap A_i}{R \cup A_i}, \quad (2)$$

where R is a automatically segmented region and A_i are all manually annotated regions. Furthermore, the best fitting annotated region must fulfill the condition

$$\frac{R \cap A_{i^*}}{R} > 0.5, \quad (3)$$

otherwise the class target of the segmented region will be set to *none*, and the segmented region will be always treated as background.

5 EXPERIMENTS

The goal of our experiments was to find appropriate features from the set of image features, which can be used for classifying our automatically segmented stable image regions. Therefore, we designed the Adaboost algorithm as follows. First, we only use weak classifiers which perform threshold classifications on a single feature. And secondly,

we perform Adaboost with only 30 iterations. We checked, that the classification error on the training samples is more or less constant after a few iterations and decreases very slowly in further iterations.

We are not only interested in reasonable classification results, but we also want to derive the most appropriate features. Therefore, we analyze the stability of the decisions of the weak classifiers. We repeated our tests V times, and then we checked the features the weak classifiers h_t use. Therefore, we determined histograms over the features which are used by the t -th weak classifier, $t = 1..30$. Furthermore, we focused on two certain characteristics. First, we analyzed, how many different features are used by t -th weak classifier in V trials: c_1 , $1 \leq c_1 \leq D$. And secondly, we evaluated, how often the best feature has been used by t -th weak classifier in V trials: c_2 , $1 \leq c_2 \leq V$. Both characteristics can be directly derived from the histograms: c_1 is the number of histogram entries which are not 0, and maximum frequency of a feature, c_2 , is the value of the highest peak.

If the number of used features is relatively small, then the weak classifiers always select the same features. We have found a set of appropriate features, if it stays consequently below a certain threshold.

5.1 Experiments on annotated regions

We selected 62 facade images from Bonn, Germany, and Prague, Czech Republic. Together, there are 5284 annotated objects in these images. Over 70% of them are *window panes*, and over 20% of the annotated objects are *windows*.

5.1.1 Facade, Roof, Sky and Vegetation The first experiment included 35 annotated *facades*, 37 *roofs*, 47 times *sky* and 70 times *vegetation*, i. e. 189 samples in total. Since the data set is quite small, we were able to perform four Leave-one-out-tests, i. e. $V = 189$, with the samples of one class as foreground and the samples from the other three classes as background. The classification errors are shown in tab. 2.

Table 2: Error rates of manually annotated regions.

facade	roof	sky	vegetation
9.5%	48.1%	1.1%	11.1%

In tab. 3 and fig. 3, we present the histograms over the features which are used by the t -th weak classifier regarding the classification of facades and roofs. With respect to the classification of facades, we notice that the first weak classifiers h_1 to h_4 distinguish facades from background by using nearly the same features over all 189 trials. Thus, the set of appropriate features should certainly contain the features f_3 , f_4 , f_6 and f_{163} . Then, number of used features, c_1 , increases for the further weak classifiers. Finally, in the 30-th iteration c_1 reaches a value of 21, i. e. the weak classifiers h_{30} of all 189 trials use 21 different features. In 91 cases, the f_3 was chosen, and f_1 was chosen in 35 cases, again. All other features which are used by h_{30} only play a minor role.

Table 3: Used features and maximum frequency of a feature with respect to the weak classifiers h_t for classifying facades in 189 trials.

h_1	h_3	h_5	h_{30}
$f_6 : 187$	$f_4 : 188$	$f_{144} : 174$	$f_3 : 91$
$f_3 : 2$	$f_1 : 1$	$f_{141} : 8$	$f_1 : 35$
$c_1 = 2$	$c_1 = 2$	$c_1 = 6$	$c_1 = 21$
$c_2 = 187$	$c_2 = 188$	$c_2 = 174$	$c_2 = 91$

The almost perfect classification of sky relies on the features f_9 , f_{151} and f_{130} . Here, even the last weak classifier h_{30} chooses in 164 or 87% of all cases the feature f_{129} , all other features which are used by h_{30} can get neglected. Regarding the classification of vegetation, the first weak classifiers mainly use the features f_9 , f_5 and f_{19} . The development of the histograms of the further weak classifiers is similar to the observation which we made after the facade classification.

