Feature Evaluation for Building Facade Images — An Empirical Study

Michael Ying Yang, Wolfgang Förstner

Department of Photogrammetry
Institute of Geodesy and Geoinformation
University of Bonn
Bonn, Germany

Abstract

Image classification are critically dependent on the features. In this paper, we perform an empirical feature evaluation task for building facade images. Feature sets we choose are basic features, color features, histogram features, peucker features, texture features, and SIFT features. We present an approach for region-wise labeling using an efficient randomized decision forest classifier and local features. We conduct our experiments with building facade image classification with eTRIMS database, where our focus is the object classes *building*, *car*, *door*, *pavement*, *road*, *sky*, *vegetation*, and *window*.
Keywords: Feature, evaluation, randomized decision forest, facade images

1. Introduction

Despite the substantial advances made during the past decade, the classification of building facade images remains a challenging problem that 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). Image classification are critically dependent on the features. Typical feature evaluation can be divided into two stages. First, image processing is used to extract a set of robust features that implicitly contains the information needed to make class-specific decisions while resisting extraneous effects such as changing object appearance, pose, illumination and background clutter. Second, a machine learning based classifier uses the features to make region-level decisions, often followed by post-processing to merge nearby decisions. Instead of using some unsupervised techniques, which bare generalization problem, it is popular way that the classifier is trained using a set of labeled training examples. The overall performance depends critically on all three elements: the feature set, the classifier & learning method, and the training set. In this paper, we focus on evaluating different feature sets.

Recently, Korč and Förstner (2009) published an image dataset showing urban buildings in their environment. It allows benchmarking of facade image classification, and therefore the repeatable comparison of different approaches. Most of the images of this data set show facades in Switzerland and Germany. Regarding terrestrial facade images, the most dominant objects are the building itself, the window, vegetation, and the sky. Fig. 1 demonstrates the variability of the object data.

In this work, we empirically investigate extended feature sets give state of the art performance on eTRIMS dataset (Korč and Förstner, 2009). We show random forest gives good classification results on building facade images, and evaluate classification results by counting corrected labeled regions. The remainder of the paper is organized as follow. Section 2 reviews some existing methods for feature

Figure 1: Example images from benchmark data set (Korč and Förstner, 2009).

evaluation and building facade image classification. Then, we introduce feature sets for evaluation in the scope of the paper in Section 3. Randomized decision forest classifier for performing image classification is described in Section 4. In Section 5, we show our results and discuss the effect of each feature sets with respect to the classification of facade images. We finally conclude with a brief summary in Section 6.

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 photogrammtery and computer vision. A feature selection scheme with Adaboost for detecting buildings and building parts is presented in Drauschke and Förstner (2008) . In recent approaches, graphical models are often used for integrating further information about the content of the whole scene (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). 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, and it is extremely slow and often the most time consuming part of the whole system, even with optimizations such as kd-trees, or hierarchical clusters (Nister and Stewenius, 2006).

Support vector machine (SVM) is 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 SVM 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 randomized decision forest as a classifier (Bosch et al., 2007; Lepetit et al., 2005; Maree et al., 2005). The use of a randomized decision forest for semantic segmentation was previously investigated in Shotton et al. (2008); Dumont et al. (2009); Fröhlich et al. (2010). These approaches utilize simple color histogram features or pixel differences. Fröhlich et al. (2010) present an approach using an randomized decision forest and local opponent-SIFT features (van de Sande et al., 2010) for pixelwise labeling of facade images. Teboul et al. (2010) perform multi-class facade segmentation by combining a machine learning approach with procedural modeling as a shape prior. 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 seven texture filter banks for the pixel-based classification of terrestrial facade images.

3. Feature sets

Image classification are critically dependent on the features that they use, which must capture the information needed to identify objects of the class despite highly variable object appearance, lighting, clutter, background texture, etc. Advances in feature sets have been a constant source of progress over the past decade. In this work, we derive 6 feature sets from each region obtained from some unsupervised segmentation algorithms, such as mean shift (Comaniciu and Meer, 2002), watershed (Vincent and Soille, 1991), or graph-based method (Felzenszwalb and Huttenlocher, 2004).

Basic features $\boldsymbol{f}_1$. First feature set $\boldsymbol{f}_1$ are basic features including (1). number of components of the region (C); (2). number of holes of the region (H); (3). Euler characteristic for planar figures (Lakatos, 1976) ($E = C - H$); (4). area (A); (5). perimeter (U); (6). form factor ($F = U^2/(4\pi A)$); (7). height of bounding box; (8). width of bounding box; (9). area ratio between region and its bounding box; (10). height portion of center; (11). width portion of center.

