eTRIMS Image Database for Interpreting Images of Man-Made Scenes

Filip Korč and Wolfgang Förstner

{filip.korc,wf}@uni-bonn.de

TR-IGG-P-2009-01

April 1, 2009

Technical Report Nr. 1, 2009

Department of Photogrammetry
Institute of Geodesy and Geoinformation
University of Bonn

Available at

http://www.ipb.uni-bonn.de/projects/etrims_db/

eTRIMS Image Database for Interpreting Images of Man-Made Scenes

Filip Korč and Wolfgang Förstner
{filip.korc,wf}@uni-bonn.de

Abstract

We describe ground truth data that we provide to serve as a basis for evaluation and comparison of supervised learning approaches to image interpretation. The provided ground truth, the eTRIMS[1] Image Database, is a collection of annotated images of real world street scenes. Typical objects in these images are variable in shape and appearance, in the number of its parts and appear in a variety of configurations. The domain of man-made scenes is thus well suited for evaluation and comparison of a variety of interpretation approaches, including those that employ structure models. The provided pixelwise ground truth assigns each image pixel both with a class label and an object label and offers thus ground truth annotation both on the level of pixels and regions. While we believe that such ground truth is of general interest in supervised learning, such data may be of further relevance in emerging real world applications involving automation of man-made scene interpretation.

[1]Name stems from supporting EU Project eTRIMS - eTraining for Interpreting Images of Man-Made Scenes. 2006 - 2009

1 Introduction

Real world applications ranging from 3D city modeling over large scale updating of geographic information systems to vision based outdoor navigation in urban environments aim to exploit large amounts of available image data. Such applications benefit from automation of man-made scene interpretation and motivate variety of approaches to the task. We describe ground truth data that we provide for training, validating and testing of supervised learning approaches to such image interpretation tasks. The provided ground truth, the eTRIMS Image Database, is a collection of annotated images of street scenes captured in European cities. Several example images are shown in the first and the third column in Figure 3.

Typical objects (e.g., buildings) in these images exhibit considerable variations in both shape and appearance and represent a challenge for a variety of object based approaches to interpretation. Further, if we regard objects as being composed of parts, we realize that the part number (e.g., the number of doors) is variable. A possible part-based recognition approach should be able to deal with such stochastic phenomena. Last, we notice that these parts appear in a variety of configurations (e.g., window arrays with variable number of rows and columns and with variable spacing between them). For this reason, the data is also suited for the evaluation of approaches that employ structure models to guide the interpretation process. The domain of man-made scenes renders thus a challenging task for variety of recognition approaches.

Pixelwise ground truth assigns each image pixel both with a class label and an object label and provides thus ground truth annotation both on the level of pixels and regions. This allows for learning on both the level of pixels and object regions.

2 Database Content

We summarize the content of the eTRIMS Image Database v1 released in March 31, 2009. The database is comprised of two datasets, the 4-Class eTRIMS Dataset with 4 annotated object classes and the 8-Class eTRIMS Dataset with 8 annotated object classes. Currently, there are 60 annotated images in each of the datasets.

Table 1: Statistics of the 4–Class eTRIMS Dataset from the eTRIMS Image Database v1 (March 31, 2009).

Class Name	**Images**	**Objects**
Building	60	142
Pavement/Road	60	127
Sky	60	71
Vegetation	56	194
Total	60	534

2.1 4–Class Dataset

In the 4–class dataset, we consider the following four object classes:

- sky
- building
- vegetation
- pavement/road

In total, there are 534 annotated objects in the dataset. Summary of the number of objects and images for each annotated class is given in Table 1.

2.2 8–Class Dataset

In the 8–class dataset, we consider the following eight object classes:

- sky
- building, window, door
- vegetation
- car, road, pavement

Table 2 summarizes the number of objects and images for each annotated class. In total, there are 1702 annotated objects in the dataset.

Table 2: Statistics of the 8–Class eTRIMS Dataset from the eTRIMS Image Database v1 (March 31, 2009).

