

Regionwise Classification of Building Facade Images

Michael Ying Yang and Wolfgang Förstner

Department of Photogrammetry
Institute of Geodesy and Geoinformation
University of Bonn
Nussallee 15, 53115 Bonn, Germany
michaelyangying@uni-bonn.de
wf@ipb.uni-bonn.de
<http://www.ipb.uni-bonn.de>

Abstract. In recent years, the classification task of building facade images receives a great deal of attention in the photogrammetry community. In this paper, we present an approach for regionwise classification using an efficient randomized decision forest classifier and local features. A conditional random field is then introduced to enforce spatial consistency between neighboring regions. Experimental results are provided to illustrate the performance of the proposed methods using image from eTRIMS database, where our focus is the object classes *building, car, door, pavement, road, sky, vegetation, and window*.

Keywords: Classification, conditional random field, random decision forest, segmentation, facade image

1 INTRODUCTION

This paper presents an investigation into the semantic scene interpretation in the context of terrestrial man-made scenes. Our goal is the interpretation of the scene contained in an image as a collection of meaningful regions. The result of such an interpretation could be a rich image description useful for 3D city models of high level of detail, which are mainly used for visualization purposes, by thematic attributes to eventually obtain truly 3D geoinformation systems for complete cities. In this paper we focus on the problem of multi-class classification of building facade image regions.

Facade classification is the task of estimating the position and size of various structural (e.g. window, door) and non-structural elements (e.g. sky, road, building) in a given image (Fröhlich et al., 2010). Despite the substantial improvements during the past decade, the classification task remains a challenging problem, which receives a great deal of attention in the photogrammetry community (Rottensteiner et al., 2007; Korč and Förstner, 2008; Fröhlich et al., 2010; Kluckner and Bischof, 2010; Teboul et al., 2010).

Recently, an increasingly popular way to solve various image interpretation problems like object segmentation, stereo and single view reconstruction is to formulate them using image regions obtained from unsupervised segmentation algorithms. These methods are inspired from the observation that pixels constituting a particular region

often have the similar property. For instance, they may belong to the same object. This approach has the benefit that higher order features based on all the pixels constituting the region can be computed and used for classification. Further, it is also much faster as inference now only needs to be performed over a small number of regions rather than all the pixels in the image.

An image dataset showing urban buildings in their environment has been made public by Korč and Förstner (2009), which is called eTRIMS dataset. It allows benchmarking of facade image classification. Most of the images of this data set show facades in Switzerland and Germany. Regarding terrestrial facade images in this dataset, the most dominant objects are building, window, vegetation, and sky.

In this work, randomized decision forest (RDF) (Breiman, 2001) is adopted as the classifier, which gives good classification results on building facade images. Because of its rather local nature, a conditional random field (CRF) (Lafferty et al., 2001) is often introduced to enforce spatial consistency. We show that CRF classification results give average performance of 65.8% accuracy on eTRIMS 8-class dataset (Korč and Förstner, 2009). The remainder of the paper is organized as follow. Sec. 2 reviews some existing methods for building facade image classification. RDF classifier for performing image classification, and refinement with a CRF are described in Sec. 3. In Sec. 4, we show our results of RDF and CRF with respect to the classification of facade images. We finally conclude with a brief summary in Sec. 5.

2 RELATED WORKS

Previous works on building facade classification mostly regard the facade classification problem as multiple object detection tasks. Building facade detection is a very active research area in photogrammetry and computer vision. Drauschke and Förstner (2008) present a feature selection scheme with Adaboost for detecting buildings and building parts, where some of these features are used for classification in this paper. Graphical models are often used for integrating further information about the content of the whole scene in recent approaches (Kumar and Hebert, 2003; 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 (Sivic et al., 2005). However, 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, and it is rather slow and the most time consuming part of the whole system.

