Object Categorization Robust to Surface Markings using Entropy-guided Codebook

Sungho Kim and In So Kweon
Korea Advanced Institute of Science and Technology
373-1 Guseong-dong Yuseong-gu Daejeon, Korea
shkim@rcv.kaist.ac.kr, iskweon@kaist.ac.kr

Abstract

Visual categorization is fundamentally important for autonomous mobile robots to get intelligence such as novel object acquisition and topological place recognition. The main difficulty of visual categorization is how to reduce the large intra-class variations. In this paper, we present a new method made robust to that problem by using intermediate blurring and entropy-guided codebook selection in a bag-of-words framework. Intermediate blurring can reduce the high frequency of surface markings and provide dominant shape information. Entropy of a hypothesized codebook can provide the necessary amount of repetition among training exemplars. A generative optimal codebook for each category is learned using the MDL (minimum description length) principle guided by entropy information. Finally, a discriminative codebook is learned using the discriminative method guided by the inter-category entropy of the codebook. We validate the effect of the proposed method using a Caltech-101 DB, which has large intra-class variations.

1 Introduction

Intelligent mobile robots should have visual perception capability akin to that provided by human eyes. Let's assume that a mobile robot is put in a strange house environment. It will wander the house and categorize each room as a living room, kitchen, or bathroom. Additionally, it will categorize novel objects such as the sofa, TV, dining table, or refrigerator. As we can see in this scenario, the two basic functions of an intelligent mobile robot are categorizing places and objects for automatic high-level learning about new environments. In the current state-of-the-art, topological localization remains at the level of image identification or matching for a specific environment [9, 13]. Object identification (recognition) of the same objects is almost matured due to the robustness of local invariant features such as SIFT and G-RIF [7, 14]. Currently, the categorization of general objects is an active research area in computer vision [6, 15].

Figure 1: Examples of surface markings for a cup category.

However, visual categorization is a very challenging problem due to large intra-class variations. Among many sources of them, such as geometric and photometric variations, surface markings are dominant. Fig. 1 shows several cups. Note the various surface markings at the interior regions of the cups. The effect of surface marking is much larger in man-made objects than in animals or plants due to creative design for beauty.

To our best knowledge, there has been no work published on the reduction of surface markings in object categorization. Most researchers have focused on how to minimize intra-class variations of object shape. We can summarize the current object representation approaches as shown in Fig. 2. As the strength of a geometric relation is weaker, the amount of intra-class variation is smaller. At the same time, the discrimination power is reduced. PCA can represent whole objects directly and is very weak to geometric variations [12]. The constellation model of visual parts can handle geometric variations more flexibly [6, 16]. Flexible shape samples can represent large variations of shapes [2]. Bag of words, derived from document indexing, is a very robust method to visual variation because it considers no geometrical relations [3]. Texton, which is a more generalized version of bag of words, can categorize textured regions such as sea, sky, and forest [20]. A compromise of both extremes is the implicit shape model, which assigns

IEEE Workshop on Applications of Computer Vision (WACV'07)
0-7695-2794-9/07 $20.00 © 2007 IEEE

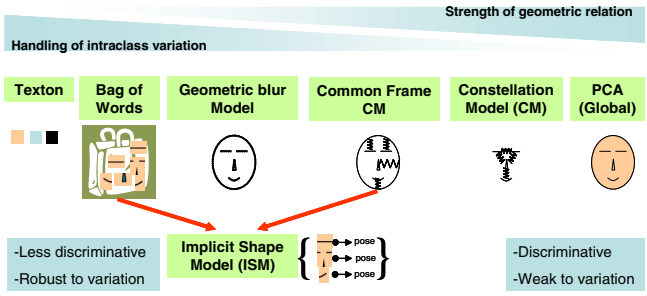


Figure 2: Summary of object representation schemes in terms of geometric strength and intra-class variation.

pose information for each codebook [11].

In this chapter, our object representation is based on the bag of words approach to take advantage of its simplicity and robustness to large geometric variations. However, we focus on reducing surface marking problems during visual word or codebook generation. Our key idea is twofold: One is to apply intermediate blurring to extract important object shape information. It is motivated from cognitive experiments showing that human visual systems can categorize blurry objects very quickly [1]. The other is based on information theory. Entropy of a hypothesized codebook among training instances should be high for surface marking reduction, and entropy among different categories should be low for discrimination.

