Indoor Place Categorization based on Adaptive Partitioning of Texture Histograms

Sven Eberhardt

Cognitive Neuroinformatics, University of Bremen

Abstract. How can we localize ourselves within a building solely using visual information, i.e. when no data about prior location or movement are available? Here, we define place categorization as a set of three distinct image classification tasks for view matching, location matching and room matching. We present a novel image descriptor built on texture statistics and dynamic image partitioning that can be used to solve all tested place classification tasks. We benchmark the descriptor by assessing performance of regularization on our own dataset as well as the established INDECS dataset, which varies lighting condition, location and viewing angle on photos taken within an office building. We show improvement on both datasets against a number of baseline algorithms.

Keywords: image processing, localization, visual features, spatial cognition, place categorization

Author Summary

Dipl. Phys. Sven Eberhardt is a PhD student in the Cognitive Neuroinformatics department of University of Bremen. He works on bio-inspired vision models from low-level texture features up to high-level hierarchical and deep learning models. He seeks to apply knowledge found in bio-inspired vision models to spatial cognition tasks in the ActionGroup[A5] of the Spatial Cognition SFB/TR8.

Public Interest Statement

As humans, we possess the remarkable ability to be almost universally able to name what place we are currently at, or at which place a picture might have been taken. Here, we show a novel method of how such one-shot place classification can be performed by a computer vision system.

1 Introduction

Humans possess the remarkable ability to reliably localize themselves in the world under a multitude of conditions: Indoors and outdoors, in unknown terrain, with reduced senses during different weather conditions and even under cognitive load while processing other tasks. But how do we solve this problem? Answering this question will not only give crucial insight into mechanisms of spatial cognition in the human brain, but may also be important for self-localization of mobile robots.

For humans, the dominant sensory information used for place recognition is vision [18]. However, the mechanisms that lead from visual input to knowledge of a place are poorly understood. What are the features we look at to determine where we are? Or, to detach the question from its anthropological context: What are the features that best discriminate between places? Optimal features for this task would have to be specific for a certain place, but invariant to different views within the location [6].

It is important to distinguish these requirements from the requirements for vision-based SLAM (Simultaneous Localization and Mapping. See section 1.1 for a more detailed discussion). The features required there ('*tracking features*') need to be stable across successive camera frames and recognizable from slightly different viewpoints, whereas the features suitable for vision-only place recognition have much stronger invariance requirements. For example, place features need to be matched across completely different views from a place and they needs to be stable over long periods of time with varying illumination conditions. Consequently, successfully matching place features is a much harder task than matching tracking features.

But what are suitable place features for this task? On larger scales, simple, global image statistics have been found to be discriminative between places. In particular, a global texture histogram descriptor has been shown to work for place classification tasks ranging from city-scale to world-scale localization [7]. For indoor localization, [26] also use histograms over relatively simple orientation descriptors on the INDECS [25] database.

However, these features are very generic and confusion may arise if the same texture feature is found in multiple areas of an image, e.g. at the ceiling and on the floor. In this paper, we present an approach that circumvents this problem by partitioning the image according to texture occurrences.

1.1 Related work

The problem of one-shot vision-based place categorization has mostly been investigated on a larger scale level for urban locations e.g. by [28], [30], [14] and [3] for discrimination within a city and [4] for discrimination between two cities. In these studies, authors usually rely on visual features useful for the recognition of house façades to enable the discrimination between locations. However, if and how the above results can be transferred to the problem of room-level indoor place classification is not clear.

In localization tasks in robotics, this question is often overshadowed by the development of vision-based SLAM (see [5] and [1] for a review). In SLAM applications, an autonomous, mobile agent is placed into and moved through an environment with no prior information about its location or surroundings. The agent then simultaneously builds a map of its surrounding and places itself within this map.

Vision-based SLAM typically operates by tracking salient features between successive camera frames and deriving self-movement from shifts and deformations of these tracked features, which is sometimes also called visual odometry [20]. Typically, tracked features are simple image patches, which are determined as unique landmarks in an environment (e.g. [8]) to avoid confusion with other locations.

