2011 Canadian Conference on Computer and Robot Vision

# Place Classification Using Visual Object Categorization and Global Information

Pooja Viswanathan, Tristram Southey, James Little, Alan Mackworth
Department of Computer Science
University of British Columbia
Vancouver, Canada
e-mail: poojav@cs.ubc.ca, tristram@cs.ubc.ca, little@cs.ubc.ca, mack@cs.ubc.ca

*Abstract*— **Places in an environment are locations where activities occur, and can be described by the objects they contain. This paper discusses the completely automated integration of object detection and global image properties for place classification. We first determine object counts in various place types based on LabelMe images, which contain annotations of places and segmented objects. We then train object detectors on some of the most frequently occurring objects. Finally, we use object detection scores as well as global image properties to perform place classification of images. We show that our object-centric method is superior and more generalizable when compared to using global properties in indoor scenes. In addition, we show enhanced performance by combining both methods. We also discuss areas for improvement and the application of this work to informed visual search. Finally, through this work we display the performance of a state-of-the-art technique trained using automatically-acquired labeled object instances (i.e., bounding boxes) to perform place classification of realistic indoor scenes.**

***Keywords-place classification; object recognition; scene recognition***

## I. INTRODUCTION

Scene understanding is a challenging and important problem in computer vision and robotics. Unlike rooms that are defined by geometric properties of the environment (e.g., walls), places are often defined by the objects that they contain and the set of related tasks that occur within them [1]. Place classification can thus benefit greatly from the ability to recognize objects in the environment. Several researchers have attempted to classify scenes using semantic place labels such as "kitchen", "bathroom", etc., using vision- and laser-based methods [2] [3] [4] [5]. However, most of these methods rely on global properties of images and also compute local image properties using feature-based methods, but do not explicitly attempt object recognition. Recently, some systems have demonstrated the ability to find query objects (specific instances as well as generic classes) in controlled indoor environments in the Semantic Robot Vision Challenge (SRVC) [6], an international competition to evaluate embodied object recognition systems. One of the winning teams constructs object recognition models based on training imagery collected from the Internet, and employs a peripheral-foveal vision system to collect a visual survey of objects in a real environment [7]. The success of these platforms in the embodied object recognition scenario presents the opportunity to leverage object detection for higher-level environment understanding.

Recognized objects and their locations can be used to automatically label places in the environment through the use of annotated databases, as demonstrated by our spatial-semantic modeling system [8]. This system, however, assumes the ability to recognize objects perfectly and does not address the recognition problem. In this paper, we use a state-of-the-art object detection technique to demonstrate the use of object detection in place classification. In addition, we also combine detected objects with global properties of the image to further enhance performance.

We seek to provide a robotic system with the ability to understand and explore the environment in an automated and scalable fashion, without extensive effort from a system designer. To this end, we use the images present in the LabelMe database [9]: a free online data source that provides a large and growing amount of human-labeled images of indoor scenes suitable for place labeling and object recognition. The use of Internet imagery gives the system access to training data for a large number of visual classes with no extra manual effort. In addition to containing object text labels, LabelMe images also contain segmentations of objects within them, which can be used to construct accurate bounding boxes. For object detection, we use a system created by Felzenszwalb et al. [10] based on mixtures of multiscale deformable part models (DPM) to perform object detection, due to its success in the PASCAL object detection challenge. Global properties of an image are computed using "Gist" and used to perform place classification, since they achieve state-of-the-art performance on several scene categories as seen in [4]. We use Boosted Decision Trees [12] (BDTs) to combine detections as well as Gist information.

The rest of the paper is structured as follows. In section II, we review related work in place classification, and state our contributions in section III. In section IV, we present the methods used to 1) perform automated data collection, 2) determine useful objects, 3) train object detectors, and 4) classify scenes. Experiments and results are discussed in section V. We conclude with future directions for this research.