Support vector machines are widely considered as a good classifier. Schnitzspan et al. (2008) propose hierarchical support vector random fields that SVM is used as a classifier for unary potentials in conditional random field framework. While the training and cross-validation steps in support vector machines are time consuming, randomized decision forest (RDF) (Breiman, 2001) is introduced to significantly speed up the learning and prediction process. Existing work has shown the power of a RDF as a useful classifier (Bosch et al., 2007; Lepetit et al., 2005; Maree et al., 2005). The use of a RDF for semantic segmentation is investigated in Shotton et al. (2008); Dumont et al. (2009); Fröhlich et al. (2010). These approaches utilize simple color histogram features or pixel differences. Multi-class facade segmentation by combining a machine learn-

ing approach with procedural modeling as a shape prior is presented by Teboul et al. (2010). Generic shape grammars are constrained so as to express buildings only. Randomized forests are used to determine a relationship between the semantic elements of the grammar and the observed image support. Drauschke and Mayer (2010) also use random forest as one of the classifiers to evaluate the potential of texture filter banks for the pixelwise classification of facade images in eTRIMS dataset. Fröhlich et al. (2010) combine an RDF and local features for pixelwise labeling of facade images. Compared to Fröhlich et al. (2010), we derive the RDF classification on the image region level, rather than pixel level, and use a CRF to refine the classification results, where we incorporate location potentials and pairwise potentials in the CRF model. Other related works using the eTRIMS dataset concentrate on reconstructing the facade with grammar (Ripperda and Brenner, 2009) or detection of structural elements (e.g. windows and balconies (Wenzel and Förstner, 2008)), while we focus on multi-class image classification.

3 CLASSIFICATION

3.1 Randomized Decision Forest

We take randomized decision forest (RDF) (Breiman, 2001) as the classifier which operates on the regions defined by some unsupervised segmentation methods. In order to train the classifier, each region is assigned the most frequent class label it contains. Then a RDF is trained on the labeled data for each of the object classes. As illustrated in Fig. 1, a RDF is an ensemble classifier that consists of T decision trees (Shotton et al. (2008)). According to a decision tree learning algorithm, a decision tree recursively splits left or right down the tree to a leaf node. We use the extremely randomized trees (Geurts et al. (2006)) as learning algorithm. A feature vector d_i of image region i is classified by going down each tree. This process gives a class distribution at the leaf nodes and also a path for each tree. The class distributions $P(x_i | d_i)$ is obtained by averaging the class distribution over the leaf nodes for all T trees.

We now describe how the features are constructed from low-level descriptors. For each region, we compute an 178-dimensional feature vector, first incorporating region area and perimeter, its compactness and its aspect ratio. For representing spectral information of the region, we use 9 color features as Barnard et al. (2003): the mean and the standard deviation of the RGB and the HSV color spaces. Texture features derived from the Walsh transform (Petrrou and Bosdogianni, 1999; Lazaridis and Petrrou, 2006) are also used. Additionally we use mean SIFT descriptors (Lowe, 2004) of the image region. SIFT descriptors are extracted for each pixel of the region at a fixed scale and orientation, which is practically the same as the HOG descriptor (Dalal and Triggs, 2005), using the fast SIFT framework in Vedaldi and Fulkerson (2008). The extracted descriptors are then averaged into one l_1 -normalized descriptor vector for each region. Other features are derived from generalization of the region's border and represent parallelity or orthogonality of the border segments.

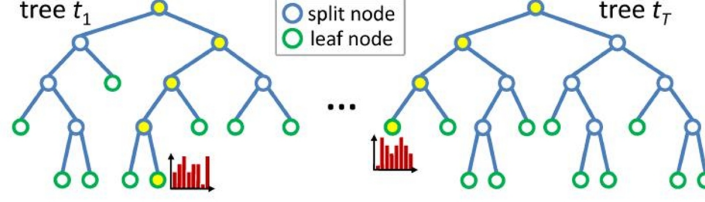


Fig. 1. Decision forest. A forest consists of T decision trees. A feature vector is classified by descending each tree. This gives, for each tree, a path from root to leaf, and a class distribution at the leaf. As an illustration, we highlight the root to leaf paths (yellow) and class distributions (red) for one input feature vector. (Figure courtesy by Jamie Shotton (Shotton et al., 2008).)

3.2 Refinement with a CRF

In order to enforce spatial consistency between neighboring regions, we must introduce constraints that allow us to reduce misclassification that occurs near the boundaries of image regions. Conditional random fields provide a natural way to incorporate such constraints by including them in the pairwise potentials of the model.