2 Background of bag of visual words

The term visual words originated from linguistics [5]. A paragraph consists of a set of words. Likewise, we can think of a scene or an object as composed of visual words. Recently, the bag of visual words approach has shown very promising results on visual categorization problems [3, 4, 15, 20]. Although it is a very simple representation, it can handle large geometric variations because it discards geometric relationships among features. The basic steps for the bag of visual words approach are visual word generation, histogram building, and classifier learning. The key issue of the visual word-based classification is how to learn the optimal set of visual words, or codebook. Csurka et al. selected the optimal set of visual words by k-means clustering [3]. The size of k is empirically selected by cross validation of the training set. Winn et al. proposed a pairwise feature clustering method that maximizes inter-class variation and minimizes intra-class variation [20]. Previous approaches do not consider surface marking problems for optimal codebook generation.

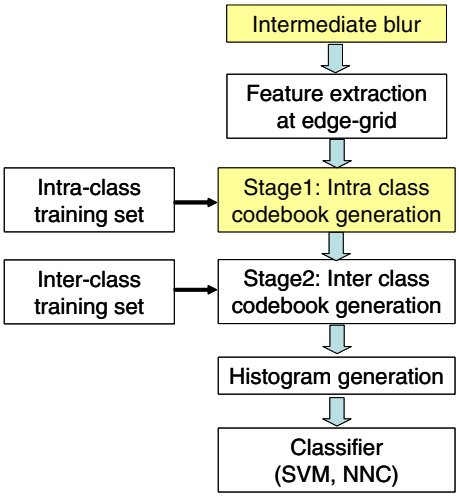


Figure 3: Overall categorization system for surface marking robustness. Intermediate blur and stage 1 blocks provide key roles for the reduction of surface markings.

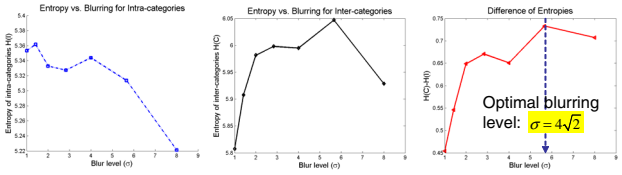
3 Robust categorization to surface markings

3.1 Overall categorization system

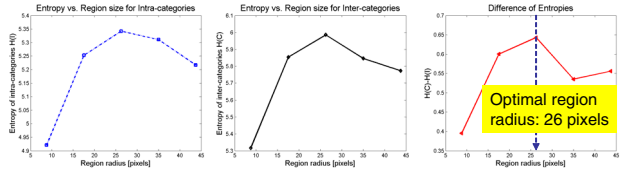
The proposed object categorization system is composed of feature extraction, codebook generation, and classification, as shown in Fig. 3. First, we extract dense features after intermediate blur. Then an intra-class codebook is learned using the model selection method of entropy-guided MDL (minimum description length) as the intra-class training set. These intra-class codebooks are further learned in a discriminative way using entropy-guided codebook selection as the inter-category training set. Then each training instance is represented by histogram using the optimal codebook learning. Finally, classification is conducted using either SVM (support vector machine) or NNC (nearest neighbor classifier) by varying distance metrics. The most important blocks for surface marking reduction are intermediate blur and stage 1. Details of the system are explained in the following sub-sections.

3.2 Information theoretic parameter selection for features

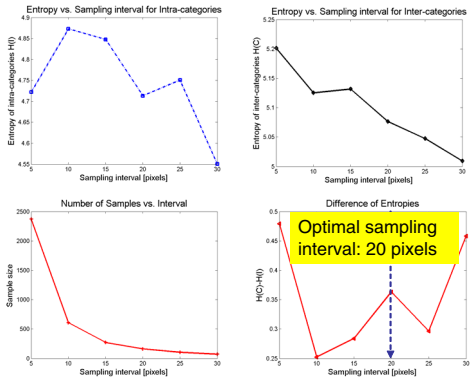
The first issue in the bag of visual words approach is how to extract local features. Direct application of sparse scale invariant features such as SIFT [14] and G-RIF [7] to Caltech-101 DB (available at http://www.vision.caltech.edu/htmlfiles/archive.html) shows very disappointing results: a 26.8% correct classification rate using 15 images for training and 15 images for testing (using Berg's evaluation method [2]). So, we need to find an opti-



(a) entropy vs. level of blurring



(b) entropy vs. level of region size



(c) entropy vs. sampling interval

Figure 4: Evaluation of feature parameters (blurring, region size, sampling interval) in terms of intra- and inter-category differences of entropy.