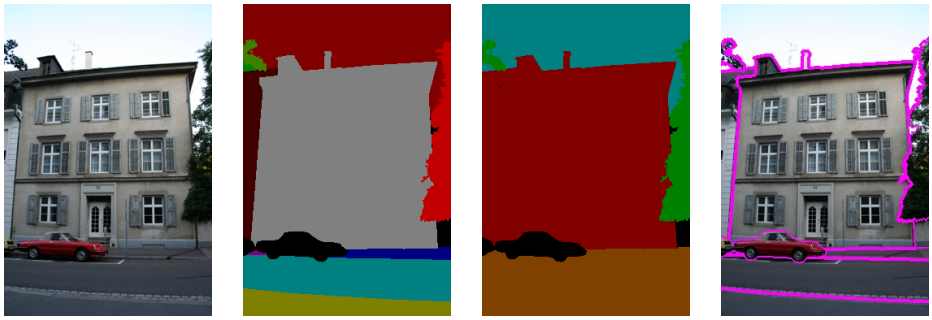
Class Name	**Images**	**Objects**
Building	60	142
Car	27	67
Door	53	85
Pavement	56	76
Road	49	51
Sky	60	71
Vegetation	56	194
Window	60	1016
Total	60	1702

3 Ground Truth

The database is comprised of images and the corresponding ground truth. Ground truth is created by human interpretation of the images, it refers to the appearance of the objects in the images, not to their 3D-structure. Therefore occluded parts of an object are not annotated as part of an object. The ground truth, each consisting of object and class segmentation, is described in the following.

3.1 Object Segmentation

Ground truth object segmentation assigns each pixel to either one of the annotated objects or background. We represent the object segmentation as an indexed image that consists of an array and a colormap matrix. Here, the pixel values $1, 2, 3, \ldots$ in the array correspond to the first, second, third object etc. The pixel value 0 corresponds to background. The pixel values in the array of an indexed image are direct indices into the colormap and allow convenient visualization the example of which can be seen in Figure 1b and 2b.



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

Figure 1: Example image from the 4–Class eTRIMS Dataset. (a) Training image. (b) Ground truth object segmentation. (c) Ground truth class segmentation showing *building, pavement/road, vegetation, sky* and *background* labels. (d) Visualization of ground truth object boundaries.

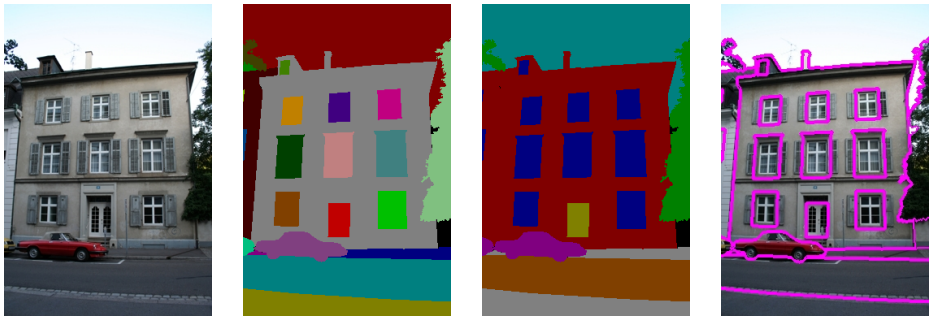
3.2 Class Segmentation

Ground truth class segmentation labels each pixel with the ground truth class or background. Again, we represent the ground truth as an indexed image. Here, the pixel values $1, 2, 3, \ldots$ correspond to class names in alphabetical order (1=building, 2=pavement/road, 3=sky, 4=vegetation in the 4–class dataset or 1=building, 2=car, 3=door, 4=pavement, 5=road, 6=sky, 7=vegetation , 8=window in the 8–class dataset). Again, the pixel value 0 corresponds to background. Class segmentation examples from the 4–class dataset and the 8–class dataset are respectively shown in Figure 1c and 2c.