Given a random field \mathbf{X} defined over a graph $\mathcal{H} = (\mathcal{V}, \mathcal{E})$, the complete model for determining the optimal labeling $\mathbf{x} = \{x_i\}$ has a distribution of the Gibbs form

$$P(\mathbf{x} \mid \mathbf{d}) = \frac{1}{Z} \exp(-E(\mathbf{x} \mid \mathbf{d})) \quad (1)$$

with the energy function defined as

$$E(\mathbf{x} \mid \mathbf{d}) = \sum_{i \in \mathcal{V}} E_1(x_i) + \alpha \sum_{\{i,j\} \in \mathcal{N}} E_2(x_i, x_j) \quad (2)$$

where α is the weighting coefficient in the model, and Z is the normalization factor. The set \mathcal{V} is the set of nodes in the complete graph. The set \mathcal{N} is the set of pairs collecting neighboring. E_1 is the unary potential, which represents relationships between variables and local observed data. E_2 is the pairwise potential, which represents relationships between variables of neighboring nodes.

The graphical model is illustrated in Fig. 2. Nodes connection and numbers correspond to segmentation of a building facade image. The blue edges between the nodes represent the neighborhoods.

The local unary potentials E_1 independently predict the label x_i based on the image \mathbf{d} :

$$E_1(x_i) = -\log P(x_i \mid \mathbf{d}) \quad (3)$$

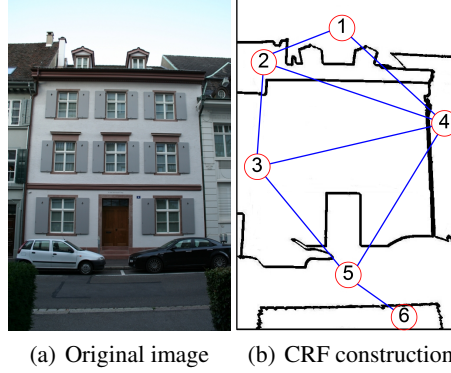


Fig. 2. Illustration of the CRF model architecture. Nodes connection and numbers correspond to segmented regions of a building facade image.

The label distribution $P(x_i | \mathbf{d})$ is usually calculated by using a classifier. The label distribution defined directly by the probability outputs provided by RDF for each region. The classifier which results from this is very specific. It finds regions which resemble regions that were seen in the training samples without considering the neighboring regions.

Based on the fact that RDF classifier does not take class label location information explicitly, we incorporate location potentials (similar to (Shotton et al., 2006)) in unary potentials. The location potential $-\log Q(x_i | \mathbf{d})$ takes the form of a look-up table with an entry for each class and region center location, where

$$Q(x_i | \mathbf{d}) = \left(\frac{N_{x_i, \hat{i}} + 1}{N_{\hat{i}} + 1} \right)^2$$

The index \hat{i} is the normalized version of the region index i . $N_{x_i, \hat{i}}$ is the number of regions of class x_i at normalized location within certain neighborhood in \hat{i} , and $N_{\hat{i}}$ is the total number of regions at location in \hat{i} . The location potentials capture the dependence of the class label on the rough location of the region in the image. Here, we use part of annotation images in 8-class eTRIMS Database (Korč and Förstner, 2009) to learn the location potentials, but ensure no overlap between these images and testing images in the experimental part. Some learned location potentials are illustrated in Fig. 3. Therefore, the unary potentials E_1 is written as

$$E_1(x_i) = -\log P(x_i | \mathbf{d}) - \log Q(x_i | \mathbf{d}) \quad (4)$$

The pairwise potentials E_2 describe category compatibility between neighboring labels x_i and x_j given the image \mathbf{d} , which take the form

$$E_2(x_i, x_j) = \frac{1 + 4 \exp(-2c_{ij})}{0.5(N_i + N_j)} \delta(x_i \neq x_j) \quad (5)$$

where c_{ij} is the norm of the color difference between regions in the LUV color space. N_i is the number of regions neighbored to region i , and N_j is the number of regions neighbored to j . The potentials E_2 are scaled by N_i and N_j to compensate for the irregularity of the graph \mathcal{H} . We refer the reader to Shotton et al. (2006); Gould et al. (2008) for more details about designing pairwise potentials.