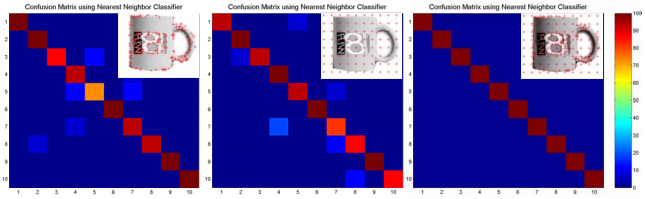


Figure 5: Evaluation (confusion matrix) of sampling type: (left) edge sampling, (middle) grid sampling, (right) edge-grid sampling. Edge-grid sampling shows better performance.

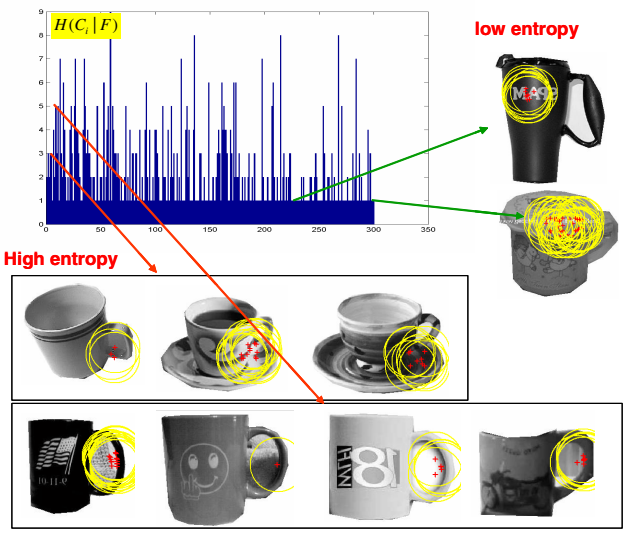


Figure 6: Observation for repeatable parts (high entropy) and surface marking parts (low entropy).

mal set of feature parameters, such as level of blur, location of sampling points, size of region, and image scale.

Motivated from the basic constraint (maxVar(inter-class)/Var(intra-class)) and entropy in information theory [17], the codebook (F) should have high entropy ($H(C|F)$) within a category , and low entropy among categories. For the evaluation, we used a PCA-GRIF descriptor (5-dim) and calculated entropy in a partitioned feature space. For a given partition $A = \{A_i\}$, entropy is $H(A) = -\sum_i p_i \log p_i$ where p_i is the relative bin count. We set the partition as 10/cell and used 10 selected categories. Due to the properties of the Caltech DB, we set the scale as fixed. The final parameter is selected at the value where the difference of inter-category entropy and intra-category entropy is maximized. Fig. 4 shows the evaluation results. According to the maximum value, we set the blurring level as $\sigma = 4\sqrt{2}$, region radius as 26 pixels, and the sampling interval at 20 pixels.

Finally, we also evaluated the edge sample and dense grid sampling types with the optimal feature parameters. The evaluation was conducted using conventional k-means clustering for codebook generation and bag of visual words framework with an NNC classifier for segmented objects. The edge-grid sampling shows upgraded performance as shown in Fig. 5. So, we used edge-grid sampling with the selected feature parameters.

3.3 Stage1: Intra class codebook generation (generative)

In stage 1, we have to minimize intra-class variations. The main cause of large intra-class variation is surface markings, which have various patterns for object instances. As shown in Fig. 6, the surface markings can be removed by finding repeatable parts or high-entropy parts.

Based on this relation of entropy and surface markings, we can conduct model selection using MDL more efficiently. The MDL criteria can provide an optimal codebook in terms of fitting distortion and model complexity, as shown by [18]

Algorithm 1 Class-specific codebook generation

```
Given:  category-specific local features
Goal:  make codebook
Step 1.  Extract edge-grid features for
each category.
Step 2.  Make initial codebook using
appearance-based clustering [8].
Step 3.  Starting from this initial
codebook.
 Evaluate MDL
 If MDL is minimum, stop.
 Else
  Remove one codebook that has lowest
entropy.  Go to 1.
```

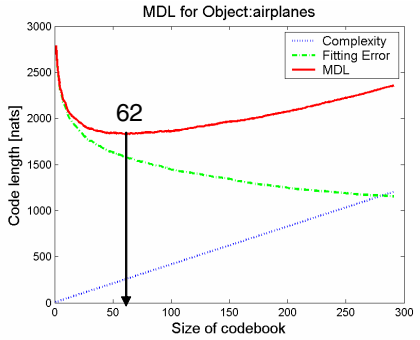
$$\hat{\Lambda}(\mathbf{X},\theta) = \arg\min\left\{-\log L(\mathbf{X},\theta) + \frac{K(V+1)}{2}\log N\right\}$$

where L is likelihood of data fitting, $\mathbf{X}$ is training features, θ is parameters (mean and var for codebook), K is the number of codebook, V is the number of parameters per codebook, and N is the number of features. Fig. 7 shows the MDL model selection curve and the properties of the selected codebook. Note that our codebook can find semantically meaningful parts for the training instances regardless of various surface markings.