Color features $\boldsymbol{f}_2$. For representing spectral information of the region, we use 9 color features (Barnard et al., 2003) as second feature set $\boldsymbol{f}_2$: the mean and the standard deviation of R-channel, G-channel and B-channel respectively in the RGB color space; and the mean of H-channel, S-channel and V-channel respectively in the HSV color space.

Histogram features $\boldsymbol{f}_3$. We also include features derived from the gradient histograms as third feature set $\boldsymbol{f}_3$, which has been proposed by Korč and Förstner (2008). We determine gradient and its orientation and its magnitude. The histograms are determined for the 3 colors R, G and B respectively in the region. Then, we derive the mean, the variance and the entropy from each histogram as features.

Peucker features $\boldsymbol{f}_4$. Peucker features are derived from generalization of the region's border as fourth feature set $\boldsymbol{f}_4$, and represent parallelity or orthogonality of the border segments. We select the four points of the boundary which are farthest away from each other. From this polygon region with four corners, we derive 3 central moments, and eigenvalues in direction of major and minor axis, aspect ratio of eigenvalues, orientation of polygon region, coverage of polygon region, and 4 angles of polygon region boundary points.

Texture features $\boldsymbol{f}_5$. We use texture features derived from the Walsh transform (Petrou and Bosdogianni, 1999; Lazaridis and Petrou, 2006) as fifth feature set $\boldsymbol{f}_5$, as features from Walsh filters are among the best texture features from the filter banks (Drauschke and Mayer, 2010). We determine the magnitude of the response of 9 Walsh filters. For each of the 9 filters, we determine mean and standard deviation for each region.

SIFT features $\boldsymbol{f}_6$. Sixth feature set $\boldsymbol{f}_6$ are mean SIFT (Scale-Invariant Feature Transform) descriptors (Lowe, 2004) of the image region. SIFT descriptors are extracted for each pixel of the region at a fixed scale and orientation using the fast SIFT framework found in Vedaldi and Fulkerson (2008). The extracted descriptors are then averaged into one l_1-normalized descriptor vector for each region.

These features are roughly listed in Table 1.

Table 1: List of derived features from image regions. The number indicates feature numbers in each feature set.

$\boldsymbol{f}_1$ basic features (11) region area and perimeter, compactness and aspect ratio, etc.
$\boldsymbol{f}_2$ color features (9) mean and standard deviation of the RGB and the HSV color spaces
$\boldsymbol{f}_3$ histogram features (9) mean, variance and entropy of histogram from region's gradients
$\boldsymbol{f}_4$ peucker features (12) moments and eigenvalues of a region as orthogonality or parallelity
$\boldsymbol{f}_5$ texture features (18) texture features derived from the Walsh transform
$\boldsymbol{f}_6$ SIFT features (128) mean SIFT descriptor features

4. Randomized decision forest

Features are evaluated by a classifier which operates on the regions defined by unsupervised segmentation. we take randomized decision forest (RDF) (Breiman, 2001) as the classifier for performing feature evaluation. In order to train the classifier, each region is assigned the most frequent class label it contains. Existing work has shown the power of decision forests as classifiers (Bosch et al., 2007; Lepetit et al., 2005; Maree et al., 2005). We begin with a brief review of randomized decision forest (Amit and Geman, 1997; Geurts et al., 2006). As illustrated in Figure 2, a decision forest is an ensemble of T decision trees. A learned class distribution $P(c \mid n)$ is associated with each node n in the tree. A decision tree works by recursively branching left or right down the tree according to a learned binary function of the feature vector, until a leaf node l is reached. The whole forest achieves an accurate and robust classification by averaging the class distributions over the leaf nodes $L = (l_1, \cdots, l_t, \cdots, l_T)$ reached for all T trees:

$$P(c \mid L) = \frac{1}{T} \sum_{t=1}^{T} P(c \mid l_t) \tag{1}$$

We use the extremely randomized trees algorithm (Geurts et al., 2006) to learn binary forests. Each tree is trained separately on a small random subset $I' \subseteq I$ of the training data I. Learning proceeds recursively, splitting the training data I_n at node n into left and right subsets I_l and I_r according to a threshold δ of some split function g of the feature vector $\boldsymbol{h}$.

