Hierarchical Object Representations in the Visual Cortex and Computer Vision

Rodriguez-Sanchez, Antonio; Fallah, Mazyar; Leonardis, Ales

DOI:
10.3389/fncom.2015.00142

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (Harvard):
Rodriguez-Sanchez, A, Fallah, M & Leonardis, A 2015, 'Hierarchical Object Representations in the Visual Cortex and Computer Vision', *Frontiers in Computational Neuroscience*, vol. 9, 142.
https://doi.org/10.3389/fncom.2015.00142

frontiers in Computational Neuroscience

EDITORIAL
published: 20 November 2015
doi: 10.3389/fncom.2015.00142

Editorial: Hierarchical Object Representations in the Visual Cortex and Computer Vision

Antonio J. Rodríguez-Sánchez[1]*, *Mazyar Fallah*[2] *and Aleš Leonardis*[3]

[1] *Intelligent and Interactive Systems, Department of Computer Science, University of Innsbruck, Innsbruck, Austria,* [2] *Visual Perception and Attention Laboratory, Centre for Vision Research, School of Kinesiology and Health Science, York University, Toronto, ON, Canada,* [3] *School of Computer Science, University of Birmingham, Birmingham, UK*

Keywords: computer model, neurophysiology, computer vision, visual cortex, computational neurosciences

Over the past 40 years, Neurobiology and Computational Neuroscience have proved that deeper understanding of visual processes in humans and non-human primates can lead to important advancements in computational perception theories and systems. One of the main difficulties that arises when designing automatic vision systems is developing a mechanism that can recognize—or simply find—an object when faced with all the possible variations that may occur in a natural scene, and with the ease of the primate visual system. The area of the brain in primates that is dedicated to analyzing visual information is the visual cortex. The visual cortex performs a wide variety of complex tasks by means of seemingly simple operations. These operations are applied to several layers of neurons organized into a hierarchy, the layers representing increasingly complex, abstract intermediate processing stages.

In this research topic we propose to bring together current efforts in Neurophysiology and Computer Vision in order to better understand (1) How the visual cortex encodes an object from a starting point where neurons respond to lines, bars or edges to the representation of an object at the top of the hierarchy that is invariant to illumination, size, location, viewpoint, rotation and robust to occlusions and clutter; and (2) How the design of automatic vision systems benefits from that knowledge to get closer to human accuracy, efficiency and robustness to variations. In fact, the primate visual system has influenced computer vision systems for decades now since Hubel and Wiesel (1968) simple and complex cells inspired the Neocognitron (Fukushima, 1980). Since then, studies about the primate and human visual systems led the way to many more works on biologically-inspired computational vision, such as Tsotsos et al. (1995); Olshausen and Field (1996); Booth and Rolls (1998); Riesenhuber and Poggio (1999); Rodríguez-Sánchez and Tsotsos (2011), to name a few.

The answers to these issues bring hypotheses that are partially addressed in this research topic, raising additional new questions:

1. What are the mechanisms involved in these visual architectures? What are the limitations of feedforward connections? When is feedback and top-down priming necessary? The classical way of seeing feedback connections is for the enhancement of neural responses through top-down attentive processes (Moran and Desimone, 1985; Rodríguez-Sánchez et al., 2006; Perry et al., 2015). But lately, other studies support a role of feedback connections related to cell selectivity through recurrent networks (Neumann and Sepp, 1999; Angelucci and Bressloff, 2006).
2. The ventral stream areas (V1, V2, V4, inferotemporal cortex) have usually been considered to be the ones involved in object recogntion and the subject of several existing models (Serre et al., 2006; Rodríguez-Sánchez and Tsotsos, 2012). But, also recently, there are new findings that relate the dorsal stream with that same task (Konen and Kastner, 2008; Perry and Fallah, 2012). What are the differences between how objects are processed in the ventral and the dorsal streams? Which areas are involved in recognition and which in localization?

Edited by:
Si Wu,
Beijing Normal University, China

Reviewed by:
Da-Hui Wang,
Beijing Normal University, China

***Correspondence:**
Antonio J. Rodríguez-Sánchez
antonio.rodriguez-sanchez@uibk.ac.at

Received: *21 August 2015*
Accepted: *06 November 2015*
Published: *20 November 2015*

Citation:
Rodríguez-Sánchez AJ, Fallah M and Leonardis A (2015) Editorial: Hierarchical Object Representations in the Visual Cortex and Computer Vision.
Front. Comput. Neurosci. 9:142.
doi: 10.3389/fncom.2015.00142

3. And finally, how much is learned and how much is genetically implemented (Rodríguez-Sánchez and Piater, 2014)? Even more, what is the relation between learning, sparse coding, selectivity and diversity (Olshausen and Field, 1996; Xiong et al., 2015) and how different learning strategies compare?

