

Invariant recognition of visual objects: some emerging computational principles

[Health & Medicine](#)



Invariant object recognition refers to recognizing an object regardless of irrelevant image variations, such as variations in viewpoint, lighting, retinal size, background, etc. The perceptual result of invariance, where the perception of a given object property is unaffected by irrelevant image variations, is often referred to as perceptual constancy ([Kofka, 1935](#) ; [Walsh and Kulikowski, 2010](#)).

Mechanisms of invariant object recognition have, to a significant extent, remained unclear. This is both because experimental and computational studies have so far largely focused on understanding object recognition without these variations, and because the underlying computational problems are profoundly difficult.

The 10 articles in this Research Topic Issue focus on some of the key computational issues in invariant object recognition. There is no pretending that the articles cover all key areas of current research exhaustively or seamlessly. For instance, none of the articles in this issue address size invariance ([Kilpatrick and Ittelson, 1953](#)) or color constancy ([Foster, 2011](#)). Nonetheless, the articles collectively paint a useful pointillist picture of current research on computational principles of invariance.

Strategies of Representing Invariance

Several articles address various strategies of exploiting or representing the information in the visual image to achieve object invariance. [Chuang et al. \(2012\)](#) show, using psychophysical experiments, that non-rigid motion provides a cue to the invariance of dynamic objects. [Groen et al. \(2012\)](#) show that low-level image statistics can cue the extent to which natural

textures are invariant across samples. Using electroencephalography (EEG), they also show that the differences in edge statistics predict the differences in the evoked neural responses to individual images. Using psychophysical experiments, [Bart and Hegdé \(2012\) 1](#) show that human subjects can use small informative fragments of an image to recognize an object regardless of variations in illumination. A more radical idea is proposed by [Edelman and Shahbazi \(2012\)](#), who argue that representing objects by their similarity to a set of prototypes can explain many properties of the visual system, including invariance.

Strategies of Learning Invariance

In a supervised setting, cues to object invariance may be provided externally (e. g., [Bart and Hegdé, 2012](#)). In unsupervised settings, finding cues to invariance is more challenging. One type of cues arises from the fact that even when an object changes in appearance, the change is generally smooth. Thus, over short, selected stretches of space and/or time, the changes in object appearance tend to be rather small, so that the visual system can, in principle, infer that the same object is changing its appearance. A theoretical approach for exploiting this contiguity is given by the continuous transformation (CT) learning ([Stringer et al., 2006](#)). A related cue arises from the fact that objects often stay in view for extended periods of time; two observations at nearby time points are therefore likely to correspond to the same object. An approach that exploits this temporal contiguity is given by the trace learning rule ([Földiák, 1991](#)).

Many articles in this issue describe models that exploit one or more of these rules to learn object invariance. The VisNet model can incorporate one or both of these strategies, depending on the particular implementation. The article by [Rolls \(2012\)](#) describes the various capabilities of VisNet. The article by [Tromans et al. \(2012\)](#) highlights the capability of VisNet to learn with clutter and occlusion. VisNet, like most neural network models, uses rate coding, in which the firing rate of a neuron determines the information coded by that neuron. The firing rate of a neuron is usually specified as a scalar, without the neuron having to actually fire spikes. The article by [Evans and Stringer \(2012\)](#) implements VisNet in which individual neurons actually fire spikes, and detail the merits of this implementation. The model by [Isik et al. \(2012\)](#) describes a different model, HMAX (also see [Serre et al., 2007](#)), that simulates many invariance properties in the primate visual system.

It is worth noting that, while it is generally thought that object invariance is represented by neurons in the higher levels of the visual pathway, such as the inferotemporal cortex, neurons in the lower levels, such as the primary visual cortex or V1, can also play key roles in implementing various aspects of invariance. The article by [Vidal-Naquet and Gepshtein \(2012\)](#) shows that populations of V1 complex cells, but not individual complex cells, can compute information about stereoscopic disparity in a spatially invariant fashion.

Some Important Caveats

It is important to emphasize a few caveats about the implications of these articles for future research. First, at the perceptual level, object invariance

neither is perfect nor needs to be ([Bülthoff and Edelman, 1992](#) ; [DiCarlo and Cox, 2007](#)). Thus, the underlying neural mechanisms need not deliver perfect invariance. Second, not all types of invariance are equal. Some types of invariance may be more important or useful to the visual system than others, depending on the behavioral context (see [Milivojevic, 2012](#)). Third, the visual system does not necessarily have to rely on prolonged supervised learning to learn invariance. It is possible that the system is able to either learn or, alternatively, infer invariance on the fly, and without any feedback (see [Rolls, 2012](#)). Fourth, top-down factors, such as the behavioral context, play an important role in object invariance and lack thereof. This is not fully addressed by the articles in this issue, which mostly focus on bottom-up processing of invariance information. Finally, for practical reasons, current research tends to deal with invariance along the various individual stimulus parameters (e. g., viewpoint, illumination, etc.) separately from each other. But in actuality, the visual system may combine invariance across multiple visual parameters, and indeed multiple sensory modalities.

Footnote

1. [^] Who are also the editors of this Research Topic Issue and the authors of this editorial.