## II. RELATED WORD

The concept of labeling areas of a 2D map, such as that captured with Simultaneous Location and Mapping (SLAM) [13], with descriptive tags has most commonly been done in

978-0-7695-4362-8/11 $26.00 © 2011 IEEE
DOI 10.1109/CRV.2011.8

topological mapping. Topological maps describe a location as a set of "places" and "connections". Kuipers [14] proposes the Spatial Semantic Hierarchy, where space is represented at many levels that contain different degrees of detail and semantic information. In work by Ranganathan et al. [15] graph-like maps are constructed where nodes are classified using visual object recognition. Kröse et al. have developed a series of practical systems [16] [17] in which the visual similarity between images is used to cluster regions in the environment. Place labels for the clusters, however, are provided by a human through speech.

For classifying scenes, Torralba et al. [18] [11] use global properties of a scene (Gist) to classify scenes. Pronobis et al. [19] [5] [20] combine Composed Receptive Field Histograms, SIFT and laser data to perform scene classification in indoor environments, under different illumination conditions. Local regions are used to infer an intermediate "theme" of an image in [21] to aid in scene classification. Several other context- and region-based approaches have been implemented, and can be found in [22]. The authors of [4] find that most methods that achieve state-of-the-art performance in classification of outdoor scenes perform significantly worse on indoor scenes. They observe that some indoor scenes are better described by the objects in them and thus combine global and local properties (Gist and spatial pyramid of visual words) to achieve increased performance. However, the reported multi-class average precision rates for the indoor dataset are still low.

Object-based methods have also been used for scene classification, as in [23], where places and functional regions of the environment are labeled based on object occurrences. However, the main drawbacks of these methods are that they are trained on specific instances of objects and tested on the same objects under different viewpoints and lighting conditions. It remains a challenge to determine which objects are strong cues for place classification. In addition, generic object class recognition has been a challenging task in computer vision research. More recently, part-based models have shown themselves to be highly effective for detection of both rigid and deformable objects [24]. This method, however, requires a large amount of training data with segmented examples of objects.

## III. CONTRIBUTIONS

In this paper, we present a novel method of place classification that uses object detections in order to perform place labeling. We pursue object-based scene classification since we believe that this method is more effective for indoor scenes and generalizable to a large number of previously-unseen indoor environments.

We use object occurrence information from the LabelMe database to inform classes of useful objects for detector training. We train these detectors automatically using DPMs on readily-available segmented and annotated images in the LabelMe database. We then train a Boosted Decision Tree that uses detection scores to predict the place. We compare this method to using Gist alone on an indoor dataset. We then present another method that combines Gist as well as object detections, and show that the two types of cues, when combined, lead to better performance than when used alone. Finally, we demonstrate our system on images taken by a robot in a real home.

## IV. AUTOMATED PLACE LABELING

We have developed a system to categorize scenes based on object detectors learned from LabelMe images. Our system is composed of four separate components. Firstly, we perform fully automated data collection from LabelMe, thus facilitating the collection of training images used to recognize a large number of object categories. We compute a Count Model that represents the number of times an object is observed in each place type in the LabelMe data based on user-provided text labels. We then use images from LabelMe to train windowed object detectors for the most frequently occurring objects. Finally, we use a Boosted Decision Tree to predict the most likely place type given the detected objects in a scene. Furthermore, we show that enhanced performance can be achieved by incorporating global cues (such as Gist in this paper) into our framework.

### A. LabelMe Data collection

LabelMe [9] is an online database of user annotated images. In LabelMe, the user can annotate an object in an image by selecting a region of the image using a polygon and giving that region a label. The entire scene can also have a description contained in the filename. Fig. 1 shows a kitchen scene from LabelMe with several labeled objects. We use LabelMe in two ways. Training of object detectors requires tight bounding boxes that we can acquire using the LabelMe polygons. Our Count Model is computed using the correspondence between labels of objects in an image and the place name descriptor found in the image filename. Note that in creating this model, we do not directly analyze the images in the dataset, and instead focus on the textual annotations in each image.