$$I_l = \{i \in I_n \mid g(\boldsymbol{h}_i) < \delta\} \tag{2}$$

$$I_r = I_n \setminus I_l \tag{3}$$

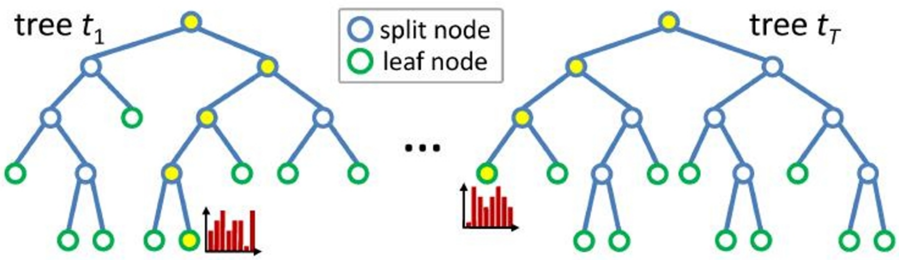


Figure 2: **Decision forest**. A forest consists of T decision trees. A feature vector is classified by descending each tree. This gives, for each tree, a path from root to leaf, and a class distribution at the leaf. As an illustration, we highlight the root to leaf paths (yellow) and class distributions (red) for one input feature vector. (Figure courtesy by Jamie Shotton (Shotton et al., 2008).)

At each split node, several candidates for function g and threshold δ are generated randomly, and the one that maximizes the expected gain in information about the node categories is chosen (Lepetit et al., 2005):

$$\Delta E = -\frac{I_l}{I_n}E(I_l) - \frac{I_r}{I_n}E(I_r) \tag{4}$$

where $E(I)$ is the Shannon entropy of the classes in the set of examples I. The recursive training continues to a maximum depth D or until no further information gain is possible. The class distributions $P(c \mid n)$ are estimated empirically as a histogram of the class labels c_i of the training examples i that reached node n.

5. Experimental Results

We conduct experiments to evaluate the performance of different image feature sets on the recently published dataset: 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.

5.1. eTRIMS Database

We start with the eTRIMS 8-class database which is a comprehensive and complex dataset consisting of 60 building facade images, mainly taken from Basel, Berlin, Bonn, and Heidelberg, labeled with 8 classes: *building*, *car*, *door*, *pavement*, *road*, *sky*, *vegetation*, *window*. These classes are typical objects which can appear in images of building facades. 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. 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. In all these experiments, we randomly divide the images into a training set with 40 images and a testing set with 20 images.

Table 2: Average accuracy of RDF classifier on each feature set of eTRIMS database (Korč and Förstner, 2009).

feature set	$\boldsymbol{f}_1$	$\boldsymbol{f}_2$	$\boldsymbol{f}_3$	$\boldsymbol{f}_4$	$\boldsymbol{f}_5$	$\boldsymbol{f}_6$
accuracy	43.8%	49.6%	36.5%	40.9%	27.9%	54.1%

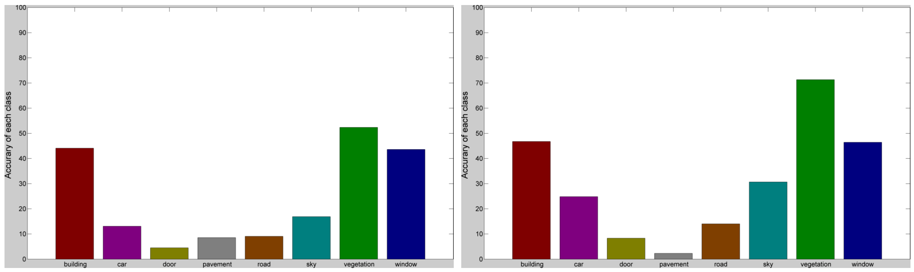
5.2. Evaluation with RDF classifier

In the following, we first evaluate with RDF classifier on each feature set $\boldsymbol{f}_1, \boldsymbol{f}_2, \boldsymbol{f}_3, \boldsymbol{f}_4, \boldsymbol{f}_5$, and $\boldsymbol{f}_6$. Then, we evaluate with RDF classifier on the combination of feature sets, and show that RDF gives fair results on building facade images.

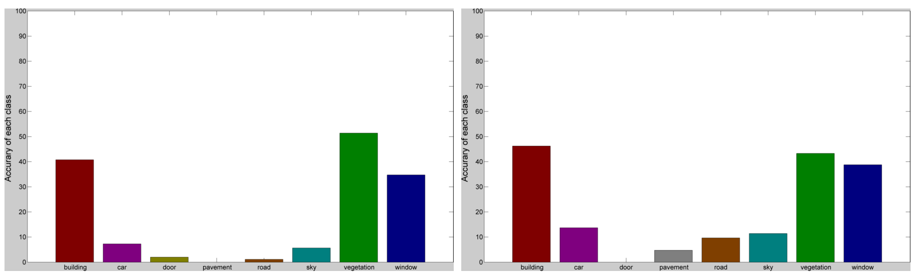
The overall classification accuracy is listed in Table 2, when applying RDF classifier on each feature set. The number of decision trees is chosen as $T = 250$. In all the following experiments, we always assume maximum depth of each decision tree $D = 7$. A random classifier for 8 classes, the expected classification accuracy is 12.5%. Fig. 3 shows the corresponding classification results over all 8 classes. Each class is normalized to 100%.

