From Pattern Recognition to Place Identification

Sven Eberhardt, Tobias Kluth, Christoph Zetzsche, and Kerstin Schill

Cognitive Neuroinformatics, University of Bremen, Enrique-Schmidt-Straße 5, 28359 Bremen, Germany,
sven2@uni-bremen.de,
homepage: http://www.informatik.uni-bremen.de/cog_neuroinf/en

Abstract. What are the ingredients required for vision-based place recognition? Pattern recognition models for localization must fulfill invariance requirements different from those of object recognition. We propose a method to evaluate the suitability of existing image processing techniques by testing their outputs against these invariances. The method is applied to several holistic and one local model. We generalize our findings and identify model properties of locality, spatial configuration and generalization as key factors for applicability to localization tasks.

Keywords: visual, model, pattern recognition, localization

1 Introduction

Although the concept of place is essential to the way humans represent and interact with spatial environments, many of its determinants are not yet completely understood. One important question is what kind of information and what computations can be used to determine a specific place. Among the different types of input suitable for this purpose pictorial information has a particularly high potential. In biological terms, the investigation of place cells, for example, indicates the importance of visual cues for the robust localization of rodents.[1]

However, the exact processing mechanisms that can enable a successful vision-based localization are still unclear. In particular, it has to be understood how the classical determinants of pattern recognition systems, invariance and generalization properties, relate to the problem of localization. Invariance properties seem to play a crucial role, since for example the activation of a place cell is primarily determined by the animal's location, whereas it is independent of the orientation and other conditions like illumination. These are typical invariance properties. It may thus be assumed that the classic invariance principles attributed to human vision, and the corresponding computer vision approaches, can also be applied to the problem of localization (or place recognition). In this paper, we will argue that this is not necessarily the case, and that successful localization requires specific properties that can be in direct in opposition to those underlying other basic visual capabilities, like for example object recognition. For this, we will first introduce a basic framework that enables the description and differentiation of image processing techniques with respect to their applicability for localization

as compared to, e.g., object recognition. We will then discuss how some established image processing techniques can be described in terms of the suggested framework. This will then motivate an investigation of the suitability of some of these techniques for the specific problem of localization, or place recognition. In particular, we will investigate whether one of the most successful models of visual object recognition, the HMAX model[2], can also be used for the task of vision based localization.

1.1 Invariance in Place Recognition

One of the difficulties in place recognition from visual input is that even minor changes in observer's orientation or location, as well as unrelated changes such as variations in illumination, can cause vast changes of retinal input. Successful models for place identification should provide output that is invariant to such small changes in the observer's view. Although this is a requirement which is shared with object recognition models, there are some fundamental differences in which kind of invariance is desired.

While changes in scale, position and occlusion of elements in a scene are often irrelevant in the context of object recognition, they correspond to movement of the observer and should elicit changes in the output for place recognition models. On the other hand, spatial shifting of a scene as a whole corresponds to rotation of the observer. Similarly, rotations within the viewing plane correspond to tilting of the viewer's head. A place detector that mimics the behavior of place cells should be invariant to such rotations.

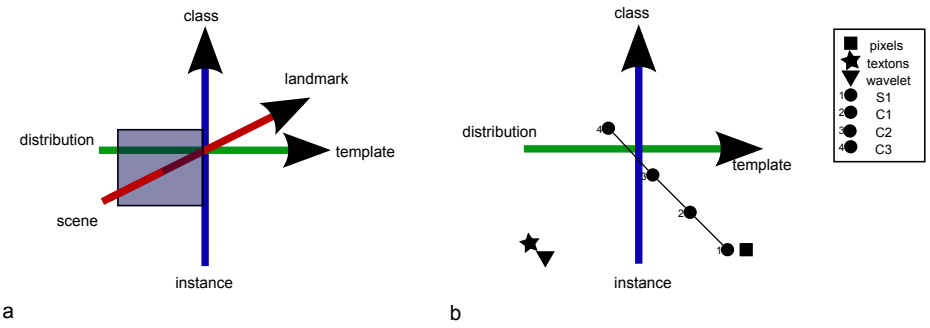

Fig. 1. a: Conceptual space for classifying pattern recognition models. b: Position of analyzed models in two dimension of our conceptual space.

Given these fundamental differences, can models for object recognition be used for place identification at all? The large amount of existing pattern recognition algorithms makes testing this hypothesis a tedious task. We therefore suggest to categorize algorithms into a conceptual space with three dimensions [3] and seek to find a systemic correspondence between the placement of models within these dimensions and their applicability to place identification.

The *first dimension* is locality. A local approach processes image data from selected image regions, whereas a global approach always takes the whole image into account. Naturally, local approaches need a detection mechanism to determine regions of interests (ROI). Such mechanisms may rely on low-level image data such as curvature[4], local brightness extrema[5] or generalized features[6]. Ideally, the detection mechanism picks out informative image regions containing objects or landmarks useful to solve the given task.

