Memory Organization for Invariant Object Recognition and Categorization

Guillermo S. Donatti

Institute for Neural Computation, Ruhr University Bochum, Germany

Advisor: Rolf P. Würtz, Institute for Neural Computation, Ruhr University Bochum, Germany
Date and location of PhD thesis defense: 21 June 2016, International Graduate School of Neuroscience,
Ruhr University Bochum, Germany

Abstract

Visual object understanding [12] relies on the comparison of currently perceived and previously acquired object models, either to discriminate among physically similar ones in the case of *recognition*, or to generalize common properties across physically different ones during *categorization*. Current research trends [16] go towards the development of artificial systems that use appearance-based object models combining feature- and correspondence-based approaches to overcome object invariance difficulties. They are structured dynamically using relatively invariant patches of information, which can be shared across objects from the same category. These patches are represented by regular-shaped graphs, whose nodes are labeled with elementary features that capture texture information from the object view. The accumulated world knowledge of these artificial systems is built on a *visual dictionary* of such patches and must be organized in a way that allows efficient object retrieval and identity or category decisions, respectively.

The present work introduces a memory framework for these artificial systems utilizing three different approaches to self-organization of visual object knowledge: the *Neural Map*, the *Neural Map Hierarchy*, and the *Semantic Correlation Graph*. The properties of the proposed computational theory are grounded in the unsupervised structural organization of these object models' components according to their visual resemblance and co-occurrence, as well as in the use of this structure for matching novel components, as illustrated in Figure 1. Both have been identified as important for achieving efficient object recognition and categorization and for providing insight into the knowledge-driven aspect of perception [2].

The *Neural Map* [3] combines a Growing Neural Gas Network [6] and a classifier inspired by the coding and decoding of information in the brain [15]. A modified winner-take-all voting scheme integrates the feature matching responses of this memory model. The performance for object recognition and categorization of the resulting Neural Map Classifier is validated using image features that derive texture information from object views with different granularity. The present work extends them by employing different representations to encode gray- and color-valued texture information from object views and evaluates alterations on self-organization caused by neural network bootstrapping [1]. The overall results indicate that medium-sized image features with
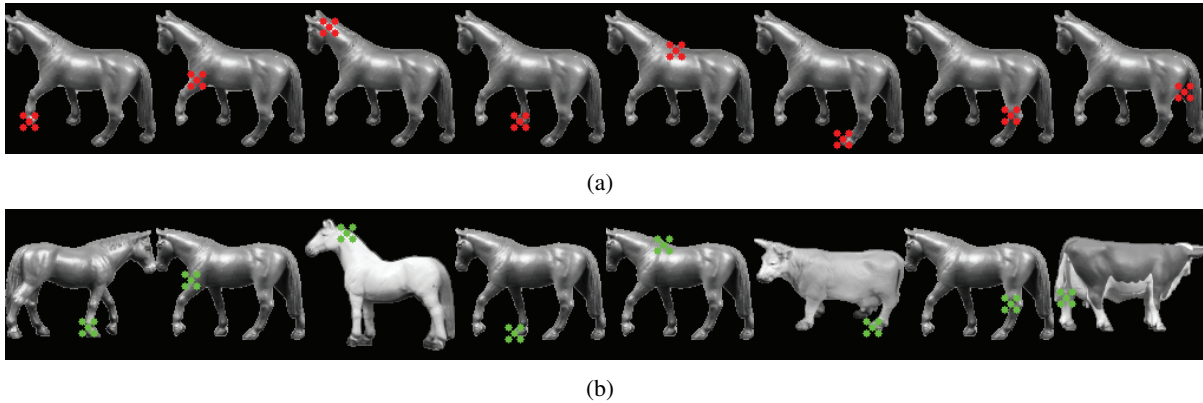
(a)

(b)

Figure 1: Compositional Approach to Object Recognition and Categorization. Artificial systems based on the self-organization approaches introduced in the present work define the signature of an unknown object view by integrating the best matching responses between extracted image features (a) and model image features (b) memorized in a visual dictionary. The world knowledge of these artificial systems is automatically learned based on the texture similarity of image features and their co-occurrence in already known object views.

the largest amount of feature descriptors maximize the informativeness and distinctiveness of texture information derived from object views. Furthermore, bootstrapping neural network topologies enhances the artificial system's performance at the cost of slightly longer learning times. This memory model is optimized by Evolutionary Algorithms exploring six different optimization approaches given by the combination of three fitness functions and two starting conditions [4]. The parameter values determined with the Sample Distance fitness function generate Neural Maps that reach the highest correct object recognition and categorization percentages. The present work also limits bootstrapping and neural growth during learning to demonstrate the Neural Map capabilities for data compression.

The *Neural Map Hierarchy* is motivated by the efficiency to store and retrieve semantic information about taxonomic hierarchies and by exploiting the tendency of features to occur in clusters across instances in the world. This memory model comprises different abstraction layers with self-organized representations of categories. The representations approximate the topological properties of clusters containing semantically equivalent samples of the feature distribution. The information stored in the more general representations determines the feature weights of more specific levels of categorization. The hierarchy of this memory model is defined according to the taxonomy of the ETH-80 image set [8]. The present work also studies alternatives employing emergent structures from the unsupervised clustering of the feature distribution using the Enhanced Tree Growing Neural Gas [13]. They attempt to overcome the curse of dimensionality through data quantization and dimensionality reduction techniques, such as Principal Component Analysis and Modified Locally Linear Embedding [18], which are tailored to the properties of the feature distribution. Although external validation criteria [17] and data separability [5] improve in these embedded distributions, the resulting clustering schemes are not meaningful enough to shape the hierarchy of this memory model. While using absolute values of feature distances may not be reliable due to the curse of dimensionality, which relates to the challenge of clustering the feature distribution, it is still viable to use rankings of these values for object recognition and categorization.