Indoor place categorization is sometimes assessed in the 'lost robot' (also called 'kidnapped robot')-problem [10], where a mobile robot is placed at a random position and needs to find its place in a previously recorded map. Typically this is done based on map features over the course of several frames e.g. by RATSLAM [24]. However, in this study we try to solve the lost robot problem in a one-shot approach using visual data only. There are other studies which employ a holistic visual descriptor, e.g. by [35] using color histograms, [26] and [36] using image statistics and [31] using SIFT (Scale-invariant feature transform [16]) descriptors.

Textons are histograms over densely sampled cluster assignments on Gabor filter responses, which have been primarily developed for image segmentation [17]. Despite that, they have been shown to be surprisingly strong in scene classification [27] as well as in some outdoor self-localization tasks [7].

One problem with using texture histograms for high-level image classification tasks is that the same texture may be found in multiple regions of an image where they belong to completely different elements of a scene. A common approach is therefore to partition an image (as e.g. in Spatial Pyramids by Lazebnik et al. [15]) and handle individual image portions separately. This is a very crude approach as partitions do not necessarily coincide with contentual segments of an image.

Here, we will address this issue by partitioning images adapted to their contents.

1.2 Task definitions

Finding a suitable measurement for the quality of a localization descriptor is not trivial. In robotics tracking context, measurements of deviation from the ground truth position is commonly used [5], but this does not make sense in one-shot place categorization problem where neither metric information nor connectivity between places is known.

We define indoor place categorization as a classification problem, where labeled images of a location are used as training data into for a classifier. Performance is evaluated as percent correct PC, sometimes also called recall [33], on

test images. With true positives of class i as tp_i and false negatives as fn_i:

$$PC = \frac{\sum_i tp_i}{\sum_i tp_i + fn_i} \tag{1}$$

Chance level of this measure is simply $c = 1/n$, where n is the number of classes.

The definition of place categorization depends very much on the definition of a place, which may be interpreted as categorization of only different views from one specific location [6] or may be defined as broadly as counting all images of a whole city as one place that is being categorized against other cities [4].

To assess indoor place categorization, we define three different conditions on which feature descriptors will be tested.

Room classification means that given a number of training images from each room, a classifier needs to determine the room of a test image. The spatial distance between sample images of each room may be up to a few meters and all viewing angles may occur, as well as different lighting conditions.

Location classification means that each location is one class, and variations within the class include only viewing direction and lighting.

View classification means that sample images are taken from the same location and in the same viewing direction. The same objects will be visible in images of the same class and only variations in illumination and slight shifts might occur. This condition is not strictly place categorization since it regards different views from the same location as different places, but serves as a sensible control condition here.

2 Methods

2.1 Adaptive partitioning of texture histograms

The proposed descriptor is calculated in four steps:

1. All image pixels are assigned into texture clusters (Textons).
2. Textons are partitioned by their occurrence in different image regions.
3. Images are partitioned according to Texton partitioning
4. Separate Texton histograms are calculated for each image partition

Textons is a term coined by Julesz in 1981 for small image patches described by second-order statistics that play a role in human peripheral vision [13]. Textons have later been introduced for image segmentation and texture classification by Malik et al. [17].

In the implementation by Malik which is used here, a vector of Gabor filter responses is assigned to each pixel in an image (for details, see [17]). A set of $n_T = 256$ clusters is precomputed on a training dataset and each pixel is assigned the cluster with the least square distance to its response vector.

We also use a generalization to colored Textons similar to the approach described by [2]. Image RGB values are transformed to opponent color space as described in [11]:

$$\begin{pmatrix} \hat{E} \\ \hat{E}_\lambda \\ \hat{E}_{\lambda\lambda} \end{pmatrix} = \begin{pmatrix} 0.06 & 0.63 & 0.27 \\ 0.30 & 0.04 & -0.35 \\ 0.34 & -0.60 & 0.17 \end{pmatrix} \cdot \begin{pmatrix} R_{x,y} \\ G_{x,y} \\ B_{x,y} \end{pmatrix} \tag{2}$$

Filters are run on all output channels (i.e. luminance ($\hat{E}$), blue-yellow ($\hat{E}_\lambda$) and green-red ($\hat{E}_{\lambda\lambda}$) channel) separately and the filter outputs are concatenated before the clustering step.

Texton partitioning In the next step we partition Textons into their typical occurrence in different vertical parts of an image, such as floor, ceiling or central area.