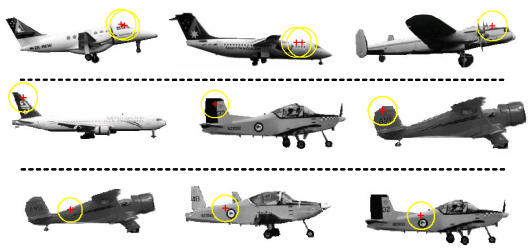
The key point for surface marking reduction is to remove codebook candidates that have low entropy. An initial codebook is generated using two steps of agglomerative clustering (bottom-up) and k-means clustering (top-down) [8]. The detailed algorithm for intra-class codebook selection is shown in Algorithm 1.

3.4 Stage2: Inter class codebook generation (discriminative)

Given the category-specific codebooks learned in stage 1, we have to select a discriminative universal codebook for bag of visual words-based classification. We can obtain a discriminative codebook (F_{opt}) by maximizing the posterior of class labels given training examples and a hypothesized universal codebook. The key point in this approach is to select a removable codebook using the inter-category entropy of a codebook that has large entropy (ambiguous codebook). If we define $\{F\}$ as a hypothesized universal codebook, I_i^c as the i-th object instance belonging to category c,and l as the category label, then the posterior can be formulated as



(a) MDL graph for airplane category



(b) Examples of selected optimal codebook

Figure 7: Entropy-guided MDL graph and its example parts corresponding to selected codebook. Note that similar parts are selected regardless of surface markings.

$$F_{opt} = \arg\max_F\{\prod_c\prod_{i\in c} p(l|I_i^c,\{F\}\}$$
$$= \arg\max_F\{log(\prod_c\prod_{i\in c} p(l|I_i^c,\{F\})\}$$

$since$

$$p(l|I_i^c) = \frac{p(I_i^c|c,\{F\})p(c,\{F\})}{\sum_{c'} p(I_i^c|c',\{F\})p(c',\{F\})},$$

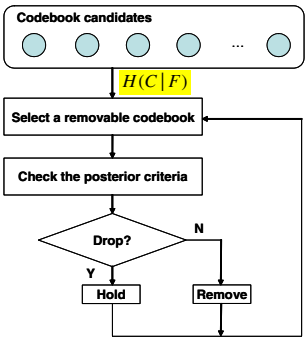
$assume\ uniform\ p(c,\{F\})$

$$F_{opt} = \arg\max_F\{\sum_c\sum_{i\in c}(\log p(I_i^c|c,\{F\}) -$$
$$\log\sum_{c'} p(I_i^c|c',\{F\}))\}$$

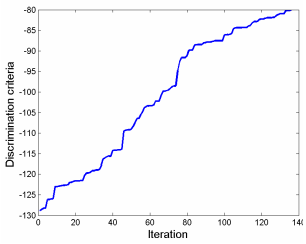
$$where\ p(I_i^c|c,\{F\}) = p(H_i^c|H_M^c) \propto \exp\left(-KL(H_i^c,H_M^c)\right)$$

$$and\ KL(H_i^c,H_M^c) = \sum_{j=1}^{|F|}(H_i^c(j) - H_M^c(j))\log\frac{H_i^c(j)}{H_M^c(j)})$$

The posterior criterion in the 4th line of above equation is used to select the optimal set for a discriminative codebook. Fig. 8 shows the codebook search algorithm and its learning graph. Fig. 9 shows the test results using only a set of the intra-class codebook ($|F| = 1062$) and the discriminatively learned universal codebook ($|F| = 926$) for



(a) optimization procedure



(b) discriminative learning graph

Figure 8: Inter-category entropy-guided universal codebook selection method.