More examples of training images and ground truth class segmentations for the 4–class dataset and the 8–class dataset are respectively shown in Figure 3 and 4.

3.3 Accuracy of Object Boundaries

The ground truth segmentations are not pixel accurate. Auxiliary visualization of the object boundaries, see Figure 1d and 2d, that we provide for each image may give user an impression.



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

Figure 2: Example image from the 8–Class eTRIMS Dataset. (a) Training image. (b) Ground truth object segmentation. (c) Ground truth class segmentation showing *building, car, door, pavement, road, sky, vegetation, window* and *background* labels. (d) Visualization of ground truth object boundaries.

3.4 Note

Data format of the eTRIMS Image Database adheres to the VOC 2008 [2] Segmentation Taster data format. Here, however, we do not make use of the *void* label indicating that the contained pixels can be any class including background.

4 Comparison with existing datasets

We compare the 8–Class eTRIMS Dataset with five existing annotated datasets that provide annotation more accurate than a bounding box: LabelMe [4], Caltech-101 [3], MSRC[5], VOC2008 Segmentation Taster dataset [2] and CBCL-Streetscenes [1]. Table 3 extends the summary of these datasets given in [4].

LabelMe, Caltech-101 and CBCL-Streetscenes provide object annotations in the form of polygonal boundaries. VOC2008 Segmentation Taster and MSRC provide annotations in the form of segmentation masks.

Table 3: Summary of annotated datasets that provide annotation more accurate than a bounding box. For the LabelMe dataset, the number of object classes with at least 30 annotated examples is given.

Dataset	Classes	Images	Objects	Annotation Type
LabelMe	183	30369	111490	Polygons
Caltech–101	101	8765	8765	Polygons
MSRC	23	591	1751	Region masks
VOC2008 Segm.	20	1023	2369	Region masks
CBCL-Streetscenes	9	3547	27666	Polygons
eTRIMS–8	8	60	1702	Region masks

Acknowledgment

The first author was supported by the EC Project FP6-IST-027113 eTRIMS. Further, we gratefully acknowledge members of the eTRIMS consortium, namely: Jan Čech, Martin Drauschke, Ondřej Drbohlav, Johannes Hartz, Daniel Heesch, Lothar Hotz, Mohammad Jahangiri, Arne Kreutzmann, Radim Šára, Hanns-F. Schuster, Jan Šochman, Robby T. Tan, Kasim Terzić and Susanne Wenzel, who contributed to the database by collecting data, providing annotations and supervising students. We thank all members of the eTRIMS consortium for valuable discussions and feedback. We further thank our students who helped programming the infractucture, namely David Schneider and Jens Uwe Michael Lange. Eventually, we wish to acknowledge the numerous other students who spent much time completing the annotation.

References

[1] S. Bileschi. CBCL streetscenes. Technical report, MIT CBCL, 2006.

[2] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge 2008 (VOC2008). http://pascallin.ecs.soton.ac.uk/challenges/VOC/voc2008/index.html.

[3] Li Fei-Fei, Rob Fergus, and Pietro Perona. One-shot learning of object categories. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 28(4):594–611, 2006.

[4] Bryan C. Russell, Antonio Torralba, Kevin P. Murphy, and William T. Freeman. Labelme: A database and web-based tool for image annotation. *Int. J. Comput. Vision*, 77(1-3):157–173, 2008.

[5] J. Winn, A. Criminisi, and T. Minka. Object categorization by learned universal visual dictionary. In *IEEE International Conference on Computer Vision*, pages 1800–1807, 2005. MSRC dataset reference.