The classification results of roofs is unacceptable, and here the selection of appropriate feature is not so stable as in the other three tests. Although the first weak classifiers always uses f_{15} as the discriminative feature, all further weak classifiers vary much more with respect to the selected feature. We show the histograms of four weak classifiers in fig. 3.

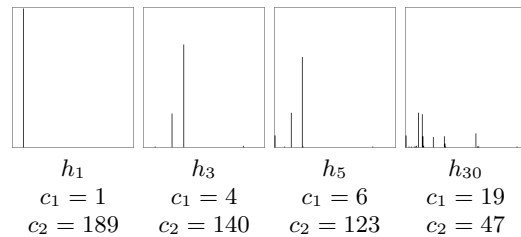


Figure 3: Histograms over the features of each weak classifier h_t and the number of used features c_1 and the maximum frequency c_2 from classifying roofs.

Due to the limited space, we do not present additional histograms in detail. But the plots of fig. 4 shows how the number of used features c_1 and the maximum frequency c_2 develop with respect to the classification target. Comparing the curves of c_2 with respect to roof and sky, it decreases much faster with bad classification results.

5.1.2 Window and Window Panes In the other two experiments, we tested the classification of windows and window panes. Since we could use 5284 annotated regions, we performed a cross validation test, where we used 10% of the regions for testing and the rest for training the classifiers. Then, we repeated all tests $V = 20$ times. The results are presented in tab. 4.

Table 4: Error rates of manually annotated regions.

class	min	median	max	total
window	15.4%	24.7%	75.3%	41.8%
window pane	3.1%	6.0%	70.0%	10.1%

The classification results of the windows and window panes make us very optimistic that we might get reasonable sets of appropriate features. Perturbingly, the classification errors vary too much between the 20 different tests. Furthermore, we cannot find any correlation between the classification results and the stability of the weak classifiers choice

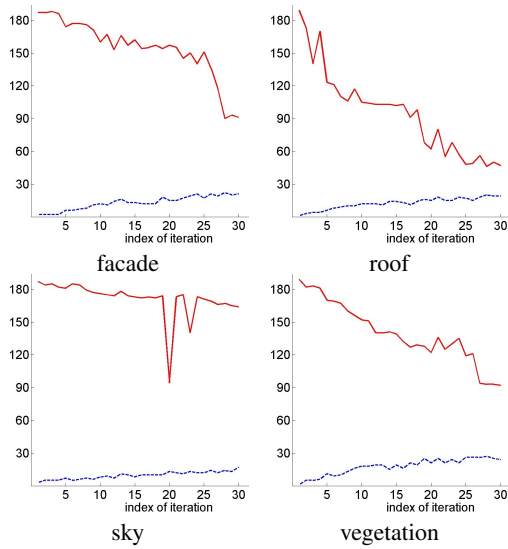


Figure 4: Development of the number of used features c_1 (dashed blue line) and the maximum frequency c_2 (solid red line) with respect to the 30 weak classifiers h_t .

on features. For window classification, the weak classifier mainly use the three features f_5 (in 28 cases!), f_4 and f_{18} . Although the classification of window panes yields much better results, the curves of c_1 and c_2 change very fast, cf. fig. 5. Nevertheless, the histograms show that only a small subset of four features is mainly used by the 30 weak classifiers, cf. fig. 6.

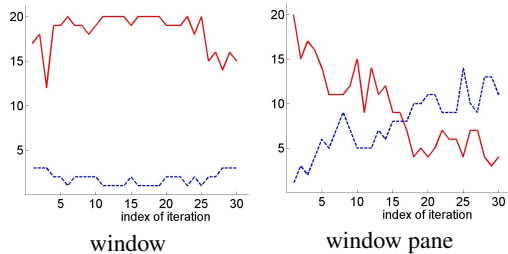


Figure 5: Development of the number of used features c_1 (dashed blue line) and the maximum frequency c_2 (solid red line) with respect to the 30 weak classifiers h_t .