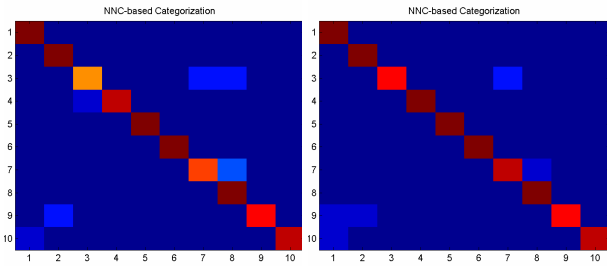


Figure 9: Confusion matrix using non-discriminative codebook and discriminative codebook. (left) Before discriminative learning, (right) after discriminative learning.

10 object categories selected from Caltech-101 DB.

3.5 Distance metrics and classification

After histogram building from the discriminative codebook for all the training instances, we have to learn classifiers with certain distance metrics. We can summarize these as follows.

Distance metrics:$D(H_t, H_m)$

- Euclidean distnace:

$D(H_t, H_m) = \sum_i (H_t(i) - H_m(i))^2$

- KL-divergence:

$D(H_t, H_m) = \sum_i (H_t(i) - H_m(i)) \cdot \log (H_t(i)/H_m(i))$

- Intersection:$D(H_t, H_m) = \sum_i \min (H_t(i), H_m(i))$

- χ^2 distance:$D(H_t, H_m) = \sum_i \frac{(H_t(i)-H_m(i))^2}{H_t(i)+H_m(i)}$

Classification

- NNC is the simplest classifier because it requires no specific learning method. Each training histogram is regarded as a single prototype. So, for an unknown test histogram, a category label is assigned with the nearest prototype in the model database.

- Support vector machine (SVM) [19] can learn classification boundaries from training samples. It has been the most powerful classifier until now. Recently, a kernel-based SVM was introduced that can learn non-linear classification boundaries for complex data. In the extended Gaussian kernel, we can use the distance metrics described above. In the experiment section, we will compare these classification methods using codebooks that are robust to surface markings and discriminative.

Table 1: Summary of classification evaluation using a generative codebook (GC).

Method	# of GC	Eucl.	KL-div	Inters.	χ^2
NNC	1,036	71.3%	**78.0%**	76.6%	77.3%
SVM	1,036	74.0%	74.6%	76.0%	76.0%

Table 2: Summary of classification evaluation in terms of classifiers, distance metrics, and type of codebook.

Method	type	# of CB	Eucl.	KL-div	Inters.	χ^2
NNC	GC	950	75.3	79.3	78.0	77.3
NNC	GC	513	70.6	77.3	76.6	78.0
NNC	DC	348	66.0	**81.3**	73.3	74.0
SVM	GC	950	74.6	74.6	78.0	75.3
SVM	GC	513	74.6	74.0	75.3	75.3
SVM	DC	348	68.0	66.0	72.6	70.6

4 Experimental results

We evaluated our categorization system using a Caltech-101 DB. It consists of 48 man-made objects and 53 animals and plants. In initial experiments, we selected 1 human face and 9 man-made objects, such as airplanes, cameras, cars, cell phones, cups, helicopters, motorbikes, scissors, and umbrellas, which have large intra-class variations due to surface markings. We randomly selected 15 examples for each category and tested 15 unlearned cluttered examples. The first experiment was conducted using codebooks obtained from stage 1 learning, which is fully generative and robust to surface markings. Table 1 shows the classification results (confusion matrix) using NNC and SVM with the different distance metrics. In this test, NNC with KL-divergence showed the best classification results. DAG-SVM for multi-category classification was worse than SVM (One vs. Rest SVM).

In the second experiment, we evaluated the performance of classifiers for a discriminative codebook (DC). Note that the NNC classifier with KL-divergence using DC showed the best performance (Table 2). The SVM classifier does not show merit because we used a small set (15) of training samples for each category.

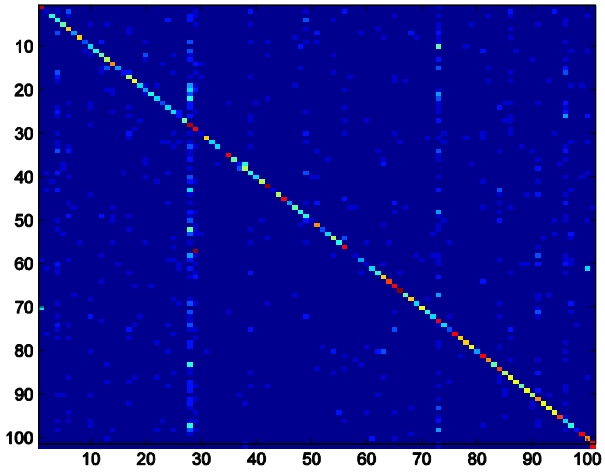


Figure 10: Extended experiment for the whole Caltech-101 DB using NNC-KL div. classifier with DC.