Let $c_{x,y,i}$ be the Texton assignment from the previous step for image position x, y on indexed image i. We then build histograms per row (y) counting how many times a Texton was assigned to Texton cluster C_j in all unlabeled test images. $1_X(x)$ is the indicator function which is 1 if $x \in X$ and 0 otherwise.

$$r_{j,y} = \sum_{x,i} 1_{C_j}(c_{x,y,i}) \tag{3}$$

From this we derive an average vertical position of occurrence $\bar{y}_j$ for each Texton cluster:

$$\bar{y}_j = \frac{\sum_y y \cdot r_{j,y}}{\sum_y r_{j,y}} \tag{4}$$

We sort the clusters by $\bar{y}_j$ and split the sorted list into n partitions such that the total histograms counts are approximately equal in each partition. Let $\hat{y}_{\hat{j}}$ be the sorted list of $\bar{y}_j$ and $\hat{j}(j)$ the sorting indices, then the normalized cumulative sums of Texton counts along their vertical positioning are:

$$R_{\hat{j}} = \frac{\sum_{\hat{j}'=1}^{\hat{j}} \sum_y r_{\hat{j}',y}}{\sum_{y,j} r_{j,y}} \tag{5}$$

$R_{\hat{j}}$ ranges from 0 to 1. This range is split into n partitions and Texton clusters C_j are assigned to a partition p based on their placement in R:

$$p(C_j) = 1 + \text{floor}\left(n \cdot R_{\hat{j}(j)}\right) \tag{6}$$

Image partitioning Depending on camera angles, field of view and objects in the scene, different portions of each image may be taken up by ceiling, floor and central areas. We therefore partition each image i according to the total amount of Textons $T_i(P)$ from each partition P.

$$T_i(P) = \sum_{x,y} 1_P(p(c_{x,y,i})) \tag{7}$$

The splits between partitions $S_i(P)$ of an image i of height h are put at:

$$S_i(P) = \text{round}\left(h \cdot \frac{\sum_{p=1}^{P} T_i(p)}{\sum_{p=1}^{n} T_i(p)}\right) \tag{8}$$

Texton histograms Texton histograms are calculated separately for each partition. Histograms are normalized by dividing them by the height of their respective sections. The resulting vectors are concatenated into one feature vector of length $n \cdot n_T$.

2.2 Baseline feature vectors

For baseline performance, we also evaluate several established visual feature descriptors derived from the biologically inspired vision models.

HMax (Hierarchical MAX-model) is an object recognition model based on the *Neocognitron* [9] and popularized by Serre et al. [32], which has been used to model neuron receptive field properties found in the ventral stream of the primate visual cortex by Hubel and Wiesel [12]. We use the CNS (Cortical Network Simulator) [19] implementation of HMax with parameter settings and dictionaries as chosen by Serre et al. [32]. Details of the implementation can be found in [32].

HMax features are included in this comparison because they have been shown to be able to discriminate well between object categories [32], which makes them a good representative of landmark-based feature vectors. Based on the assumption that the presence or absence of object classes (like e.g. dishes in a kitchen, chairs in a conference room, etc.) are discriminative for places, HMax features might perform satisfactory for self-localization as well.

Gist is a low-dimensional feature vector designed to capture the gist of a scene developed by Oliva et al. It consists of the first few principal components of spectral components on a very coarse grid (8x8) as well as on the whole image. Gist features are of special interest here because the algorithm has developed in for scene recognition task. For instance Oliva et al. have found in [21] that Gist features successfully distinguish between scene categories like *forest* and *city* and it has been hypothesized that humans use similar features for rapid scene classification tasks [29].

Since there is a strong relation between scenes and locations, Gist features are promising candidates for localization.

2.3 Spatial pyramids

have been introduced by Lazebnik et al. [15]. In this approach, histograms over low-level features over image regions of different size are calculated and concatenated them to one large feature vector. The features used here are densely

sampled SIFT [16] descriptors and we test a three-level pyramid. For the sake of comparability, we omit the custom histogram matching support vector machine (SVM) kernel used by Lazebnik and use the same linear kernel and regression method also employed for other models.

Luminance histogram As an additional control to test whether the task can be completed on very simple features, we also test a simple luminance histogram over the grayscale values of the image.