5.2 Experiments on segmented data

We determined the stable regions from 82 facade images from Berlin and Munich, Germany, and Prague, Czech Republic. We changed the image data that was the basis of the previous experiments a bit, so we only worked with buildings of similar scale and with images of these buildings that have a similar resolution. We extracted over 13 000 stable regions, and then we repeated the cross validation tests on five selections of regions. First, we considered all scales, and then we only considered regions which were stable in the scale-space layers with scales $\sigma = 1, 2, 4, 8$, respectively.

In the tabs. 5 and 6, we show the results of our classifications. With respect to the experiment on classifying roofs, we are surprised that the classification of regions over all scales yields much better results than the classification of

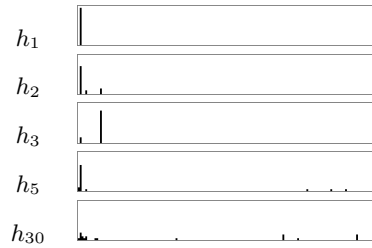


Figure 6: Histograms over the features of each weak classifier h_t : f_2 , f_5 and f_{13} are the most dominant features for classifying window panes.

regions of the same scale. When choosing regions from all scales, all weak classifiers choose one of the three features f_1 , f_2 and f_5 . Thus, $c_1 \leq 3vt$, cf. fig. 7. Surprisingly, the bad classification results of roof regions in scale $\sigma = 1$ is also based on stable decisions of the weak classifiers since only 11 features are used by all weak classifiers, cf. fig. 8.

Table 5: Error rates on automatically segmented regions.

scale	facade	roof	sky	vegetation
all	39.67%	15.88%	44.44%	20.75%
1	35.00%	81.06%	44.81%	33.17%
2	42.94%	28.44%	47.55%	45.27%
4	53.96%	39.76%	50.43%	43.26%
8	44.71%	25.14%	51.71%	58.86%

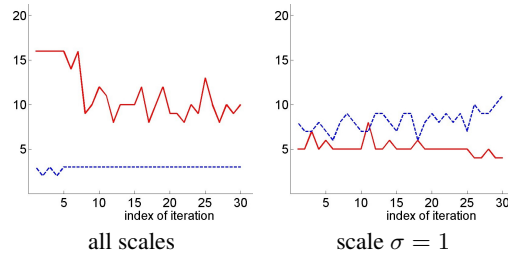


Figure 7: Development of the number of used features c_1 (dashed blue line) and the maximum frequency c_2 (solid red line) with respect to the 30 weak classifiers h_t .

6 CONCLUSION AND OUTLOOK

In this paper, we presented a feature selection scheme which is connected to the classification framework of Adaboost. We chose very simple weak classifiers which only work on single features. Thus, we were able to derive information on the relevance of features for the classification process. We could show, that the weak classifiers favour the use of the same features. So, we obtained sets of appropriate features even for the cases where we did not find good classification results. Therefore, we resume that we should improve the classification, e. g. by using more complex weak classifiers on 2-dimensional feature planes as $f_d \times f_d^2$. Additionally, we should expand the feature space, further features might be derived from texture.

ACKNOWLEDGEMENTS

This work was done within the project *eTraining for Interpreting Images of Man-Made Scenes (eTRIMS)*, *STREP*

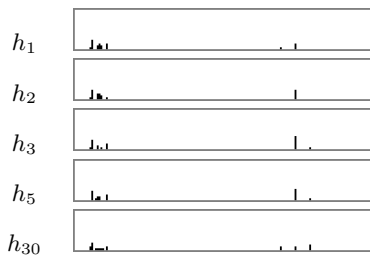


Figure 8: Histograms over the features of each weak classifier h_t : $f_9, f_{10}, f_{12}, f_{13}, f_{14}, f_{15}, f_{16}, f_{18}, f_{113}, f_{121}$ and f_{129} are the only features that are involved in classifying roof regions at scale $\sigma = 1$.

Table 6: Error rates on automatically segmented regions.

scale	window	window pane
all	21.0%	51.4%
1	18.2%	13.1%
2	19.9%	19.0%
4	86.4%	31.3%
8	38.6%	7.5%

027113 which is funded by The European Union. The authors would also like to thank Marko Pilger for implementing the functions for feature extraction.