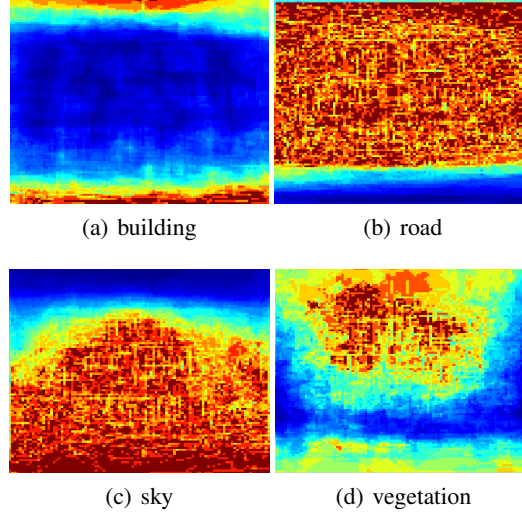


Fig. 3. Example location potentials. *Sky* tends to occur at the top part of images, while *road* tends to occur at the bottom part of images, and *building* tends to occur in the middle part of images. Here, the dark blue area indicates the most likely locations of one class, while the dark red area indicates the most unlikely locations.

3.3 Learning and Inference

The weighting parameter α , represents the tradeoff between pairwise regularization and the confidence in unary potentials, is the model parameter that should be learned. We estimate α by cross validation on the training data. Once the model has been learned, inference is carried out with the multi-label graph optimization library of (Boykov et al., 2001; Kolmogorov and Zabih, 2004; Boykov and Kolmogorov, 2004) using α -expansion. Since the CRF is defined on the graph of regions, inference is very efficient, taking less than half a second per image.

4 EXPERIMENTAL RESULTS

We conduct experiments to evaluate the performance of the proposed model on the the 8-class eTRIMS Database (Korč and Förstner, 2009). In the experiments, we take the

ground-truth label of a region to be the majority vote of the ground-truth pixel labels. We randomly divide the images into training and test data sets.

4.1 eTRIMS Database

We start with the eTRIMS 8-class Database which consists of 60 annotated building facade images, labeled with 8 classes: *building*, *car*, *door*, *pavement*, *road*, *sky*, *vegetation*, *window*. The ground-truth labeling is approximate (with foreground labels often overlapping background objects).

We segment the facade images using mean shift algorithm (Comaniciu and Meer, 2002), tuned to give approximately 480 regions per image. In all 60 images, we extract around 29 600 regions. We have following statistics. Compared to the ground-truth labeling, almost 33% of all the segmented regions get the classlabel *building*. 25% of all regions get the classlabel *window*. These statistics are very comprehensive, because facade images typically show buildings typically contain many windows. Furthermore, 19% of the regions get the classlabel *vegetation*, and 2% belong to *sky*, and the last 21% of the regions are spread over most of other classes. We randomly divide the images into a training set with 40 images and a testing set with 20 images.

4.2 RDF Classification Results

We take the RDF classifier, run experiments 5 times, and obtain overall averaging classification accuracy 58.8%. The number of decision trees is chosen as $T = 250$. Random chance would give $1/8 = 12.5\%$, and thus RDF results are about 5 times better than chance. In the whole paper, accuracy values are computed as the percentage of image pixels assigned to the correct class label, ignoring pixels labeled as void in the ground truth. Fig. 4 *Left* shows the classification results over all 8 classes. The classification accuracy with respect to numbers of decision trees T for training are shown in Fig. 4 *Right*. While increasing the number of decision trees, the classification accuracy also increases. After 250 iteration, the accuracy converges. So we choose $T = 250$ for performing experiments above.

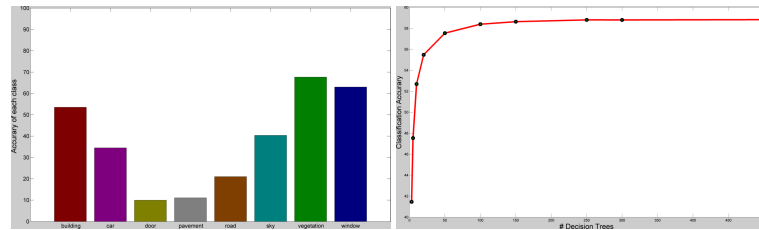


Fig. 4. *Left*: accuracy of each class of RDF classifier. The horizontal-axis reads *building*, *car*, *door*, *pavement*, *road*, *sky*, *vegetation*, *window*, while the vertical-axis reads $[0, 100]$. *Right*: classification accuracy with respect to numbers of decision trees for training. The horizontal-axis reads $[0, 500]$, while the vertical-axis reads $[0, 60]$.