2.4 Room, place and view categorization

To test how well potential place feature discriminate between rooms, places and views, we evaluate performance as percent correct labels in one-versus-all classification, where all images taken from one room/place/view (see section 1.2) are samples of one class. Classification is performed as linear regression with leave-one-out cross-validation for parameters on the feature vector, which has been reduced to 128 components using principal component analysis (PCA) to ensure all descriptors have the same dimensionality. Each test is repeated 50 times with different random splits between test and training data to determine mean performance and mean error. All classifications are performed using the GURLS (Grand Unied Regularized Least Squares) classification package for MATLAB [34].

All source codes, datasets and trained dictionaries required to reproduce the results of this study can be freely downloaded from `http://www.informatik.uni-bremen.de/cog_neuroinf/indoorstudy`. We also provide data on parameter tuning as applicable to the models in the supplemental materials.

2.5 Test datasets

Fig. 1. Example images from the INDECS database for cloudy (A), night (B) and sunny (C) condition.

INDECS database The INDECS database (Indoor Environment under Changing conditionS) by Pronobis et al. [25] is a collection of 3252 indoor photos taken

from five different under three different lighting conditions (sunny, cloudy, night). Photos are sorted by position they are taken from, viewing angle and lighting condition (see figure 1). We use subsets of the database sorted into different categories to test the different place categorization conditions:

For the *room classification* task, five rooms with 216 images per room are randomly selected. 10 images per class are used for training and the rest for testing. The *place classification* task picks 90 location with 12 images per class, of which 5 are used for training and the rest for testing. The The *view classification* task runs on 50 different views, where location and viewing angle are fixed so only three samples per class are available. One sample is used for training and the rest for testing.

Fig. 2. Example images from kitchen (A), hallway (B) and printer area (C)

3Rooms To ensure that our results are not specific to one dataset, we created an additional image dataset consisting of pictures taken from three connected rooms (see figure 2). For the classification, we took 543 pictures at human eye level from different locations and different view directions from the three rooms.

3 Results

3.1 INDECS classification

For the classification tasks on the INDECS database, there is a strong dependence of performance on task condition (see figure 3). Although samples were taken from the same image dataset, feature vectors rank differently depending on how the classes are defined. Performance on luminance histogram is generally very low and near chance level for all three tasks.

Our proposed model (APTH) as well as the variant running on colored Textons (APCTH) performs well on all three place categorization tasks. For room classification (figure 3 left) we achieve $48.10 \pm 0.53\%$ correct on APCTH, which outperforms the colored Texton approach of $46.34 \pm 0.56\%$ and lies well above all other control models. Both runs were performed with $n = 2$ partitions.

On location classification, performance is much lower, which can be explained by the higher number of classes and also the higher confusion between locations within the same room. Adaptive partitioning again leads to stronger performance

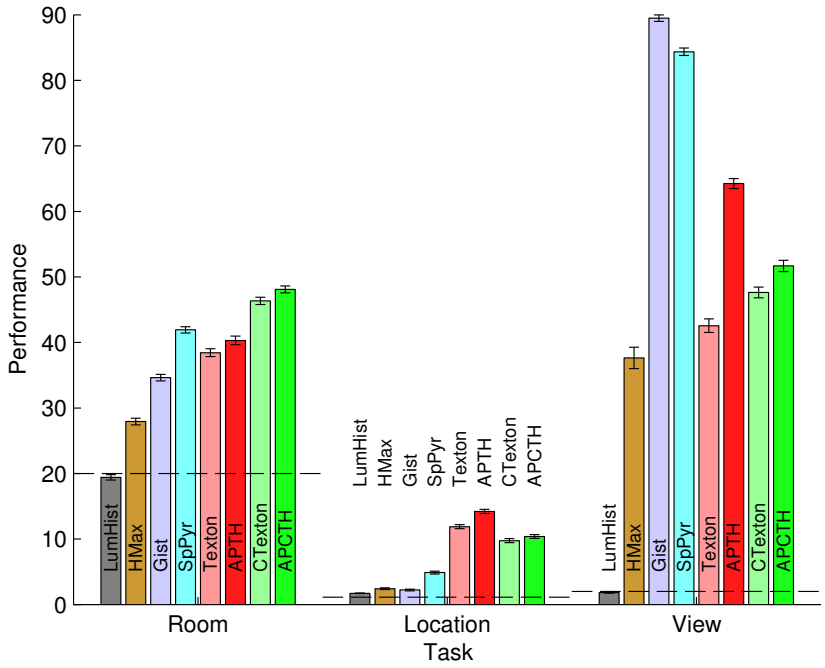

Fig. 3. Classification performance of room, location and view classification task in percent correct by feature vector. Dashed lines mark chance levels. APTH: Adaptive partitioning of texture histograms. APCTH: Adaptive partitioning of color texture histograms. SpPyr: Spatial Pyramids. LumHist: Luminance histogram. CTexton: Colored Textons.