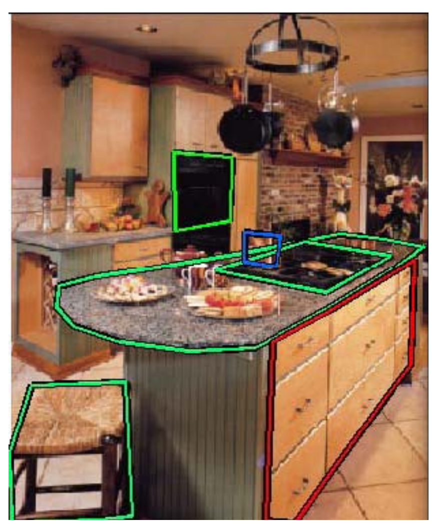
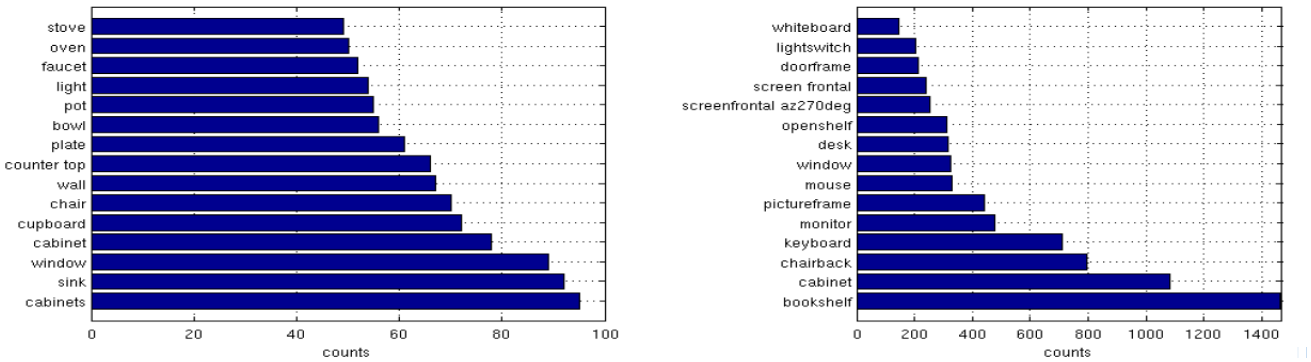


Figure 1. A kitchen scene from the LabelMe database. The polygons used to segment objects in the scene are shown as colored lines.




(a) Kitchens (b) Offices

Figure 2. Counts of the types of objects found in kitchen and office scenes

### *B. Count Model*

In order to perform place classification based on objects, we first need to learn a model of objects and their number of occurrences in each place type. We can obtain this information from the LabelMe database by querying for scenes and recording the number of annotated occurrences of each object in the scene, as in Vasudevan et al. [23] and [8]. The counts table $ct_p(o)$ contains the number of times object $o$ occurs in images of place type $p$. If the number of images of place type $p$ is $n_p$, the likelihood of observing object $o$ in place $p$ is computed as

$$P(o|p) = ct_p(o)/n_p \tag{1}$$

We refer to this likelihood as the Count Model, which is used to inform detector training and learning of the Place Model described below.

### *C. Useful Objects*

The most informative objects are often the most frequently occurring, in the domain of place labeling. In addition, since most LabelMe scenes contain objects in realistic home settings, objects that have high counts in the learned model are more likely to be present in the intended test environment. Thus, we compute the most frequently occurring objects in each place type to train object detectors. A histogram of these objects for some place types can be found in the Fig. 2.