From Fig. 3, we observe that each feature set performs reasonable results on *building*, *window*, and *vegetation* classes. Color features $\boldsymbol{f}_2$ perform better than other features on *vegetation* class because most *vegetation* parts are homogeneous regions. For other classes, each feature set performs not good. Relatively, peucker features $\boldsymbol{f}_4$ perform better than other feature sets on minor classes. SIFT features $\boldsymbol{f}_6$ perform better than other features on average.

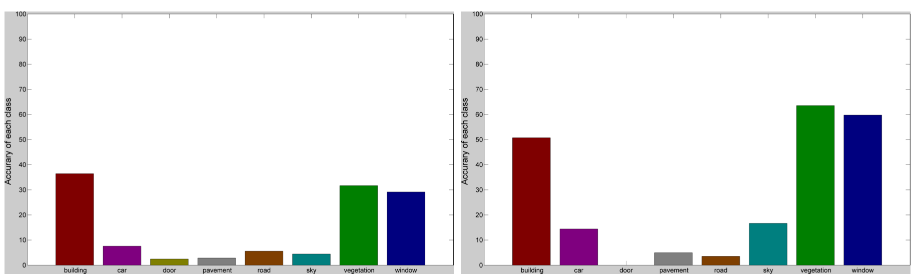
We also make the experiments using leave-one-out method. The overall clas-



(a) On feature set $\boldsymbol{f}_1$ (Left), $\boldsymbol{f}_2$ (Right)



(b) On feature set $\boldsymbol{f}_3$ (Left), $\boldsymbol{f}_4$ (Right)



(c) On feature set $\boldsymbol{f}_5$ (Left), $\boldsymbol{f}_6$ (Right)

Figure 3: **Accuracy of each class** on each feature set, with each class is normalized to 100.

Table 3: Average accuracy of RDF classifier on the feature sets. Feature sets $-\boldsymbol{f}_i$ mean the rest of all 6 feature sets except $\boldsymbol{f}_i$ is used, $i = 1, \cdots, 6$.

feature sets	$-\boldsymbol{f}_1$	$-\boldsymbol{f}_2$	$-\boldsymbol{f}_3$	$-\boldsymbol{f}_4$	$-\boldsymbol{f}_5$	$-\boldsymbol{f}_6$
accuracy	58.1%	57.2%	58.8%	58.1%	58.3%	53.0%

sification accuracy is listed in Table 3. Feature sets $-\boldsymbol{f}_i$ mean the rest of all 6 feature sets except $\boldsymbol{f}_i$ is used, $i = 1, \cdots, 6$. The number of decision trees is chosen as $T = 250$.

In the following, we make use of all the feature sets $\boldsymbol{f}_1, \boldsymbol{f}_2, \boldsymbol{f}_3, \boldsymbol{f}_4, \boldsymbol{f}_5, \boldsymbol{f}_6$. We run experiments 5 times, and obtain overall averaging classification accuracy 58.8% (± 0.24). The number of decision trees is also chosen as $T = 250$. Fig. 4 shows the classification results over all 8 classes. The classification accuracy with

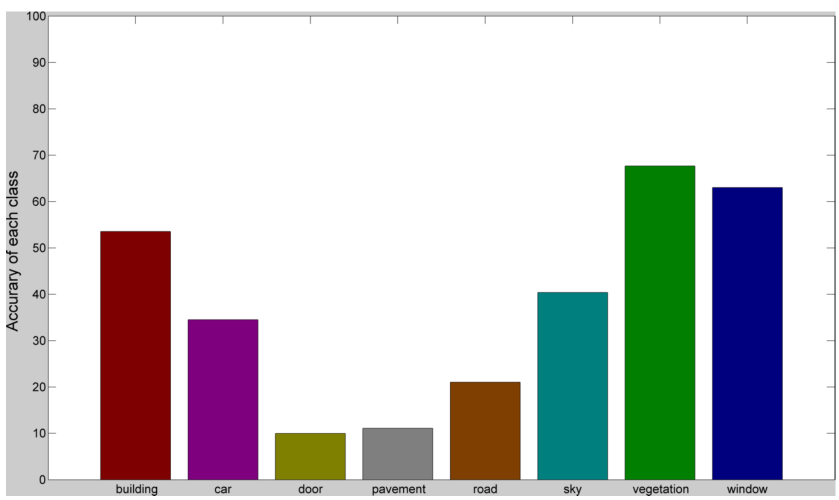


Figure 4: **Accuracy of each class** (a median accuracy result is shown here). All feature sets $\boldsymbol{f}_1, \boldsymbol{f}_2, \boldsymbol{f}_3, \boldsymbol{f}_4, \boldsymbol{f}_5, \boldsymbol{f}_6$ are used, with each class is normalized to 100.

respect to numbers of decision trees T for training are shown in Fig. 5. 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.