(APTH: $14.20 \pm 0.32\%$, APCTH: $10.39 \pm 0.29\%$) than classification on holistic Texton histograms (Texton: $11.88 \pm 0.31\%$, Colored Texton: $9.76 \pm 0.31\%$). Interestingly, using color leads to higher performance on room classification but is derogatory for performance on location classification.

On view classification, all Texton-based approaches are outperformed by the control models that employ fixed image partitioning (Gist: $89.48 \pm 0.49\%$, Spatial Pyramids: $84.36 \pm 0.57\%$). However, adaptive partitioning of Textons still leads to a huge improvement over raw Texton histograms (APTH: $64.24 \pm 0.76\%$, Texton: $42.56 \pm 1.05\%$).

3.2 3Rooms classification

In the 3Rooms dataset, we test room categorization performance on a dataset that has fewer (3) but more diverse rooms, where a larger number of image samples per room is available. Performance results for the room classification task are shown in figure 4. Again, room classification fails on the luminance histogram, for which performance remains at chance level. For low sample count (figure 4 left),

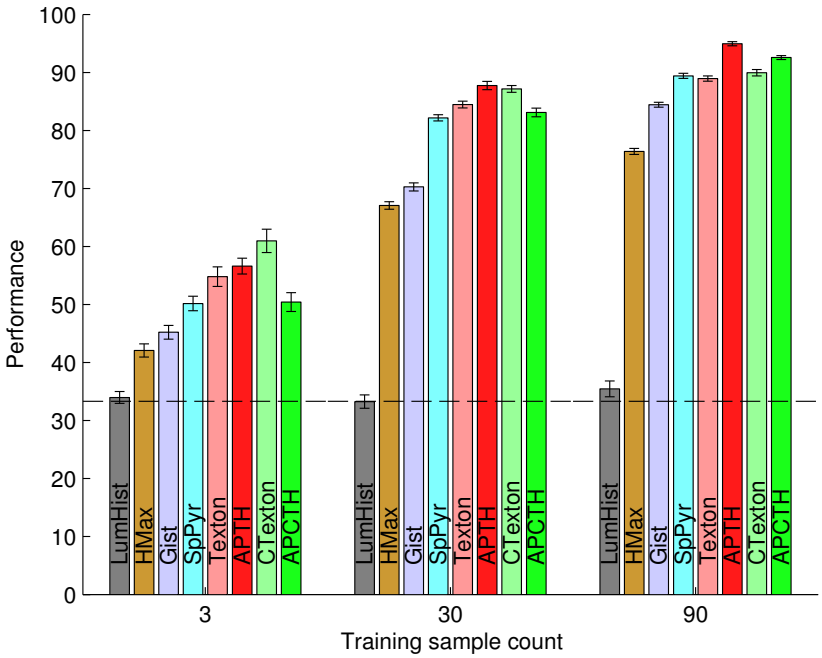

Fig. 4. Classification performance of room classification task in percent correct by feature vector for different number of training samples. Dashed line marks chance level.

APTH yields only a slight improvement over Textons (APTH: $56.62 \pm 1.37\%$, Textons: $54.79 \pm 1.69\%$) and both are outperformed by unpartitioned colored Textons ($60.96 \pm 2.00\%$).

However, as more samples become available (figure 4 right), APTH performance picks up a leads to near perfect labeling ($94.94 \pm 0.36\%$), hinting that after partitioning, more task-relevant information is encoded in the descriptor.

4 Discussion

We have shown that adaptive partitioning of texture histograms can provide a powerful image descriptor to perform several different place recognition tasks and find that our descriptor outperforms all tested baseline models in most of the tested place classification conditions.