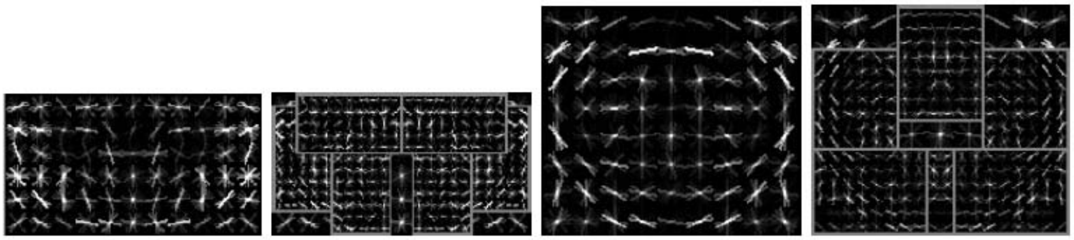
### *D. Detector Training*

To train object detectors, we use the mixtures of multiscale DPM described in [10]. The DPM models both the entire object as well as its parts. It learns the number of parts comprising an object, the positions of these parts, and the variation in positions. The system employs a margin-sensitive technique for data-mining hard negative examples to improve classification. The underlying model is called latent SVM, which is a reformulation of MI-SVM [26] in terms of latent variables. The approach alternates between fixing latent values for positive examples and optimizing the SVM object function.

The approach described above must be trained on images of the target objects with accurate bounding boxes, making many conventional data sources unusable. However, LabelMe provides user-defined bounding polygons that we use to determine training image bounding boxes. We optimized the parameters of the DPM to train each detector in less than three hours on machines with 2 Intel quad-core Xeon 3.2 GHz processors with 32GB of memory. We trained detectors for a subset of the most frequently occurring objects based on the object counts. A total number of 61 objects were used (corresponding to approximately 15 objects in each place type). The precision-recall rates for a few categories, as well as visualizations of a detector model can be found in the Experiments section.

### *E. Place Labeling using Boosted Decision Trees*

Decision trees are a hierarchical model for supervised learning that identify local regions in the input space using a sequence of recursive splits. Boosting is a learning technique that determines how to combine and weight many weak classifiers to produce a single strong classification result. Adaboost (adaptive boosting) by Freund and Schapire [27] was an improvement on earlier boosting techniques that built classifiers iteratively and adapted them based on the performance of previous iterations. The Adaboost algorithm focuses on so called “hard to solve” samples from the training distribution by increasing weights on samples incorrectly labeled by previous iterations and decreasing weights on those that were successfully classified. Significantly, boosted decision trees [12] require no knowledge about the properties of the weak classifiers used and can be combined with any weak classifier that is more accurate than random guess, allowing us to test a wide variety of features. Also, they are not prone to over-fitting and have no parameters to tune except for the number of rounds they run for and can usually just be run until the test accuracy (on a hold-out set) plateaus. To achieve multiclass classification, we trained a decision tree for each place type using a 1 vs. all approach. We describe the inputs to the boosted decision trees in section V since they vary between experiments.



(a) bowl side view (b) bowl side view parts (c) bowl top view (d) bowl top view parts

Figure 3. Visualizations of the Felzenszwalb et al. object classifier. Images a and c show the expected intensity of gradients in a grid pattern for the entire object. Images b and d show the gradients in the parts model.

