From Perceptual Relations to Scene Gist Recognition

Ilan Kadar and Ohad Ben-Shahar
Dept. of Computer Science, Ben-Gurion University
Beer-Sheva, Israel
{ilankad,ben-shahar}@cs.bgu.ac.il

Abstract

The ability to recognize visual scenes quickly and accurately is highly constructive for both biological and machine vision. In this work we study the process of scene gist recognition from a novel point of view and investigate whether prior knowledge of the perceptual relations between the different scene categories may help facilitate better computational models for scene gist recognition. We first introduce a psychophysical paradigm that probes human scene gist recognition and extracts perceptual relations between scene categories. Then, we show that these perceptual relations do not always conform the semantic structure between categories. Next, we incorporate the obtained perceptual relations into a computational classification scheme, which takes inter-class relationships into account to obtain better scene recognition regardless of the particular descriptors with which scenes are represented. We present such improved recognition performance using several popular descriptors, we discuss why the contribution of inter-class perceptual relations is particularly pronounced for under-sampled training sets, and we argue that this mechanism may explain the ability of the human visual system to perform well under similar conditions. Finally, we introduce an online experimental system for obtaining perceptual relations for large collections of scene categories.

1. Perceptual Relations

A key issue in the context of scene gist recognition is perceptual relations (as opposed to semantic relations; see below), a possibility that has been rarely considered either in the perceptual literature or the computational literature [7, 2, 8]. However, even intuitively, when our visual system observes a bedroom scene for a fraction of a second and "deliberates" how to categorize it, what possibly comes to mind in addition to "bedroom" are perhaps classes like "living room" or "kitchen". It appears as if our visual system does not even consider possibilities such as "coast" or "highway", or more generally, scenes which are perceptually "distant" from the observable reference class. Put differently, prior knowledge about the perceptual relations between the different categories of scenes may help facilitate more accurate and more efficient scene gist recognition. Knowledge of such relationships could also partly explain the fact that humans are often able to learn and process hundreds of scene categories from very few training examples while computational models usually need at least tens of training examples per category before achieving reasonable recognition performance. Exploring relations between categories is not new and was recently promoted by exploiting *WordNet* as a *semantic* relationships database for object recognition [3]. Indeed, *semantic* relationships can be extracted quite conveniently from *WordNet*. Still, we found several examples to suggest that *semantic* relationships between categories do not necessarily agree with their *perceptual* relationships. For example, our experimental setup reveals that the "highway" category is perceptually closer to "coast" than to "kitchen", although semantically the opposite holds. Once relations between visual categories are considered based on perceptual criteria, two questions immediately arise. First, how can perceptual difference or distance between scene categories be determined or inferred directly from human vision? Put differently, can the perceptual distance between categories be measured psychophysically in a robust and unbiased way? Second, once determined, how could these perceptual relations be incorporated into a computational classification scheme.

We introduce a psychophysical paradigm where we briefly present two natural scene stimuli *simultaneously* and ask human observers whether they belong to the same scene category or not (i.e., same/different forced choice task). Collected from 79 human subjects, we analyze subjects' average performance over to provide an unbiased objective measure regarding the perceptual "distance" between the different scene categories. In particular, we calculate subjects' probability to respond *Different* for each pair of categories. Since this probability is expected to increase when such judgment is easier, and since the latter case is expected when scenes become more "perceptually different", this probability is termed as the "perceptual distance" (PD)

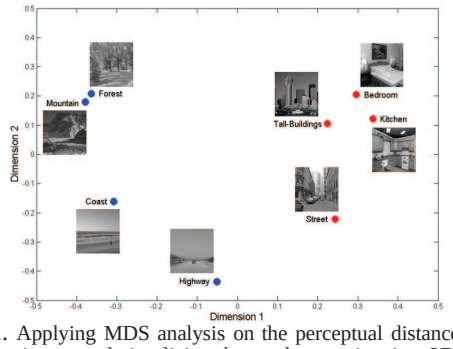

Figure 1. Applying MDS analysis on the perceptual distances obtained in our experiment, and visualizing the results as points in a 2D perceptual space.

between pair of visual scene categories. Figure 1 illustrates the result of the *Multidimensional Scaling (MDS)* analysis on the perceptual distances obtained in our experiment.