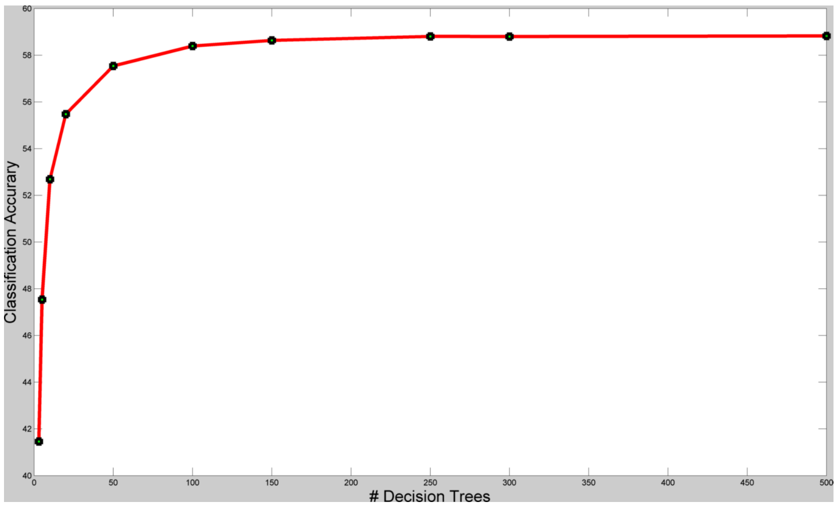


Figure 5: Classification accuracy with respect to numbers of decision trees for training. All feature sets $\boldsymbol{f}_1, \boldsymbol{f}_2, \boldsymbol{f}_3, \boldsymbol{f}_4, \boldsymbol{f}_5, \boldsymbol{f}_6$ are used.

Fig. 6 and Fig. 7 present 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. 6 shows that RDF classifier yields good results. In Fig. 7, 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 row in Fig. 7). This can be resolved simply by introducing some kind of spatial prior (Gould et al., 2008), such as *sky*

Table 4: Accuracy of RDF classifier on the eTRIMS 8-class database (Korč and Förstner, 2009). The confusion matrix shows the classification accuracy for each class (columns) and is column-normalized to sum to 100%. Column labels indicate the true class, and row labels the predicted class.

	build.	car	door	pave.	road	sky	veget.	window
build.	**60**	22	46	40	40	29	11	24
car	8	**40**	0	16	20	2	5	1
door	2	1	**15**	0	0	0	1	2
pave.	2	3	0	**12**	14	5	1	0
road	2	1	0	4	**23**	2	1	0
sky	1	2	0	4	3	**48**	0	1
veget.	9	29	8	16	0	7	**76**	4
window	16	2	31	8	0	7	5	**68**

is above the *building*, *road* and *pavement* are below the *building*, *car* is above the *road*, and *window* is surrounded by *building*. A full confusion matrix summarizing RDF classification results over all 8 classes is given in Table 4, showing the performance of this method.

5.3. Classification results with watershed segmentation and RDF classifier

To test whether the classification result mainly benefits from the mean shift segmentation method, not the feature sets we use, we also employ another unsupervised segmentation method, i.e. watershed algorithm (Vincent and Soille, 1991), to segment the facade images.