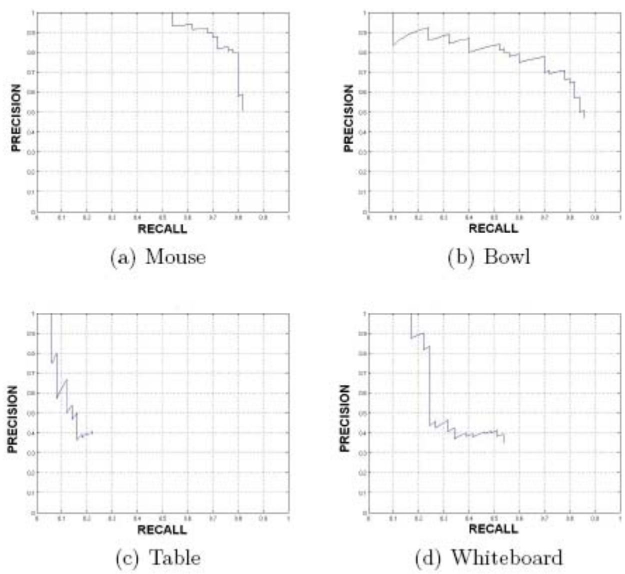



Figure 4. The precision/recall rates of object detectors. The top row shows 2 of the most successful classifiers and the bottom row shows 2 of the least successful classifiers.

## V. EXPERIMENTS

In this paper, we attempt to classify kitchens, offices, bathrooms and bedrooms. However, due to the automated nature of our system, we can easily learn models for other types of places including living rooms and dining rooms by simply querying LabelMe for more scenes.

### *A. Count Model*

Fig. 2 shows the object counts learned for kitchens and offices. We display the 15 most frequently occurring objects in each place type. As seen in the figures, some of the objects have unusable labels due to ambiguous user entries. We thus select a limited number of the objects, and show later on that these are in fact sufficient for the task of place classification. Future work could involve language processing to eliminate ambiguous labels as well as combine synonymous labels.

### *B. Detection*

Fig. 3 provides visualizations of the bowl detector using the DPM. We trained detectors using at most 600 positive examples, and 1000 negative examples for each class. We set the number of components of the mixture model, *n*, based on the size of the training data for each class (classes with few training examples were trained on 1 component, while classes with more training data were trained on up to 3 components). Thus, training examples are split into *n* components based on the aspect ratio of the bounding boxes they contain. DPMs are trained on each component individually and merged together to form the final model. In order to produce precision-recall and average precision rates for each category, we validated the models on images of LabelMe objects that were not used in training. We used loosely cropped versions of these images to prevent unannotated true positive examples in an image from being detected as false positives. Fig. 4 shows some of the most and least successful detection results. As seen, objects that are usually fairly obscured by other objects (furniture such as desks and tables) tend to perform the worst, because the training images for these classes mostly contain views of cluttered table/desk tops.

### C. *Place Classification*

We designed experiments to test place classification in three different scenarios. In the first experiment, we attempt classification of places based on perfect labels of all annotated objects. In the second, we classify full images that depict a scene containing different types and numbers of objects, using real detection results. Finally, we evaluate place classification on images taken by a robot in a real home.

#### 1) *Place Classification with Perfect Labels*

In this experiment, we want to determine the performance of the place classification system using BDTs if all objects can be recognized perfectly, that is, we used place names and object names from annotated objects in LabelMe. We run 10-fold-cross validation on all examples of bedrooms (37), bathrooms (52), offices (647) and kitchens (190). The number of examples for each place type is indicated in brackets. Only objects that occur in at least 10 images are used as features in the BDTs. In this experiment, inputs are binary, indicating the presence or absence of an object. As seen in Table 1, our place classification algorithm produces assignments that closely match the ground-truth place labels for all place types, and outperform our weighted voting method used in [8]. Bathrooms produce the lowest recall rate due to the limited number of example images currently in LabelMe. This demonstrates that objects present in a scene are very useful in place classification.

#### 2) *Place Classification on Indoor Dataset*

Our second experiment determines place labels using real, noisy objects detections in the image. These images only contain a portion of the scene, and can contain few to many objects. The highest SVM detector scores, produced by running each learned object detector on the image, are fed as input to the BDT, which then infers the place label. We found that using a higher number of detections per object class resulted in negligible improvement, and thus use only the top detection for each object class in this paper.