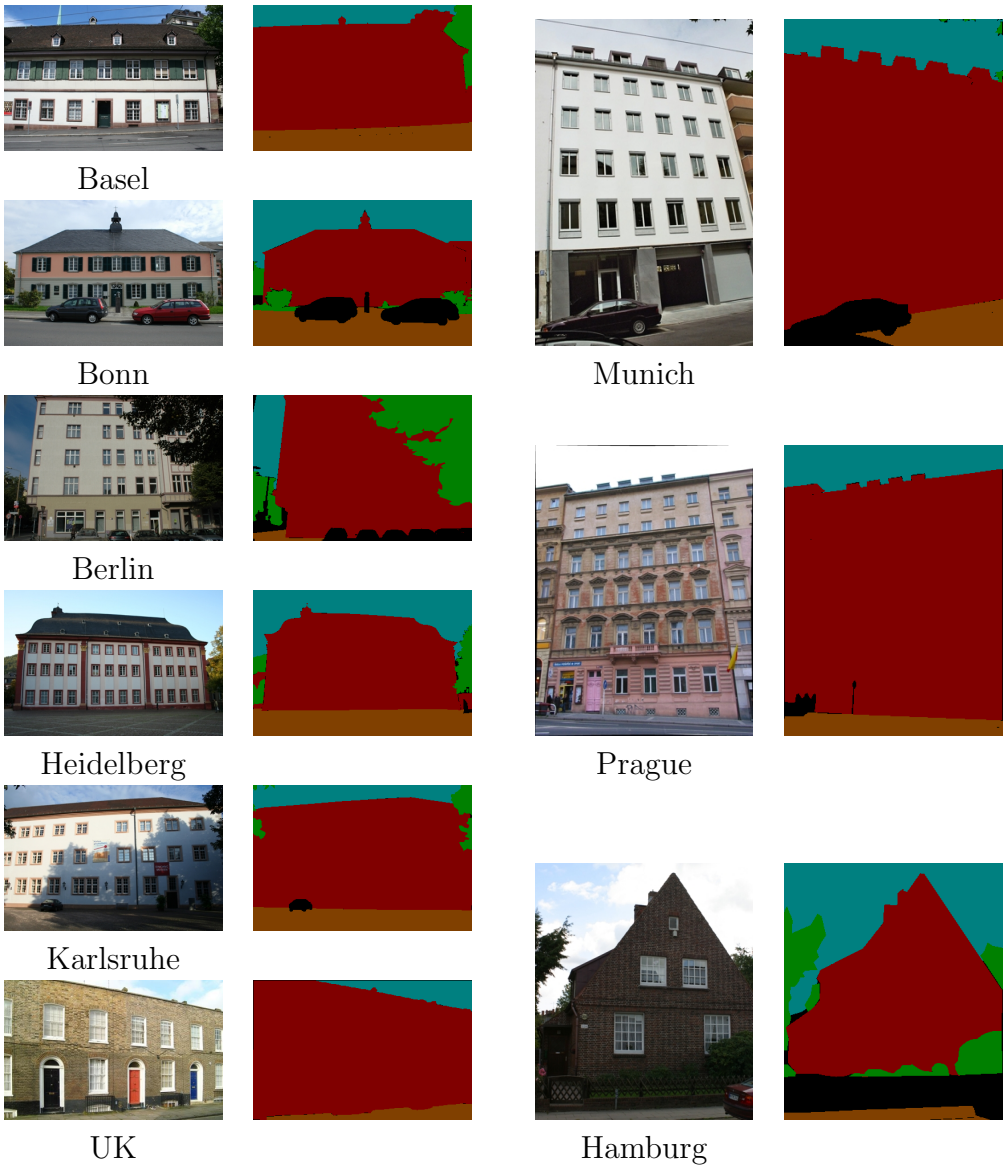


Figure 3: Example images from the 4–Class eTRIMS Dataset. Column 1 and 3 show training images. Column 2 and 4 show ground truth class segmentation with *building, pavement/road, vegetation, sky* and *background* labels. City names of origin are given below training images.