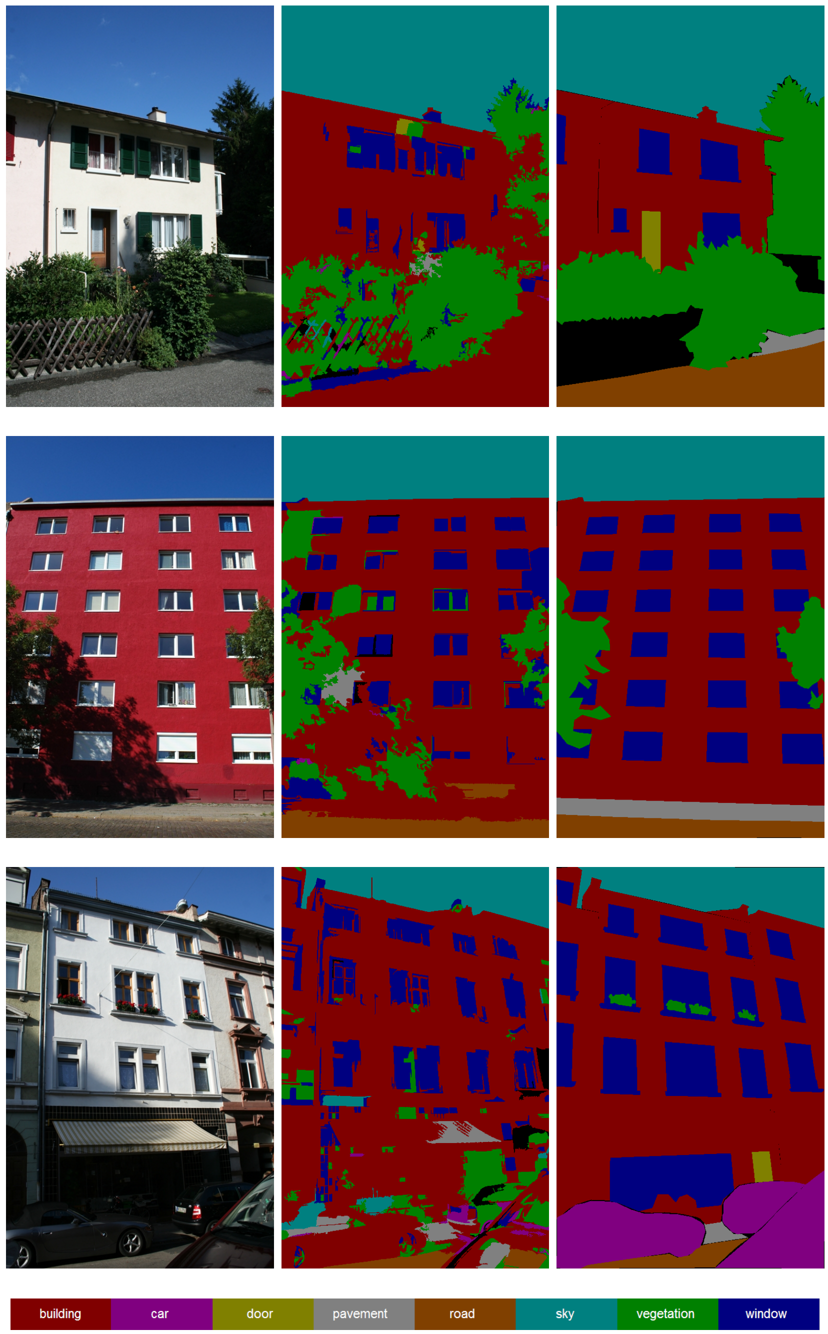


Figure 6: Example images of eTRIMS database and classification results based on a randomized decision forest. (*Left*: test image, *middle*: result, *right*: ground truth.)