TABLE I. CLASSIFICATION RESULTS FOR PLACE CLASSIFICATION WITH PERFECT OBJECT LABELS.

| **Place** | **Precision** | **Recall** |
|---|---|---|
| Bathroom | 0.96 | 0.88 |
| Bedroom | 0.92 | 0.92 |
| Kitchen | 0.95 | 0.93 |
| Office | 0.98 | 1.0 |
| Average | 0.95 | 0.93 |

We expanded the indoor dataset in [4] with 50 more images for each place type acquired from Photobucket to minimize overlap between the Torralba dataset (which uses LabelMe as a data source) and the images used in detector training. Our new database will be made available online soon. We compare our object-centric method with the technique in [11], which uses Gist for place classification using an SVM. We perform 10-fold cross-validation on the entire dataset for both methods. Results of place classification on the combined indoor dataset are shown in Table 2, in columns 1 and 2. Given the difficulty of the task, our model performs extremely well at distinguishing between the various place types based on a relatively small set of trained detectors. This demonstrates object detections using DPM with our automated training method are reliable for the task of place classification. We also notice that our object-centric method is comparable to Gist (columns 3 and 4), outperforming it in most cases.

We also show results of combining object detections and Gist in the fifth and sixth columns. The input features in the BDT in this experiment are the highest SVM object detector scores as well as SVM output from Gist. We see that due to the complementary strengths of both methods, combining them results in enhanced performance for all place types in the indoor dataset.

Finally, we assess the performance of the system as we vary the number of detectors. In Table 3, we report the precision and recall rates when the top 45, 30, and 15 detectors of 61 are used. The results show that place classification requires a relatively small number of object detectors; however, using more detectors can improve performance.

TABLE II. PLACE CLASSIFICATION RESULTS USING OBJECT DETECTIONS, GIST AND BOTH COMBINED ON IMAGES ACQUIRED BY HUMANS

| **Place** | **Object** | | **Gist** | | **Object+Gist** | |
|---|---|---|---|---|---|---|
| | ***Prec.*** | ***Recall*** | ***Prec.*** | ***Recall*** | ***Prec.*** | ***Recall*** |
| Bathroom | 0.56 | 0.58 | 0.63 | 0.56 | 0.68 | 0.63 |
| Bedroom | 0.57 | 0.53 | 0.53 | 0.53 | 0.57 | 0.64 |
| Kitchen | 0.61 | 0.62 | 0.53 | 0.57 | 0.64 | 0.67 |
| Office | 0.58 | 0.60 | 0.53 | 0.52 | 0.66 | 0.60 |
| Average | 0.58 | 0.58 | 0.56 | 0.55 | 0.63 | 0.64 |

TABLE III. PLACE CLASSIFICATION RESULTS WITH VARYING NUMBER OF DETECTORS

| **Number of detectors** | **Precision** | **Recall** |
|---|---|---|
| 45 | 0.59 | 0.59 |
| 30 | 0.54 | 0.54 |
| 15 | 0.53 | 0.53 |

TABLE IV. PLACE CLASSIFICATION RESULTS USING OBJECT DETECTIONS, GIST AND BOTH COMBINED ON IMAGES ACQUIRED BY A ROBOT

| Place | Object | | Gist | | Object+Gist | |
|---|---|---|---|---|---|---|
| | *Prec.* | *Recall* | *Prec.* | *Recall* | *Prec.* | *Recall* |
| Bathroom | **0.41** | 0.87 | 0.37 | **1.0** | 0.33 | 0.87 |
| Bedroom | **0.50** | **0.47** | 0.11 | 0.07 | 0.33 | 0.40 |
| Kitchen | **1.0** | **0.07** | 0.0 | 0.0 | 0.0 | 0.0 |
| Office | **0.69** | **0.60** | 0.0 | 0.0 | 0.50 | 0.07 |
| Average | **0.65** | **0.50** | 0.12 | 0.27 | 0.29 | 0.33 |