References

Bart, E., and Hegdé, J. (2012). Invariant object recognition based on extended fragments. *Front. Comput. Neurosci.* 6: 56. doi: 10.3389/fncom.2012.00056

[CrossRef Full Text](#)

<https://assignbuster.com/invariant-recognition-of-visual-objects-some-emerging-computational-principles/>

Bülthoff, H. H., and Edelman, S. (1992). Psychophysical support for a 2-D view interpolation theory of object recognition. *Proc. Natl. Acad. Sci. U. S. A.* 89, 60–64.

[Pubmed Abstract](#) | [Pubmed Full Text](#) | [CrossRef Full Text](#)

Chuang, L. L., Vuong, Q. C., and Bülthoff, H. H. (2012). Learned non-rigid object motion is a view-invariant cue to recognizing novel objects. *Front. Comput. Neurosci.* 6: 26. doi: 10.3389/fncom.2012.00026

[Pubmed Abstract](#) | [Pubmed Full Text](#) | [CrossRef Full Text](#)

DiCarlo, J. J., and Cox, D. D. (2007). Untangling invariant object recognition. *Trends Cogn. Sci. (Regul. Ed.)* 11, 333–341.

[Pubmed Abstract](#) | [Pubmed Full Text](#) | [CrossRef Full Text](#)

Edelman, S., and Shahbazi, R. (2012). Renewing the respect for similarity. *Front. Comput. Neurosci.* 6: 45. doi: 10.3389/fncom.2012.00045

[Pubmed Abstract](#) | [Pubmed Full Text](#) | [CrossRef Full Text](#)

Evans, B., and Stringer, S. (2012). Transform-invariant visual representations in self-organizing spiking neural networks. *Front. Comput. Neurosci.* 6: 46. doi: 10.3389/fncom.2012.00046

[Pubmed Abstract](#) | [Pubmed Full Text](#) | [CrossRef Full Text](#)

Földiák, P. (1991). Learning invariance from transformation sequences. *Neural Comput.* 3. 2, 194–200.

<https://assignbuster.com/invariant-recognition-of-visual-objects-some-emerging-computational-principles/>

[CrossRef Full Text](#)

Foster, D. H. (2011). Color constancy. *Vision Res.* 51, 674–700.

[PubMed Abstract](#) | [PubMed Full Text](#) | [CrossRef Full Text](#)

Groen, I. I. A., Ghebreab, S., Lamme, V. A. F., and Scholte, H. S. (2012). Low-level edge statistics predict invariance of natural textures. *Front. Comput. Neurosci.* 6: 34. doi: 10.3389/fncom.2012.00034

[PubMed Abstract](#) | [PubMed Full Text](#) | [CrossRef Full Text](#)

Isik, L., Leibo, J. Z., and Poggio, P. (2012). Learning and disrupting invariance in visual recognition with a temporal association rule. *Front. Comput. Neurosci.* 6: 37. doi: 10.3389/fncom.2012.00037

[PubMed Abstract](#) | [PubMed Full Text](#) | [CrossRef Full Text](#)

Kilpatrick, F. P., and Ittelson, W. H. (1953). The size-distance invariance hypothesis. *Psychol. Rev.* 60, 223–231.

[PubMed Abstract](#) | [PubMed Full Text](#) | [CrossRef Full Text](#)

Kofka, K. (1935). *Principles of Gestalt Psychology*. New York, NY: Harcourt, Brace and Company.

Milivojevic, B. (2012). Object recognition can be viewpoint dependent or invariant – it's just a matter of time and task. *Front. Comput. Neurosci.* 6: 27. doi: 10.3389/fncom.2012.00027

[PubMed Abstract](#) | [PubMed Full Text](#) | [CrossRef Full Text](#)

<https://assignbuster.com/invariant-recognition-of-visual-objects-some-emerging-computational-principles/>

Rolls, E. T. (2012). Invariant visual object and face recognition: neural and computational bases, and a model, VisNet. *Front. Comput. Neurosci.* 6: 35. doi: 10.3389/fncom.2012.00035

[Pubmed Abstract](#) | [Pubmed Full Text](#) | [CrossRef Full Text](#)

Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M., and Poggio, T. (2007). Robust object recognition with cortex-like mechanisms. *IEEE Trans. Pattern Anal. Mach. Intell.* 29. 3, 411–426.

[Pubmed Abstract](#) | [Pubmed Full Text](#) | [CrossRef Full Text](#)

Stringer, S. M., Perry, G., Rolls, E. T., and Proske, J. H. (2006). Learning invariant object recognition in the visual system with continuous transformations. *Biol. Cybern.* 94, 128–142.

[Pubmed Abstract](#) | [Pubmed Full Text](#) | [CrossRef Full Text](#)

Tromans, J. M., Higgins, I., and Stringer, S. M. (2012). Learning view invariant recognition with partially occluded objects. *Front. Comput. Neurosci.* 6: 48. doi: 10.3389/fncom.2012.00048

[Pubmed Abstract](#) | [Pubmed Full Text](#) | [CrossRef Full Text](#)

Vidal-Naquet, M., and Gepshtein, S. (2012). Spatially invariant computations in stereoscopic vision. *Front. Comput. Neurosci.* 6: 47. doi: 10.3389/fncom.2012.00047

[Pubmed Abstract](#) | [Pubmed Full Text](#) | [CrossRef Full Text](#)

Walsh, V., and Kulikowski, J. (eds). (2010). *Perceptual Constancy: Why Things Look as They Do* . New York, NY: Cambridge University Press.