The *Semantic Correlation Graph* self-organizes image features depending on their co-occurrence in object views. Initially, the texture information represented by a limited set of prototype features, together with categorical identifiers, which are the object signatures of their originating object views, determine co-occurrences during a learning procedure. Finally, the co-occurrence of prototype features registered during learning is scrutinized to define categorical identifiers for unseen object views during a recall procedure, which employs a configurable weight distribution between occurrences and co-occurrences of image features. Experimental results indicate this memory model is robust against different weight distributions.

The performance of feature-based object recognition and categorization systems equipped with this memory

framework are cross-compared to those of state-of-the-art approaches found in literature [7–10, 14, 16] utilizing the ETH-80 and the COIL-100 [11] image sets. In general, the results for the ETH-80 image set show a gradual decrease in accuracy from the more abstract to the more concrete levels of the taxonomy, thus reinforcing the concept that the categories from higher abstraction levels are more accurately discriminable, since the characteristics shared by their intra-category individuals are significantly dissimilar in comparison to the ones observed inter-category. In both cases, the artificial systems that employ sequences of object views, either during learning or recall phases, attain the best object recognition and categorization, in agreement with an intuitive interpretation of visual object understanding in biological systems.

References

[1] B. Bolder, *Coordination of an Artificial Visual System with Biological Models*, Shaker Verlag, 2006.

[2] G.S. Donatti and R.P. Würtz, "Memory Organization for Invariant Object Recognition and Categorization", *Tech. Post. of 20th Brazilian Symp. on Comp. Grap. and Image Proc.*, B. Horizonte, 1:11-12, 2007.

[3] G.S. Donatti and R.P. Würtz, "Using Growing Neural Gas Networks to Represent Visual Object Knowledge", *Proc. of 21st IEEE Intl. Conf. on Tools with Artificial Intelligence*, Newark, 1:54-58, 2009.

[4] G.S. Donatti, O. Lomp, and R.P. Würtz, "Evolutionary Optimization of Growing Neural Gas Parameters for Object Categorization and Recognition", *Proceedings of 2010 International Joint Conference on Neural Networks*, Barcelona, 1:1862-1869, 2010.

[5] R.O. Duda, P.E. Hart, and D.G. Stork, *Pattern Classification (2nd Edition)*, Wiley-Interscience, 2000.

[6] B. Fritzke, "A Growing Neural Gas Network Learns Topologies", *Proceedings of Advances in Neural Information Processing Systems 7*, Denver, 1:625-632, 1994.

[7] D. George and J. Hawkins, "Towards a Mathematical Theory of Cortical Micro-circuits", *PLoS Computational Biology*, 5(10):1-26, 2009.

[8] B. Leibe and B. Schiele, "Analyzing Appearance and Contour Based Methods for Object Categorization", *Proc. of Computer Vision and Pattern Recognition Conference*, Madison, 1:409-415, 2003.

[9] M. Leßmann and R.P. Würtz, "Learning Invariant Object Recognition from Temporal Correlation in a Hierarchical Network", *Neural Networks*, 54:70-84, 2014.

[10] O. Linde and T. Lindeberg, "Composed Complex-cue Histograms: An Investigation of the Information Content in Receptive Field Based Image Descriptors for Object Recognition", *Computer Vision and Image Understanding*, 116(4):538-560, 2012.

[11] S.A. Nene, S.K. Nayar, and H. Murase, *Columbia Object Image Library (COIL-100)*, Technical Report CUCS-006-96, Department of Computer Science, Columbia University, 1996.

[12] T.J. Palmeri and I. Gauthier, "Visual Object Understanding", *Nat. Rev. Neuroscience*, 5(4):291-303, 2004.

[13] M. Richter, *Using Enhanced Tree Growing Neural Gas Networks to Represent Knowledge Derived from Artificial and Real-world Data*, B.Sc. Thesis, ET-IT Department, Ruhr University Bochum, 2009.

[14] F. Suard, A. Rakotomamonjy, and A. Bensrhair, "Object Categorization Using Kernels Combining Graphs and Histograms of Gradients", *Proc. of 3rd Intl. Conf. Img. Ana. and Recog.*, P. de Varzim, 1:23-24, 2006.

[15] T. Trappenberg, *Fundamentals of Computational Neuroscience*, Oxford University Press, 2002.

[16] G. Westphal and R.P. Würtz, "Combining Feature- and Correspondence-Based Methods for Visual Object Recognition", *Neural Computation*, 21(7):1952-1989, 2009.

[17] R. Xu and D. Wunsch, *Clustering*, Wiley-IEEE Press, 2009.

[18] Z. Zhang and J. Wang, "MLLE: Modified Locally Linear Embedding Using Multiple Weights", *Proc. of Advances in Neural Information Processing Systems 19*, Vancouver, 1:1593-1600, 2006.