### 3) *Place Classification on Robot Images*

In order to demonstrate the effectiveness of the system in a robotic platform, we evaluate the BDT trained on the indoor dataset above to perform place classification on images captured by a tele-operated Pioneer robot in an apartment. Since several images in the video sequence were difficult to classify even by a human due to the lack of objects in the image, we report results on a subset of frames (classifiable by a human user) captured by the robot, and assume that in the future we can propagate these labels to nearby images based on location estimates (using SLAM or odometry data), or by using visual similarity as in [28]. The results in Table 4 show that although both object-centric and Gist methods produce poor results on the data collected by the robot, the object detector is more generalizable to images that are significantly different from training data with regards to viewpoints, framing, lighting, etc. Here we see the benefit of using local properties of the image (specifically, the objects detected in it) that are more consistent across various environments rather than global properties such as Gist, which are more environment-specific. Since the BDT was trained on the indoor dataset (where Gist is found to be more reliable) the combined model results in poor performance on the test data, where Gist performs significantly worse than object-based classification.

## VI. DISCUSSION

In the future, we intend to extend this work to a larger number of place types using more object detections. One of the greatest challenges we encounter in this work is acquiring good training data from LabelMe. For many classes, such as "pot", the images in LabelMe are of different types of sub-classes (such as flower pots, ornamental pots and cooking pots). In future work, we would like to automate clustering of objects into different types/views using techniques such as comparing the differences of LabelMe polygons.

We would like to attempt place classification on data collected by a robot that can acquire 3D layouts of the environment using a panning laser range finder. This 3D data allows us to identify structures such as tables and desks and automatically segment and acquire multiple images of objects on their surface. We can also use the place labels to guide visual search of novel objects using our Location Model described in [8]. In addition, we need to investigate the use of place labels as context to enhance recognition of objects that are currently difficult to recognize. Finally, we need to incorporate spatial relationships between objects to enhance place classification and informed search.

## VII. CONCLUSION

In conclusion, we have demonstrated a system that can perform place classification using object detection on both segmented images of objects from the environment and full scene images. In addition, we have shown that, with state-of-the-art object detectors trained with large, freely available data sources like LabelMe, we can effectively both detect and classify a wide variety of objects in realistic indoor images.

## ACKNOWLEDGMENTS

We would like to thank NSERC and the GEOIDE NCE for funding this research. We would also like to thank Walter Wohlkinger for providing the robot images used in this paper.

## REFERENCES

[1] Southey, T., Little, J.J.: Object discovery through motion, appearance and shape. In: AAAI Workshop on Cognitive Robotics, Technical Report WS-06-03, AAAI Press (2006)

[2] Wu, J., Christensen, H.I., Rehg, J.M.: Visual place categorization: Problem, dataset, and algorithm. In: IROS, IEEE (2009) 4763{4770

[3] Wu, J., Rehg, J.M.: Where am I: Place instance and category recognition using spatial pact. In: Proceedings of 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. (2008)

[4] Quattoni, A., Torralba, A.B.: Recognizing indoor scenes. In: CVPR. (2009) 413-420

[5] Pronobis, A., Mozos, O.M., Caputo, B., , Jensfelt, P.: Multi-modal semantic place classification. The International Journal of Robotics Research (IJRR) **29** (2010) 298-320

[6] Efros, A., Rybski, P. Website: http://www.semantic-robot-vision-challenge.org/.

[7] Meger, D., Forss_en, P.E., Lai, K., Helmer, S., McCann, S., Southey, T., Baumann, M., Little, J.J., Lowe, D.G.: Curious george: An attentive semantic robot. Robotics and Autonomous Systems Journal, Special Issue From Sensors to Human Spatial Concepts (2008)

[8] Viswanathan, P., Meger, D., Southey, T., Little, J.J., Mackworth, A.K.: Automated spatial-semantic modeling with applications to place labeling and informed search. In: Canadian Conference on Computer and Robot Vision, IEEE Computer Society (2009) 284-291