2. Scene Recognition with Perceptual Relations

With perceptual relations established via experimental analysis as above, we turn to discuss how they may be exploited for scene categorization, especially when only few labeled examples are available for each class. In contrast to existing computational algorithms, our everyday experience indicates that humans can learn new scene categories from only few examples, so it seems unlikely that a large set of training examples per-category is a necessity. We suggest that the solution resides in leveraging perceptual relations *between* categories in order to characterize each scene category in a more informative way. To introduce the idea, consider that after learning few examples from each category, we are given a query scene and asked to decide whether or not it belongs to one of the categories, say "coast". Due to the visual complexity of real-scenes and the high visual variability within each scene category, it seems practically impossible that the few training coast examples would contain all the perceptual properties of a characteristic coast scene. However, additional perceptual properties of such scenes may possibly be obtained from training examples of other, perceptually related categories, e.g., the "highway" category. Still, while doing so, highway examples should be considered as making smaller contribution compared to coast examples. More generally, we suggest that the perceptual properties of the given category can be learned or inferred not only from its designated exemplars, but also from all training examples that belong to other perceptually related categories, weighted according to their perceptual distance to the given category. In effect, such strategy increases the pool of useful training examples manyfold, thus facilitating categorization performance similar to well-sampled classes. We develop this idea more formally and incorporate perceptual distance into a practical classifier. Although this can be done with almost any classifier, here we chose the *Naive-Bayes Nearest-Neighbor* (NBNN) algorithm [1] due to its excellent trade off between simplicity (or complexity) and performance. In particular, we extend the NBNN classifier to a new classifier that exploits inter-class relations and (abbreviated as NBNN-ICR). In order to explore the possible contribution of inter-class perceptual relations, we combine our measured perceptual relations and the NBNN-ICR classifier into a computational scene recognition framework. The resultant classifier, termed here as NBNN-PR, is then compared to NBNN (see Fig. 2) to show how the use of the measured perceptual relations facilitates significant improvements in scene recognition performance, especially when the number of training scenes in each category is small. We also show that this improvement does not result from the mere inclusion of any class relations, but from the very particular relations that were inferred experimentally and reflect human perception. Toward this end we compared performance to instances of NBNN-ICR where the inter-class relations are selected randomly (cf. NBNN-Rand), or are estimated computationally from a trained classifier confronted with the same experimental procedure as described for humans (cf. NBNN-CR).

Finally, seeking to apply this theory on large collections of scene categories, we introduce an online experimental system [4] that allows users from all over the world to participate in our experiment for mass acquisition of perceptual relations. With this online experimental system, we believe that the perceptual relations between many scene categories from the SUN dataset [8] could be established in reasonable period (for more details, see [6, 5])

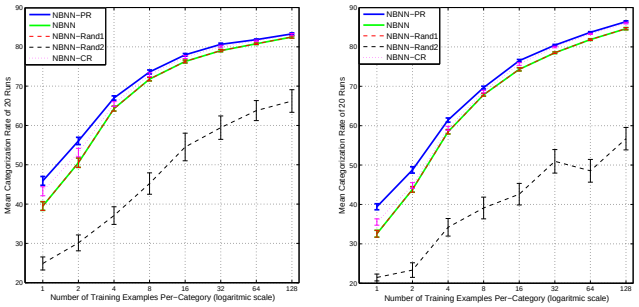

Figure 2. Performance of all discussed classifiers (NBNN, NBNN-PR, and all other variants of NBNN-ICR), based on the GIST and the LBP descriptors, as the number of training examples is increased.

References

[1] O. Boiman, E. Shechtman, and M. Irani. In defense of nearest-neighbor based image classification. In *Proc. CVPR*, 2008. 2

[2] L. Fei-Fei and P. Perona. A bayesian hierarchy model for learning natural scene categories. In *Proc. CVPR*, 2005. 1

[3] R. Fergus, H. Bernal, Y. Weiss, and A. Torralba. Semantic label sharing for learning with many categories. In *ECCV*, 2010. 1

[4] I. Kadar and O. Ben-Shahar. An online experimental system: `http://www.cs.bgu.ac.il/~vision/pmc`, 2012. 2

[5] I. Kadar and O. Ben-Shahar. A perceptual paradigm and psychophysical evidence for hierarchy in scene gist processing. *Journal of Vision*, 2012. 2

[6] I. Kadar and O. Ben-Shahar. Small sample scene categorization from perceptual relations. In *Proc. CVPR*, 2012. 2

[7] A. Oliva and A. Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *Int. J. Comput. Vision*, 2001. 1

[8] J. Xiao, J. Hays, K. Ehinger, A. Oliva, and A. Torralba. Sun database: Large scale scene recognition from abbey to zoo. In *Proc. CVPR*, 2010. 1, 2