The performance gain over vanilla global Texton histograms can be attributed to reduced confusion when the same Texton assignment appears in different areas of the image where they represent different contentual elements.

The rejection of both Gist and Spatial Pyramid features may be surprising, as they have both been introduced in a scene recognition context, which is often closely linked to place recognition. A possible explanation is that Gist has been

tested on scenes downloaded by keyword from photographic image databases [21]. However, when a photographer pictures a certain scene, there is often a default view that is taken which is specific to the scene at hand [22, 23]. For example, a tropical beach is commonly portrait with the vanishing point along the shore, the sun across the sea and a number of palm trees on the opposing side. Such artificial features may be caught by Gist, but they are not inherent to the location when photos are taken in random angles. A similar argument may be made for the Spatial Pyramid descriptor.

The poor performance on HMax descriptor may be attributed to the fact that the hierarchical HMax architecture picks on complex features that are unique to certain objects [32]. For example in a kitchen, a typical HMax feature might be responsive to a pot or an oven. Since these are valuable tracking features, they lead to a strong performance in the view matching task. But individual, tracked objects would only be present in a limited number of views at the location and so they do not generalize well to other views from the same place.

Our results therefore highlight that place recognition is a task which is different from other tasks typically solved in pattern recognition problems such as object detection, scene categorization or feature tracking between successive camera frames as found e.g. in SLAM applications.

But why do Texton histograms in general and adaptively partitioned Texton histograms in particular generalize? The mechanisms are still poorly understood and should be subject of future studies. However, one possible explanation is that Textons are designed to discriminate between surface textures and therefore attributes such as material types of the surrounding environment. Surface types are common to a more generic class of objects found at certain places than individual objects themselves. For example, the existence of metallic surfaces may be indicative of a kitchen because many metallic objects may be found there and they are distributed across the whole room.

Therefore, we hypothesize that a suitable visual place descriptor should be geared at recognizing surface structure instead of objects.

Acknowledgments. This work was supported by DFG, SFB/TR8 Spatial Cognition, project A5-[ActionSpace].

References

1. Bailey, T., Durrant-Whyte, H.: Simultaneous localization and mapping (SLAM): Part II. IEEE Robotics & Automation Magazine 13(3), 108–117 (2006)
2. Burghouts, G.J., Geusebroek, J.m.: Color Textons for Texture Recognition. In: British Machine Vision Conference. pp. 1099–1108 (2006)
3. Chen, D., Baatz, G., Koser, K.: City-scale landmark identification on mobile devices. Computer Vision and Pattern Recognition. IEEE Conference on (2011)
4. Doersch, C., Singh, S., Gupta, A.: What makes Paris look like Paris? ACM Transactions on Graphics (2012)
5. Durrant-Whyte, H., Bailey, T.: Simultaneous localization and mapping: part I. IEEE Robotics & Automation Magazine 13(2), 99–110 (2006)