Fig. 5 presents some result images of RDF method. The black regions in all the result images and ground truth images correspond to background. The quality inspection of the results in Fig. 5 shows that RDF classifier yields fair results. However, there exists some misclassification for each class. For example, the incorrect results at windows are often due to the reflectance of vegetation and sky in the window panes. Most sky regions are classified correctly. However, sky region is assigned label *car* in one image (last image in Fig. 5).

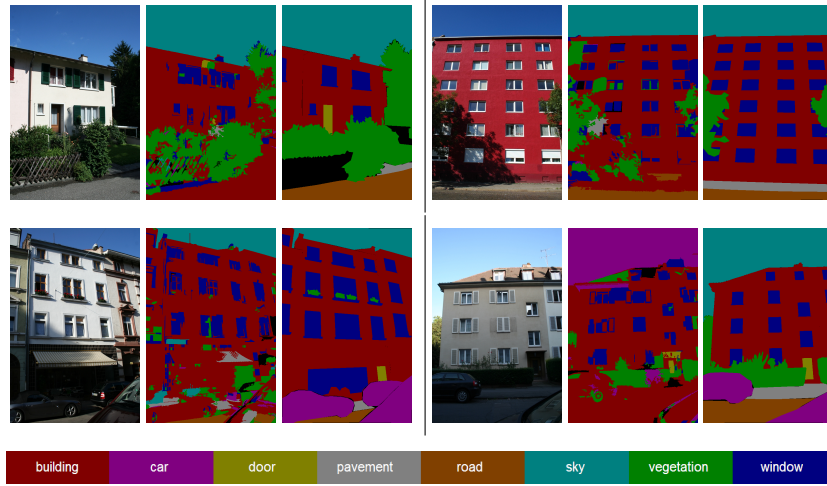


Fig. 5. Example images of eTRIMS database and classification results based on a randomized decision forest. (For each image triplet, *left*: test image, *middle*: result, *right*: ground truth.)

4.3 CRF Refinement Results

We run experiments 5 times, and obtain overall averaging classification accuracy 65.8% by applying CRF to the whole test set. For comparison, the RDF classifier alone gives an overall accuracy of 58.8%. Therefore, the location, and pairwise potentials increase the accuracy by 7%. For correct recovery of most *sky* regions, the location potentials increase the accuracy by 3%. This seemingly small numerical improvement corresponds to a large perceptual improvement (cf. Fig. 6). The weighting parameter, learned by cross validation on the training data, is $\alpha = 0.8$. Table 1 shows the confusion matrix.

Qualitative results of CRF on the eTRIMS dataset are presented in Fig. 6. The quality inspection of the results in these images shows that CRF yields good results. The greatest accuracies are for classes which have low visual variability and many training examples (such as window, vegetation, building, and sky) whilst the lowest accuracies are for classes with high visual variability or few training examples (for example door, car, and pavement). We expect more training data and the use of features with better

lighting invariance properties will improve the classification accuracy. Objects such as car, door, pavement, and window are sometimes incorrectly classified as *building*, due to the dominant presence of building in the image. This kind of problem may be resolved by some detection schemes. For example, detecting windows, cars, and doors should resolve some of such ambiguities. Furthermore, 36% of pavement is misclassified as road, and 44% of road is misclassified as pavement. Objects, such as pavement and road, can be confused with each other. This effect is partially attributable to inaccuracies in the manual ground truth labeling, where pixels are often mislabeled near object boundaries.