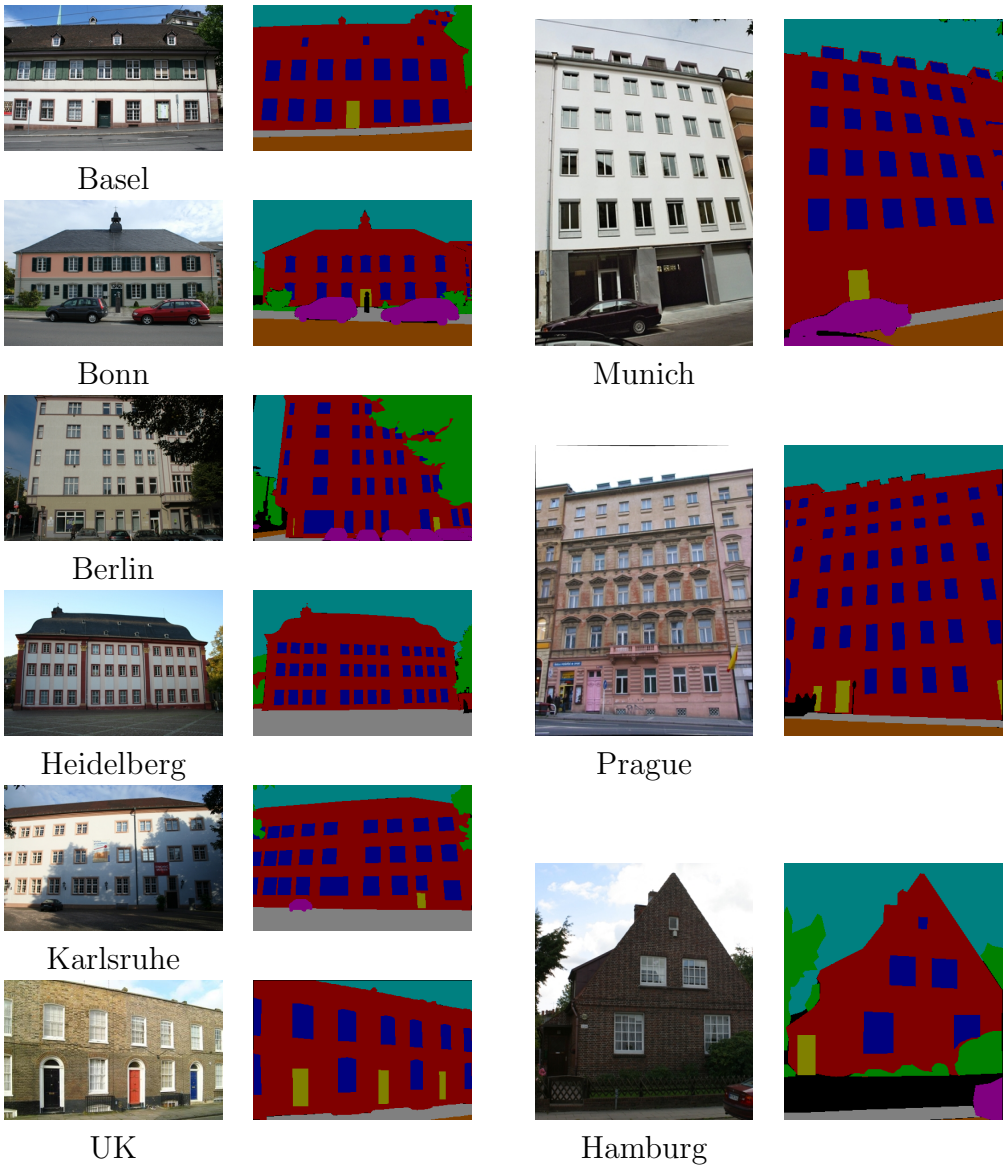


Figure 4: Example images from the 8–Class eTRIMS Dataset. Column 1 and 3 show training images. Column 2 and 4 show ground truth class segmentation with *building, car, door, pavement, road, sky, vegetation, window* and *background* labels. City names of origin are given below training images.