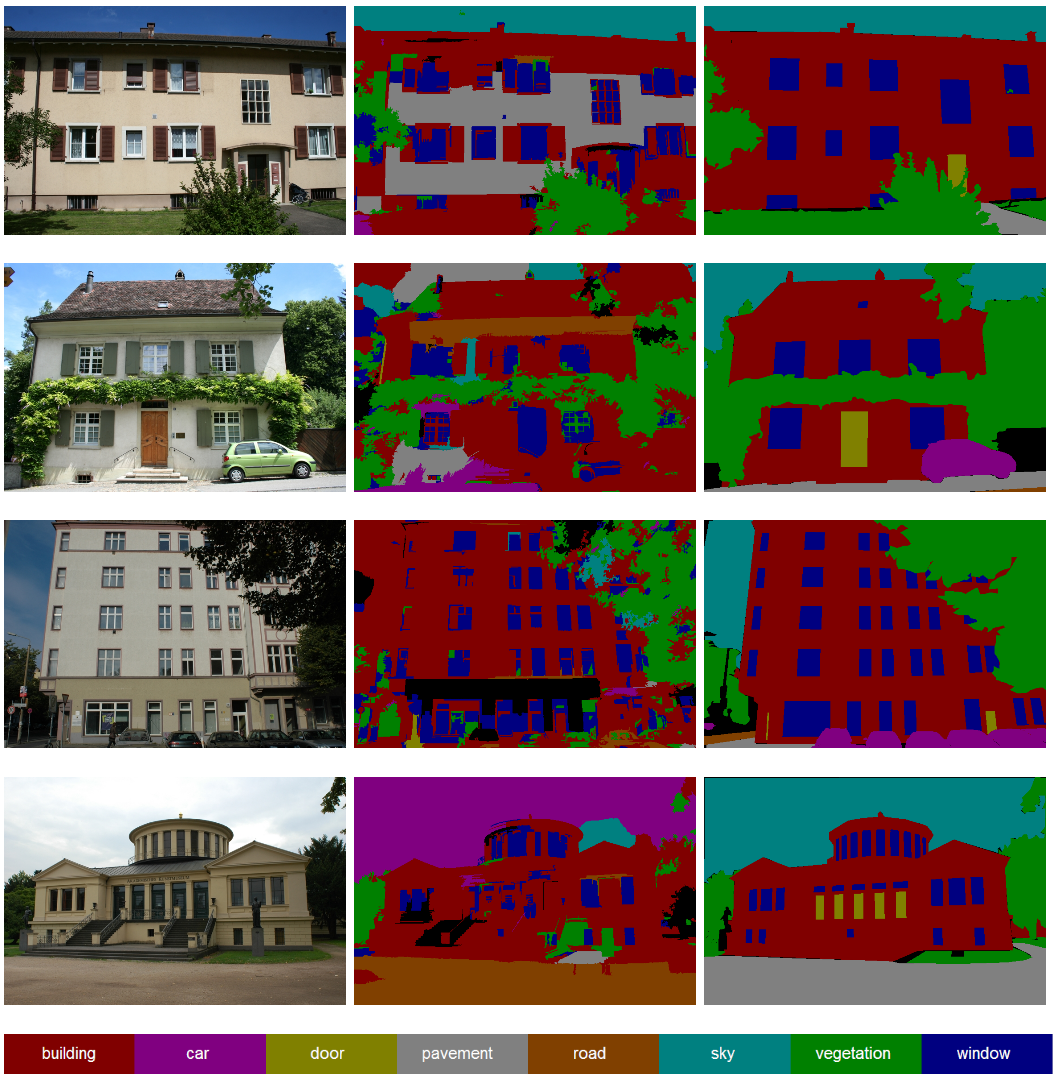


Figure 7: Example images of eTRIMS database and classification results based on a randomized decision forest. (*Left*: test image, *middle*: result, *right*: ground truth.)

The overall classification accuracy is 60.3%, with RDF classifier on all the feature sets and the number of decision trees chosen as $T = 250$. Fig. 8 shows the classification results over all 8 classes. In comparison with Fig. 4, accuracy for each class remains similar, which shows that feature sets are robust to produce good classification results.

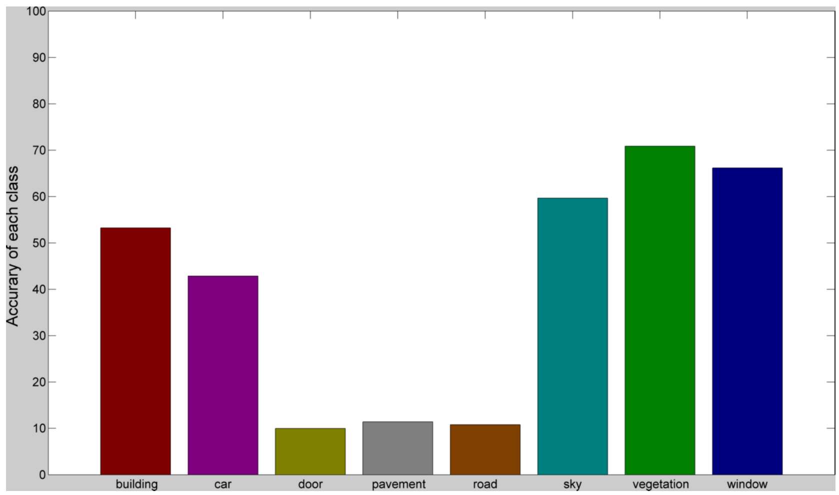


Figure 8: **Accuracy of each class**. All feature sets $\boldsymbol{f}_1, \boldsymbol{f}_2, \boldsymbol{f}_3, \boldsymbol{f}_4, \boldsymbol{f}_5, \boldsymbol{f}_6$ are used, with each class is normalized to 100.

5.4. Discussion

With respect to the three most important classes *building*, *window*, and *vegetation*, we are satisfied with our classification results. But our multi-class approach does not perform very well for most of the other classes. Our classification scheme is faced with a dramatic inequality between the sizes of the classes. Almost 60% of the data is covered by only 2 classes, and the rest is spread over the rest classes. And for the classes like *car* and *door*, Gestalt features (Bileschi and Wolf, 2007) may play major role in a good classification performance. We also believe symmetry and repetition features are vital for classifying *window* class.

In this paper, features are extracted at local scale. Classification results are achieved from bottom up on these local features by classifiers. This factor leads to noisy boundaries in the example images. To enforce consistency, a Markov or conditional random field (Shotton et al., 2006) is often introduced for refinement, which would likely improve the performance.

6. Conclusions

We evaluate the performance of seven feature sets with respect to region-based classification of facade images. The feature sets include basic features, color features, histogram features, peucker features, texture features, and SIFT features. We use randomized decision forest (RDF) to perform the classification scheme. In our experiments on eTRIMS dataset (Korč and Förstner, 2009), we have shown that RDF produces fair classification results.