Based on this finding, we extended the experiment to the whole database. We selected the NNC classifier with KL-div. distance. The DC was learned from each category-specific GC. The average classification of our system was 48.58% for a cluttered test set. The current state-of-the-art for the same database using the bag of visual words (single level, L=0, 15 training) shows 41.2% [10]. Most incorrect classifications are for animals and plants.

5 Conclusion

In this paper, we presented an object categorization method focusing on surface markings in the bag of visual words framework. We can minimize the effect of surface markings based on the entropy of the codebooks. High entropy in the intra-class codebook can remove surface marking parts (low entropy) in stage 1 learning. Additionally, a discriminative codebook is also selected from the category-specific codebook guided by the entropy of the inter-class codebook. The high entropy codebook is removed first because it gives ambiguous class labels. Finally, we evaluated those codebooks using NNC and SVM classifiers with different distance metrics. With the optimal set of features, codebooks, and classifiers, we can get upgraded performance in the bag of visual words framework. This work for codebook selection and classification can be applied to other complex categorization methods.

Acknowledgements

This research has been partially supported by the Korean MOST for National Research Laboratory Program (Grant number M1-0302-00-0064), by the MIC for the project, "Development of Cooperative Network-based Humanoids' Technology" of Korea, and by Microsoft Research Asia.

References

[1] M. Bar. Visual objects in context. *Nature Reviews: Neuroscience*, 5:617–629, August 2004.

[2] A. Berg, T. Berg, and J. Malik. Shape matching and object recognition using low distortion correspondences. In *CVPR*, pages 26–33, 2005.

[3] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray. Visual categorization with bags of keypoints. In *ECCV Workshop on Stat. Learn. in Comp. Vis.*, 2004.

[4] G. Dorkó and C. Schmid. Selection of scale-invariant parts for object class recognition. In *ICCV*, pages 634–640, 2003.

[5] S. Dumais, J. Platt, D. Heckerman, and M. Sahami. Inductive learning algorithms and representations for text categorization. In *Int'l conf. on Info. and knowledge management (CIKM'98)*, pages 148–155, 1998.

[6] R. Fergus, P. Perona, and A. Zisserman. Object class recognition by unsupervised scale-invariant learning. In *CVPR*, pages 264–271, 2003.

[7] S. Kim and I.-S. Kweon. Biologically motivated perceptual feature: Generalized robust invariant feature. In *ACCV*, pages 305–314, 2006.

[8] S. Kim and I.-S. Kweon. Simultaneous classification and visualword selection using entropy-based minimum description length. In *ICPR*, pages 650–653, 2006.

[9] J. Kosecka and F. Li. Vision based topological markov localization. In *ICRA*, 2005.

[10] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *CVPR*, pages 2169–2178, 2006.

[11] B. Leibe, A. Leonardis, and B. Schiele. Combined object categorization and segmentation with an implicit shape model. In *Workshop on Stat. Learn. in Comp. Vis.*, 2004.

[12] B. Leibe and B. Schiele. Analyzing appearance and contour based methods for object categorization. In *CVPR*, volume 2, pages 409–415, 2003.

[13] Z. Lin, S. Kim, and I. Kweon. Recognition-based indoor topological navigation using robust invariant features. In *IROS*, 2005.

[14] D. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004.

[15] K. Mikolajczyk, B. Leibe, and B. Schiele. Multiple object class detection with a generative model. In *CVPR*, pages 26–36, 2006.

[16] P. Moreels, M. Maire, and P. Perona. Recognition by probabilistic hypothesis construction. In *ECCV*, pages 55–68, 2004.

[17] C. Schmid, R. Mohr, and C. Bauckhage. Evaluation of interest point detectors. *IJCV*, 37(2):151–172, 2000.

[18] A. Vailaya, M. A. T. Figueiredo, A. K. Jain, and H.-J. Zhang. Image classification for context-based indexing. *IEEE Trans. Image Processing*, 10(1):117–130, 2001.

[19] V. Vapnik. *The nature of statistical learning theory*. Springer-Verlag, 1995.

[20] J. Winn, A. Criminisi, and T. Minka. Object categorization by learned universal visual dictionary. In *ICCV*, pages 1800–1807, 2005.