References

Bangham, J. A., Moravec, K., Harvey, R. and Fisher, M., 1999. Scale-space trees and applications as filters for stereo vision and image retrieval. In: *BMVC*, pp. 113–143.

Bishop, C. M., 2006. *Pattern Recognition and Machine Learning*. Springer.

Breiman, L., 2001. Random forests. *Machine Learning* 45, pp. 5–32.

Douglas, D. H. and Peucker, T. K., 1973. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Canadian Cartographer* 10(2), pp. 112–122.

Drauschke, M., 2008. Description of stable regions ipm. Technical report, Department of Photogrammetry, University of Bonn.

Drauschke, M. and Förstner, W., 2008. Comparison of adaboost and adtboost for feature subset selection. In: *Proc. 8th PRIS 2008*.

Drauschke, M., Schuster, H.-F. and Förstner, W., 2006. Detectability of buildings in aerial images over scale space. In: *PCV06, IAPRS 36 (3)*, pp. 7–12.

Epshtein, B. and Ullman, S., 2005. Feature hierarchies for object classification. In: *Proc. 10th ICCV, Beijing, China*, pp. 220–227.

Fischer, A., Kolbe, T. H., Lang, F., Cremers, A. B., Förstner, W., Plümer, L. and Steinhage, V., 1998. Extracting buildings from aerial images using hierarchical aggregation in 2d and 3d. *CVIU* 72(2), pp. 185–203.

Guyon, I. and Elisseeff, A., 2003. An introduction to variable and feature selection. *Journal of Machine Learning Research* 3, pp. 1157–1182.

Hartz, J. and Neumann, B., 2007. Learning a knowledge base of ontological concepts for high-level scene interpretation. In: *ICMLA*.

Ho, T. K., 1998. The random subspace method for constructing decision forests. *PAMI* 20(8), pp. 832–844.

Kumar, S. and Hebert, M., 2003. Man-made structure detection in natural images using a causal multiscale random field. In: *CVPR*.

Lifschitz, I., 2005. Image interpretation using bottom-up top-down cycle on fragment trees. Master’s thesis, Weizmann Institute of Science, Rehovot, Israel.

Liu, H. and Motoda, H., 1998. *Feature Selection for Knowledge Discovery and Data Mining*. Kluwer Academic.

Mayer, H., 1999. Automatic object extraction from aerial imagery - a survey focusing on buildings. *CVIU* 74, pp. 138–149.

McGlone, J. C. (ed.), 2004. *Manual of Photogrammetry*. 5th edition edn, ASPRS.

Nevatia, R., Lin, C. and Huertas, A., 1997. A system for building detection from aerial images. In: *USC Computer Vision*.

Rätsch, G., Onoda, T. and Müller, K.-R., 2001. Soft margins for adaboost. *Machine Learning* 43(3), pp. 287–320.

Reznik, S. and Mayer, H., 2007. Implicit shape models, model selection, and plane sweeping for 3d facade interpretation. In: *PIA07, IAPRS 36 (3/W49A)*, pp. 173–178.

Schapire, R. E., 1990. The strength of weak learnability. *Machine Learning* 5(2), pp. 197–227.

Schapire, R. E. and Singer, Y., 1999. Improved boosting algorithms using confidence-rated predictions. *Machine Learning* 37(3), pp. 297–336.

Sivic, J., Russell, B. C., Efros, A. A., Zisserman, A. and Freeman, W. T., 2005. Discovering objects and their location in images. In: *Proc. 10th ICCV, Beijing, China*.

Tuytelaars, T., Turina, A. and Van Gool, L., 2003. Noncombinatorial detection of regular repetitions under perspective skew. *PAMI* 25(4), pp. 418–432.

Verbeek, J. and Triggs, B., 2007. Region classification with markov field aspect models. In: *VCVPR*.

Wang, X., Totaro, S., Taillandier, F., Hanson, A. R. and Teller, S., 2002. Recovering facade texture and microstructure from real-world images. In: *Proc. 2nd IW-TAS*, pp. 145–149.