The results show that these features and a local classifier are not sufficient. As future work, we are interested in evaluating more features, such as Gestalt features (Bileschi and Wolf, 2007) and other descriptor features (van de Sande et al., 2010), for building facade images. In order to recover more precise boundaries, we will put our current work into conditional random field framework (Shotton et al., 2006) by including neiboring region information in the pairwise potential of the model, which allows us to reduce misclassification that occurs near the edges of objects.

References

Amit, Y., Geman, D., 1997. Shape quantization and recognition with randomized trees. Neural Computation 9 (7), 1545–1588.

Barnard, K., Duygulu, P., Freitas, N. D., Forsyth, D., Blei, D., Jordan, M., 2003. Matching words and pictures. Journal of Machine Learning Research 3, 1107–1135.

Bileschi, S., Wolf, L., 2007. Image representations beyond histograms of gradients: The role of gestalt descriptors. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 1–8.

Bosch, A., Zisserman, A., Muñoz, X., 2007. Image classification using random forests and ferns. In: IEEE International Conference on Computer Vision. pp. 1–8.

Breiman, L., 2001. Random forests. Machine Learning 45 (1), 5–32.

Comaniciu, D., Meer, P., 2002. Mean shift: A robust approach toward feature space analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (5), 603–619.

Drauschke, M., 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). pp. 447–452.

Drauschke, M., 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. pp. 257–262.

Dumont, M., Marée, R., Wehenkel, L., 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). Vol. 2. pp. 196–203.

Felzenszwalb, P. F., Huttenlocher, D. P., 2004. Efficient graph-based image segmentation. International Journal of Computer Vision 59 (2), 167–181.

Fröhlich, B., Rodner, E., Denzler, J., 2010. A fast approach for pixelwise labeling of facade images. In: International Conference on Pattern Recognition. pp. 3029–3032.

Geurts, P., Ernst, D., Wehenkel, L., 2006. Extremely randomized trees. Machine Learning 63 (1), 3–42.

Gould, S., Rodgers, J., Cohen, D., Elidan, G., Koller, D., 2008. Multi-class segmentation with relative location prior. International Journal of Computer Vision 80 (3), 300–316.

Kluckner, S., 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. pp. 233–238.

Korč, F., 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). pp. 291–296.

Korč, F., Förstner, W., 2009. eTRIMS image database for interpreting images of man-made scenes. Tech. Rep. TR-IGG-P-2009-01, `http://www.ipb.uni-bonn.de/projects/etrims_db`.

Kumar, S., Hebert, M., 2003. Discriminative random fields: a discriminative framework for contextual interaction in classification. In: IEEE International Conference on Computer Vision. pp. 1150–1157.

Lakatos, I., 1976. Proofs and Refutations. Cambridge: Cambridge University Press.

Lazaridis, G., Petrou, M., 2006. Image registration using the Walsh transform. IEEE Transactions on Image Processing 15 (8), 2343–2357.

Lepetit, V., Lagger, P., Fua, P., 2005. Randomized trees for real-time keypoint recognition. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 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., Wehenkel, L., 2005. Random subwindows for robust image classification. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 34–40.

Nister, D., Stewenius, H., 2006. Scalable recognition with a vocabulary tree. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 2161–2168.

Petrou, M., Bosdogianni, P., 1999. Image Processing: The Fundamentals. Wiley.

Rottensteiner, F., Trinder, J., Clode, S., 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., Schiele, B., 2008. Hierarchical support vector random fields: joint training to combine local and global features. In: European Conference on Computer Vision, Part II. pp. 527–540.

Shotton, J., Johnson, M., Cipolla, R., 2008. Semantic texton forests for image categorization and segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 1–8.

Shotton, J., Winnand, J., Rother, C., 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. pp. 1–15.

Sivic, J., Russell, B. C., Efros, A. A., Zisserman, A., Freeman, W. T., 2005. Discovering objects and their locations in images. In: IEEE International Conference on Computer Vision. pp. 1–8.

Teboul, O., Simon, L., Koutsourakis, P., Paragios, N., 2010. Segmentation of building facades using procedural shape priors. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 3105–3112.

van de Sande, K. E. A., Gevers, T., Snoek, C. G. M., 2010. Evaluating color descriptors for object and scene recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 32 (9), 1582–1596.

Vedaldi, A., Fulkerson, B., 2008. Vlfeat: An open and portable library of computer vision algorithms. http://www.vlfeat.org/.

Verbeek, J., Triggs, B., 2007. Region classification with markov field aspect models. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 1–8.

Vincent, L., Soille, P., 1991. Watersheds in digital spaces: An efficient algorithm based on immersion simulations. IEEE Transactions on Pattern Analysis and Machine Intelligence 13 (6), 583–598.