6. Eberhardt, S., Kluth, T., Zetzsche, C., Schill, K.: From pattern recognition to place identification. In: Spatial cognition, international workshop on place-related knowledge acquisition research. pp. 39–44 (2012)
7. Eberhardt, S., Zetzsche, C.: Low-level global features for vision-based localization. In: Proceedings of the KI 2013 Workshop on Visual and Spatial Cognition. pp. 5–13 (2013)
8. Fitzgibbons, T., Nebot, E.: Bearing only SLAM using colour-based feature tracking. Proceedings of the Australasian Conference on Robotics and Automation, Auckland (2002)
9. Fukushima, K.: Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biological Cybernetics 36, 193–202 (1980)
10. Gerecke, U., Sharkey, N.: Quick and dirty localization for a lost robot. In: Computational Intelligence in Robotics and Automation. IEEE International Symposium on. pp. 262–267 (1999)
11. Geusebroek, J.M., van den Boomgaard, R., Smeulders, A., Geerts, H.: Color invariance. IEEE Transactions on Pattern Analysis and Machine Intelligence 23(12), 1338–1350 (2001)
12. Hubel, D., Wiesel, T.: Receptive fields and functional architecture of monkey striate cortex. The Journal of physiology pp. 215–243 (1968)
13. Julesz, B.: Textons, the elements of texture perception, and their interactions. Nature 290, 91–97 (1981)
14. Knopp, J., Sivic, J., Pajdla, T.: Avoiding confusing features in place recognition. In: Computer Vision - ECCV. pp. 748–761 (2010)
15. Lazebnik, S., Schmid, C., Ponce, J.: Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In: Computer vision and pattern recognition, IEEE computer society conference on. vol. 2, pp. 2169–2178 (2006)
16. Lowe, D.: Object recognition from local scale-invariant features. In: Computer vision. IEEE seventh international conference on. vol. 2, pp. 1150–1157 (1999)
17. Malik, J., Belongie, S., Leung, T., Shi, J.: Contour and texture analysis for image segmentation. International journal of computer vision 43(1), 7–27 (2001)
18. Markus, E.J., Barnes, C.a., McNaughton, B.L., Gladden, V.L., Skaggs, W.E.: Spatial information content and reliability of hippocampal CA1 neurons: effects of visual input. Hippocampus 4(4), 410–421 (1994)
19. Mutch, J., Knoblich, U., Poggio, T.: CNS: a GPU-based framework for simulating cortically-organized networks. Tech. rep., Massachusetts Institute of Technology, Cambridge, MA (2010)
20. Nistér, D., Naroditsky, O., Bergen, J.: Visual odometry. Computer Vision and Pattern Recognition. Proceedings of the IEEE Computer Society Conference on 1, 652–659 (2004)
21. Oliva, A., Hospital, W., Ave, L.: Modeling the shape of the scene: A holistic representation of the spatial envelope. International journal of computer vision 42(3), 145–175 (2001)
22. Pinto, N., Cox, D.D., DiCarlo, J.J.: Why is real-world visual object recognition hard? PLoS computational biology 4(1), e27 (2008)
23. Ponce, J., Berg, T.L., Everingham, M., Forsyth, D.A., Hebert, M., Lazebnik, S., Marszalek, M., Schmid, C., Russell, B.C., Torralba, A., Williams, C.K.I., Zhang, J., Zisserman, A.: Dataset issues in object recognition. Springer Berlin Heidelberg (2006)
24. Prasser, D., Milford, M., Wyeth, G.: Outdoor simultaneous localisation and mapping using RatSLAM. Field and Service Robotics pp. 143–154 (2006)

25. Pronobis, A., Caputo, B.: A discriminative approach to robust visual place recognition. Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on pp. 3829–3836 (2006)
26. Pronobis, A., Caputo, B.: A realistic benchmark for visual indoor place recognition. Robotics and Autonomous Systems 58(1), 81–96 (2010)
27. Renninger, L.W., Malik, J.: When is scene identification just texture recognition? Vision research 44(19), 2301–2311 (2004)
28. Robertson, D., Cipolla, R.: An Image-Based System for Urban Navigation. British Machine Vision Conference pp. 1–10 (2004)
29. Rousselet, G., Joubert, O., Fabre-Thorpe, M.: How long to get to the gist of real-world natural scenes? Visual Cognition 12(6), 852–877 (2005)
30. Schindler, G., Brown, M., Szeliski, R.: City-scale location recognition. Computer Vision and Pattern Recognition. IEEE Conference on (2007)
31. Se, S., Lowe, D., Little, J.: Global localization using distinctive visual features. In: Intelligent Robots and Systems. IEEE/RSJ International Conference on. pp. 226–231 (2002)
32. Serre, T., Oliva, A., Poggio, T.: A feedforward architecture accounts for rapid categorization. Proceedings of the national academy of sciences 104(15), 6424–6429 (2007)
33. Sokolova, M., Lapalme, G.: A systematic analysis of performance measures for classification tasks. Information Processing & Management 45(4), 427–437 (2009)
34. Tacchetti, A., Mallapragada, P.K., Santoro, M., Rosasco, L.: GURLS: a toolbox for large scale multiclass learning. In: Big learning workshop at NIPS (2011), `http://cbcl.mit.edu/gurls/`
35. Werner, F., Sitte, J., Maire, F.: Visual topological mapping and localisation using colour histograms. In: Proceedings of the International Conference on Control, Automation, Robotics and Vision (2008)
36. Zhou, C., Wei, Y., Tan, T.: Mobile robot self-localization based on global visual appearance features. In: Robotics and Automation. Proceedings. IEEE International Conference on. pp. 1271–1276 (2003)