The *second dimension* measures the invariance to changes in spatial configuration. Algorithms that are sensitive to spatial layout match templates of stored objects against the input image, but fail to generalize if object components are rearranged or scrambled. On the other hand, the largest invariance to spatial layout is provided by models relying on image statistics [7] or bags-of-features like [8]. The class of HMAX models by [2] follow an intermediate approach where invariance to feature locations is increased step-by-step in a multi-layer hierarchy.

The *third dimension* describes how well a model generalizes among several instances of a class. Most local descriptor-based algorithms such as [5, 6] only store patterns specific to the particular instance and view of an object, so multiple patterns are required to describe a class. Usually, category-level generalization can be achieved by clustering specific descriptors into broader categories [9].

These dimensions describe key attributes required for a model to be suitable for place identification. The first dimension, locality, is certainly useful to determine place. If each detected feature is attributed to a position, the relation of these positions provides valuable information in determining the position of an observer[10]. For the second dimension, spatial configuration, the requirements are not so clear. On the one hand, changes in spatial configuration result from changes in position of an observer and invariance to such changes is not desired. On the other hand, invariance to small changes in configuration increase robustness of the detection of features, and could improve detection when scenes are presented under slightly different conditions. The third dimension, generalization properties, are probably required to some extent to generalize different views from the same place onto the same class. Too much generalization is not desirable, because it might project locations that look similar onto the same place.

In the following study, we investigated the invariance properties of models that vary within the second and third dimension. In particular, we varied the two parameters of location and orientation. We judged algorithms based on how well they stayed invariant to changes in orientation compared to their variation induced by changes in location. We tested two holistic models, wavelet-like histograms[7] and texture descriptors called 'textons'[11]. In comparison, we chose the HMAX model as a hierarchical model of which we analyzed each model step separately. Finally, performance on raw pixel values has been checked as a baseline.

2 Methods

We developed a test setup to evaluate the applicability of pattern recognition methods for place identification. We recorded input images x_α^L at $n_L = 10$ different locations L, and $n_\alpha = 181$ different observer rotation angles α spanning 180 degrees of rotation. If a model S is applied to two input images, the dissimilarity of output vectors can be written as their euclidean distance d^S.

$$d^S(x_{\alpha_1}^{L_1}, x_{\alpha_2}^{L_2}) := \left\| S\left(x_{\alpha_1}^{L_1}\right) - S\left(x_{\alpha_2}^{L_2}\right) \right\|_2 \tag{1}$$

We measure the invariance to rotation $\tilde{I}_{rot}^S(\alpha)$ of a model by averaging the dissimilarity to a midpoint rotation over all locations.

$$\tilde{I}_{rot}^S(\alpha) := \frac{1}{n_L} \sum_L d^S(x_\alpha^L, x_0^L) \tag{2}$$

A low value of $\tilde{I}_{rot}^S(\alpha)$ means the output of the model is highly invariant to the given rotation α. In order to measure usability for place identification, we need to put this value in relation to variations of model outputs achieved by changing the place. We define a relative orientation invariance measure $I_{rot}^S(\alpha)$ as:

$$I_{rot}^S(\alpha) := \frac{1}{n_L} \sum_L \frac{d^S(x_\alpha^L, x_0^L)}{\frac{1}{n_\alpha} \sum_{\alpha'} \min_{L' \neq L} d^S(x_{\alpha'}^L, x_{\alpha'}^{L'})} \tag{3}$$

Values larger than 1 for $\tilde{I}_{rot}^S$ for a given angle mean that the model produces more dissimilar outputs under rotation by that angle than it would by switching the place. We therefore define the maximum angle of invariance α_I^S as the largest value under which this condition is met:

$$\alpha_I^S := \max \left\{ |\alpha| \;\middle|\; I_{rot}^S(\alpha) < 1 \right\} \tag{4}$$

Large values of α_I^S stand for good invariance to rotation compared to changes in place, which attributes the model as suitable for place recognition.

We applied this method to the raw input pixels, as well as outputs from the texton algorithm, wavelet descriptors and the HMAX model at various stages. For the HMAX model, we were particularly interested in how the rotational invariance properties vary with increasing layers. We extracted values at the gabor filter layer (S1), as well as the fist and second local invariance layers (C1, C2) and the final, global invariance layer (C3) At each layer, a maximum of 500 features was extracted. For non-global layers, a random sub sampling over features and locations was done. The same features at the same locations were subtracted for all images.