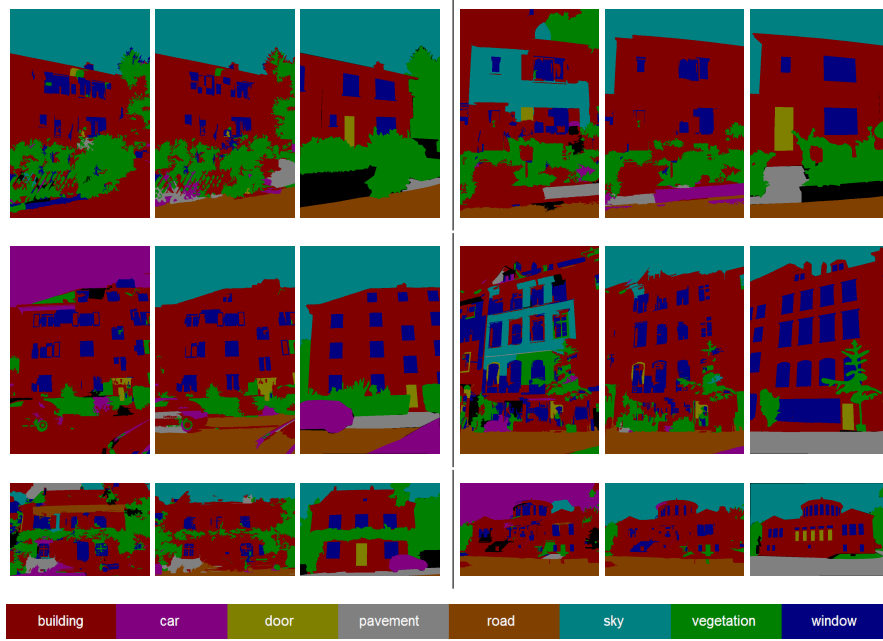


Fig. 6. Qualitative classification results of CRF on the eTRIMS database. (For each image triplet, *left*: RDF classification result, *middle*: CRF refinement result, *right*: ground truth.)

5 CONCLUSIONS AND FUTURE WORK

In this paper, we demonstrate an approach to regionwise classification of building facade images based on a randomized decision forest (RDF) and local features. In our experiments on eTRIMS dataset (Korč and Förstner, 2009), we have shown that this approach produces fair classification results. When we refine the classification with a conditional random field (CRF), which penalizes pairs of regions that are very different

Pr \ Tr								
	b	c	d	p	r	s	v	w
b	71	2	1	1	1	2	10	12
c	12	35	0	12	11	0	30	0
d	42	0	16	1	6	0	8	27
p	11	15	0	22	36	0	14	2
r	4	8	0	44	35	0	9	0
s	13	0	0	0	0	78	8	1
v	18	5	2	1	1	0	66	7
w	19	1	1	0	0	1	3	75

Table 1. Pixelwise accuracy of image classification using CRF on the eTRIMS 8-class database. The confusion matrix shows the classification accuracy for each class (rows) and is row-normalized to sum to 100%. Row labels indicate the true class (Tr), and column labels the predicted class (Pr). (b = *building*, c = *car*, d = *door*, p = *pavement*, r = *road*, s = *sky*, v = *vegetation*, w = *window*.)

in color, we consistently improve both our quantitative and especially our qualitative results.

Since neighboring regions in facade images are not independent from each other, CRF models these dependencies directly by introducing pairwise potentials. However, standard CRF works on a local level and long range dependencies are not addressed explicitly in the CRF. Our future work will try to set up a hierarchical model to integrate neighborhood and partonomy relations in a unified CRF model. A similar concept has been raised in Yang et al. (2010), where the model is denoted as hierarchical conditional random field (HCRF).

ACKNOWLEDGEMENTS

The work is funded by Deutsche Forschungsgemeinschaft (German Research Foundation) FO 180/16-1. The authors gratefully acknowledge the support.

References

- Barnard, K., Duygulu, P., Freitas, N. D., Forsyth, D., Blei, D., and Jordan, M. (2003). Matching words and pictures. *Journal of Machine Learning Research*, 3:1107–1135.
- Bosch, A., Zisserman, A., and Muñoz, X. (2007). Image classification using random forests and ferns. In *IEEE International Conference on Computer Vision*, pages 1–8.
- Boykov, Y. and Kolmogorov, V. (2004). An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26:1124–1137.
- Boykov, Y., Veksler, O., and Zabih, R. (2001). Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23:1222–1239.

- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.
- Comaniciu, D. and Meer, P. (2002). Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603–619.
- Dalal, N. and Triggs, B. (2005). Histograms of oriented gradients for human detection. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 886–893.
- Drauschke, M. and Förstner, W. (2008). Selecting appropriate features for detecting buildings and building parts. In *Proceeding of the 21st Congress of the International Society for Photogrammetry and Remote Sensing (ISPRS)*, number B3b, pages 447–452.
- Drauschke, M. and Mayer, H. (2010). Evaluation of texture energies for classification of facade images. In *ISPRS Technical Commission III Symposium on Photogrammetry Computer Vision and Image Analysis*, pages 257–262.
- Dumont, M., Marée, R., Wehenkel, L., and Geurts, P. (2009). Fast multi-class image annotation with random subwindows and multiple output randomized trees. In *International Conference on Computer Vision Theory and Applications (VISAPP)*, volume 2, pages 196–203.
- Fröhlich, B., Rodner, E., and Denzler, J. (2010). A fast approach for pixelwise labeling of facade images. In *International Conference on Pattern Recognition*, pages 3029–3032.
- Geurts, P., Ernst, D., and Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, 63(1):3–42.
- Gould, S., Rodgers, J., Cohen, D., Elidan, G., and Koller, D. (2008). Multi-class segmentation with relative location prior. *International Journal of Computer Vision*, 80(3):300–316.
- Gluckner, S. and Bischof, H. (2010). Image-based building classification and 3d modeling with super-pixels. In *ISPRS Technical Commission III Symposium on Photogrammetry Computer Vision and Image Analysis*, pages 233–238.
- Kolmogorov, V. and Zabih, R. (2004). What energy functions can be minimized via graph cuts? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(2):147–159.
- Korč, F. and Förstner, W. (2008). Interpreting terrestrial images of urban scenes using discriminative random fields. In *Proceeding of the 21st Congress of the International Society for Photogrammetry and Remote Sensing (ISPRS)*, pages 291–296.
- Korč, F. and Förstner, W. (2009). eTRIMS image database for interpreting images of man-made scenes. Technical Report TR-IGG-P-2009-01. <http://www.ipb.uni-bonn.de/projects/etrimdb>.
- Kumar, S. and Hebert, M. (2003). Discriminative random fields: a discriminative framework for contextual interaction in classification. In *IEEE International Conference on Computer Vision*, pages 1150–1157.
- Lafferty, J., McCallum, A., and Pereira, F. (2001). Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. In *International Conference on Machine Learning*, pages 282–289.
- Lazaridis, G. and Petrou, M. (2006). Image registration using the Walsh transform. *IEEE Transactions on Image Processing*, 15(8):2343–2357.

- Lepetit, V., Laguerre, P., and Fua, P. (2005). Randomized trees for real-time keypoint recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 775–781.
- Lowe, D. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110.
- Maree, R., Geurts, P., Piater, J., and Wehenkel, L. (2005). Random subwindows for robust image classification. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 34–40.
- Petrou, M. and Bosdogianni, P. (1999). *Image Processing: The Fundamentals*. Wiley.
- Ripperda, N. and Brenner, C. (2009). Evaluation of structure recognition using labelled facade images. In *Proceedings of the 31st DAGM Symposium on Pattern Recognition*, pages 532–541.
- Rottensteiner, F., Trinder, J., Clode, S., and Kubik, K. (2007). Building detection by fusion of airborne laser scanner data and multi-spectral images: Performance evaluation and sensitivity analysis. *ISPRS Journal of Photogrammetry and Remote Sensing*, 62(2):135–149.
- Schnitzspan, P., Fritz, M., and Schiele, B. (2008). Hierarchical support vector random fields: joint training to combine local and global features. In *European Conference on Computer Vision, Part II*, pages 527–540.
- Shotton, J., Johnson, M., and Cipolla, R. (2008). Semantic texton forests for image categorization and segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8.
- Shotton, J., Winnand, J., Rother, C., and Criminisi, A. (2006). Textonboost: joint appearance, shape and context modeling for multi-class object recognition and segmentation. In *European Conference on Computer Vision, Part I*, pages 1–15.
- Sivic, J., Russell, B. C., Efros, A. A., Zisserman, A., and Freeman, W. T. (2005). Discovering objects and their locations in images. In *IEEE International Conference on Computer Vision*, pages 1–8.
- Teboul, O., Simon, L., Koutsourakis, P., and Paragios, N. (2010). Segmentation of building facades using procedural shape priors. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3105–3112.
- Vedaldi, A. and Fulkerson, B. (2008). Vlfeat: An open and portable library of computer vision algorithms. <http://www.vlfeat.org/>.
- Verbeek, J. and Triggs, B. (2007). Region classification with markov field aspect models. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8.
- Wenzel, S. and Förstner, W. (2008). Semi-supervised incremental learning of hierarchical appearance models. In *Proceeding of the 21st Congress of the International Society for Photogrammetry and Remote Sensing (ISPRS)*, number B3b, pages 399 – 405.
- Yang, M. Y., Förstner, W., and Drauschke, M. (2010). Hierarchical conditional random field for multi-class image classification. In *Proceedings of the Fifth International Conference on Computer Vision Theory and Applications - Volume 2*, pages 464–469.