[9] Russell, B., Torralba, A., Murphy, K., Freeman, W.: Labelme: a database and web-based tool for image annotation. International Journal of Computer Vision (special issue on vision and learning) **77** (2008) 157 - 173

[10] Felzenszwalb, P., Mcallester, D., Ramanan, D.: A discriminatively trained, multiscale, deformable part model. Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on (2008)

[11] Oliva, A., Torralba, A.: Modeling the shape of the scene: A holistic representation of the spatial envelope. Int. J. Comput. Vision **42** (2001) 145-175

[12] Freund, Y., Mason, L.: The alternating decision tree learning algorithm. In: ICML '99: Proceedings of the Sixteenth International Conference on Machine Learning, San Francisco, CA, USA, Morgan Kaufmann Publishers Inc. (1999) 124-133

[13] Choset, H., Nagatani, K.: Topological simultaneous localization and mapping (slam): Toward exact localization without explicit

localization. IEEE Transactions on Robotics and Automation **17** (2001) 125 - 137

[14] Kuipers, B.: The spatial semantic hierarchy. Arti_cial Intelligence **119** (2000) 191 - 233

[15] Ranganathan, A., Dellaert, F.: Semantic modeling of places using objects. In: Proceedings of Robotics: Science and Systems (RSS). (2007)

[16] Krose, B., Booij, O., Zivkovic, Z.: A geometrically constrained image similarity measure for visual mapping, localization and navigation. In: Proceedings of the 3rd European Conference on Mobile Robots, Freiburg, Germany (2007) 168 - 174

[17] Spexard, T., Li, S., Wrede, B., Fritsch, J., Sagerer, G., Booij, O., Zivkovic, Z., Terwijn, B., Krose, B.: Biron, where are you? - enabling a robot to learn new places in a real home environment by integrating spoken dialog and visual localization. In: Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, IEEE (2006)

[18] Torralba, A., Murphy, K., Freeman, W., Rubin, M.: Context-based vision system for place and object recognition. In: International Conference on Computer Vision. (2003)

[19] Pronobis, A., Caputo, B., Jensfelt, P., Christensen, H.I.: A discriminative approach to robust visual place recognition. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS06), Beijing, China (2006)

[20] Pronobis, A., Martinez Mozos, O., Caputo, B.: SVM-based discriminative accumulation scheme for place recognition. In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA08), Pasadena, CA, USA (2008)

[21] Li, F.F., Perona, P.: A bayesian hierarchical model for learning natural scene categories. In: CVPR '05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 2, Washington, DC, USA, IEEE Computer Society (2005) 524-531

[22] Bosch, A., Mu noz, X., Marti, R.: Review: Which is the best way to organize/classify images by content? Image Vision Computation **25** (2007) 778-791

[23] Vasudevan, S., Siegwart, R.: Bayesian space conceptualization and place classification for semantic maps in mobile robotics. Robotics and Autonomous Systems **56** (2008) 522 - 537

[24] Crandall, D., Felzenszwalb, P., Huttenlocher, D.: Spatial priors for part-based recognition using statistical models. In: CVPR '05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 1, Washington, DC, USA, IEEE Computer Society (2005) 10-17

[25] Yang, Y., Pedersen, J.O.: A comparative study on feature selection in text categorization, Morgan Kaufmann Publishers (1997) 412-420

[26] Andrews, S., Hofmann, T., Tsochantaridis, I.: Multiple instance learning with generalized support vector machines. In: Eighteenth national conference on Artificial intelligence, Menlo Park, CA, USA, American Association for Artificial Intelligence (2002) 943-944

[27] Freund, Y., Schapire, R.E.: A brief introduction to boosting. In: Sixteenth International Joint Conference on Artificial Intelligence. Volume 14. (1999) 1401-1406

[28] Booij, O., B. Krose, J. Peltason, T.S.M.H.: Moving from augmented to interactive mapping. In: Proceedings of the Robotics: Science and Systems workshop Interactive Robot Learning. (2008) 21-23