3 Results

We find that, in accordance with our predictions, pattern recognition models display vastly different performances when investigated for their applicability in

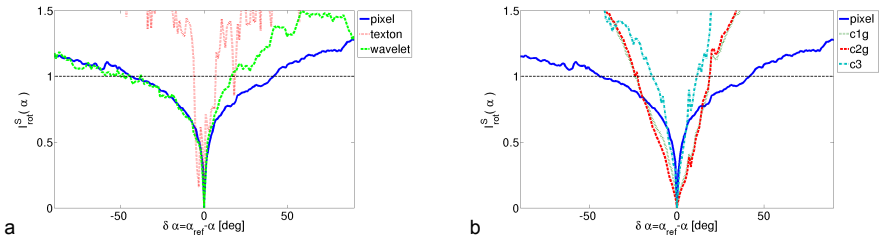

Fig. 2. Relative orientation invariance measures $I^S_{rot}(\alpha)$ for (a) raw pixels, textons and wavelets and (b) different layers of the HMAX model.

place identification. Relative orientation invariance measures $I^S_{rot}(\alpha)$ for holistic models (textons and wavelets) as well as raw pixels are shown in fig. 2a. The maximum angle of invariance for texton outputs ($\alpha_I^{\text{Texton}} = 7°$) is actually lower than for raw pixels ($\alpha_I^{\text{Pix}} = 44°$), which shows that these models are more invariant to changes in location than to changes in rotation compared to raw pixels. Invariance to wavelet transformation is only slightly lower than pixels ($\alpha_I^{\text{Wavelets}} = 38°$).

For HMAX, performance for each layer is shown in figure fig. 2b. Again, α_I sinks below performance on raw pixel down to ($\alpha_I^{\text{Wavelets}} = 24°$) for the successive layers C1 and C2 and further down to ($\alpha_I^{\text{Wavelets}} = 14°$) for the final layer C3. This decay of performance in higher stages of the model show that invariance to place increases faster than invariance to orientation.

4 Discussion

We have investigated the question of how a place can be characterized in terms of visual properties. In particular, we have investigated which invariances are required to uniquely determine a place and how these are related to the invariance properties commonly attributed to visual processing. We have evaluated different models asking how they are able to generalize across all possible views of a place while still being selective enough to guarantee a unique localization.

We have shown that the invariance requirements for place recognition are not necessarily met by models popular for object recognition, such as texton outputs or HMAX. Further, we found that higher layers in the hierarchy of the model, which correspond to more complex features and higher levels of invariance to spatial configuration, lead to a reduced level of invariance to rotation. This yields the hypothesis that invariance to spatial layout, i.e. the second dimension of our conceptual space in fig. 1a, is a detrimental ingredient for invariant place recognition in general. However, since we have explored only a small part of the space of approaches, a more comprehensive study needs to be done.

How much generalization is needed to perform localization? Being able to generalize across different views of the same location is certainly helpful. How-

ever, if generalization leads to higher invariance across different locations, as happens in the higher stages of the HMAX model in our case, reliable place identification performance decreases.

Interestingly, [12] proposes a hierarchical model architecture for place cells very similar to that of the HMAX model. In his model, cells are repeated across locations and pooled over increasingly receptive fields in higher stages. The main difference to HMAX lies in that features are trained explicitly to be invariant to rotations using slow feature analysis. This shows that the invariance properties wired into a model greatly affect its suitability for localization, as long as the learning stage is tuned generalize across views, but not across places.

These results suggest that a universal vision system for both object recognition and localization methods is unfeasible. While some of the processing mechanisms may be shared between architectures for the two tasks, specific mechanisms are required to uniquely determine a place.

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

References

1. O'Keefe, J., Dostrovsky, J.: The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. Brain research **34**(1) (1971) 171
2. Serre, T., Oliva, A., Poggio, T.: A feedforward architecture accounts for rapid categorization. PNAS **104**(15) (2007) 6424–6429
3. Wolter, J., Reineking, T., Zetzsche, C., Schill, K.: From visual perception to place. Cognitive Processing **10** (2009) 351–354
4. Zetzsche, C., Barth, E.: Fundamental limits of linear filters in the visual processing of two-dimensional signals. Vision Research **30** (1990) 1111–1117
5. Lowe, D.: Distinctive image features from scale-invariant keypoints. IJCV **60**(2) (2004) 91–110
6. Kim, S., Kweon, I.: Biologically motivated perceptual feature: Generalized robust invariant feature. In Narayanan, P., Nayar, S., Shum, H.Y., eds.: Computer Vision ACCV 2006. Volume 3852 of Lecture Notes in Computer Science. Springer Berlin / Heidelberg (2006) 305–314
7. Oliva, A., Torralba, A.: Modeling the shape of the scene: A holistic representation of the spatial envelope. IJCV **42**(3) (2001) 145–175
8. Grauman, K., Darrell, T.: Efficient image matching with distributions of local invariant features. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005. CVPR 2005. Volume 2. (2005)
9. Griffin, G., Holub, A., Perona, P.: Caltech-256 object category dataset. Technical Report 7694, California Institute of Technology (2007)
10. Se, S., Lowe, D., Little, J.: Vision-based global localization and mapping for mobile robots. IEEE Transactions on Robotics **21**(3) (2005) 364–375
11. Renninger, L., Malik, J.: When is scene identification just texture recognition? Vision Research **44**(19) (2004) 2301–2311
12. Franzius, M., Sprekeler, H., Wiskott, L.: Slowness and sparseness lead to place, head-direction, and spatial-view cells. PLoS Comput Biol **3**(8) (2007) e166