We present a total of 19 papers related to those questions. The following five papers deal with the questions related to visual architectures and their mechanisms. Ghodrati et al. (2014) studied whether recent relative successes in object recognition on various image datasets based on sparse representations applied in a feedforward fashion represented a breakthrough in invariant object recognition. In their study they showed, using a carefully designed parametrically controlled image database consisting of several object categories, that these approaches fail when the complexity of image variations is high and that their performance is still poor compared to humans. This suggests that learning sparse informative visual features may be one of the necessary components but definitely not a complete solution for a human-like object recognition system. A classical feedforward filtering approach is also challenged in the paper by Herzog and Clarke (2014), where the authors provided ample evidence, stemming from experiments from crowding research, to support their arguments that the computations are not purely local and feedforward, but rather global and iterative. On the same topic, Tal and Bar (2014) explored the role of top-down mechanisms which bias the processing of the incoming visual information and facilitate fast and robust recognition. This work specifically addresses the question of what happens to initial predictions that eventually get rejected in a competitive selection process. The work by Marfil et al. (2014) brings into focus another important aspect of biological visual sytems, namely attention. The authors studied a bidirectional relationship between segmentation and attention processes. They presented a bottom-up foveal attention model that demonstrates how the attention process influences the selection of the next position of the fovea and how segmentation, in turn, guides the extraction of units of attention. In Han and Vasconcelos (2014) the authors also researched the role of attention models, but this time in connection to object recognition. Using their recognition model, hierarchical discriminant saliency network (HDSN), they clearly demonstrated the benefits of integrating attention and recognition.

We provide an interesting discussion on the role of ventral and dorsal streams with a total of 10 articles. Kubilius et al. (2014) discusses the importance of surface representation and reviews recent work on mid-level visual areas in the ventral stream. We include here two models of shape related to those intermediate visual areas. The first approach is a recurrent network that achieves figure-ground segregation by assigning border ownership through the interaction between feedforward and feedback inputs (Tschechne and Neumann, 2014). The second approach is a trainable set of shape detectors that can be applied as a filter bank to recognize letters and keywords as well finding objects in complex scenes (Azzopardi and Petkov, 2014). The question that arises regarding computational models is of course, how faithful they are? This is what Ramakrishnan et al. (2015) answers by comparing the fMRI responses from 20 subjects to two different types of computer vision models: the classical bag of words and the biologically-inspired HMAX. HMAX is also the subject of study in Zeman et al. (2014), here the authors use that model to compare the robustness of complex cells to simple cells in the Müller-Lyer illusion. The final stage in the object recognition pathway is the inferotemporal cortex (IT), Leeds et al. (2014) present an fMRI study that tries to answers the problem of how starting from simple edge-like features in V1 we obtain neurons at the top of the hierarchy that respond to complex features as parts, textures or shapes. Using feed-forward object detection and classification modeling, Khosla et al. (2014) developed a neuromorphic system that also efficiently produces automated video object recognition. However, the visual system is not limited to only detecting objects, but can also detect the spatial relationships between objects and even between parts of the same object. The dorsal stream areas are thus also important for object representation with a focus on action via effectors such as the eyes or the hand. Theys et al. (2014) reviews how 3D shape for grasping is processed along the dorsal stream, focusing on the representations in the anterior intraparietal area (AIP) and ventral premotor cortex (PMv). Rezai et al. (2014) advances this by modeling the curvature and gradient input from the caudal intraparietal area (CIP) to visual neurons in AIP, using superquadric fits—used in robotics for grasp planning—or Isomap dimension reductions of object surface distances. They found that both models fit responses from primate AIP neurons. However, Isomaps better approximated the feedforward input from CIP making it the more promising model of how the dorsal stream produces shape representations for grasping. Yet the features used for grasping are only a subset of an object's features. While the integration of features along the ventral stream to form object representations is well-known, Perry and Fallah (2014) review recent findings supporting dorsal stream object representations and propose a framework for the integration of features along the dorsal stream.

Finally, four papers address the problem of learning and sparse coding. Rinkus (2014) shows that a hierarchical sparse distributed code network provides the foundation for the storage and retrieval of associative memory on top of building up an object representation. The end point of object processing is recognition, which the human visual system is very efficient at and many computational models are based upon. Webb and Rolls (2014) investigated how recognition of the identity of individuals and their poses can be separated. They showed that a model of the ventral visual system using temporal continuity, VisNet, can through learning develop pose-specific and identity-specific representations that are invariant to the other factor. In their biologically inspired study, Kermani Kolankeh et al. (2015) researched different computational principles (sparse coding, biased competition, Hebbian learning) capable of developing receptive fields comparable to those of V1 simple-cells and discovered that methods which employ competitive mechanisms achieve higher levels of robustness against loss of information which may be important to achieve better performance on classification tasks. While these studies have focused on using

biologically-inspired visual processing in computational models, Bertalmío (2014) worked in reverse by taking an image processing technique used for local histogram equalization and applying it to a neural activity model. The resultant model predicts spectrum whitening, contrast enhancement and lightness induction, all behavioral aspects of visual processing. Time will tell if neuronal studies bear out this process.

We are bringing together two seemingly different disciplines: Neuroscience and Computer Vision. We show in this research topic that each one can benefit from the other. The latter can aid Neuroscience for testing hypotheses regarding the visual cortex in a non-invasive way, or otherwise when we reach technical limitations, e.g., how the information flows along the visual architectures (see Rodríguez-Sánchez, 2010 for a recent example). On the other hand, Computer Vision can benefit from Neuroscience in order to develop better, more robust, efficient and general systems than the ones present to date (Krüger et al., 2013).

Due to the complexity of vision (Tsotsos, 1987), objects/locations are considered to *compete* for the visual system's resources. The studies presented here show that—among other aspects—feedforward hierarchies are insufficient, supporting the need for top-down priming or attention. The interaction between feedforward and feedback inputs have an impact in neural encoding as shown in the models presented in this research topic. Not only competition, sparsity is another important mechanism. The aim is achieving efficient codes that represent and store object classes efficiently into memory since not every possible combination of features/parameters is feasible to be stored. Finally, a number of studies stress on the importance of the dorsal stream in shape and identity-object representation in order to interact with specific objects, e.g., grasping.

REFERENCES

Angelucci, A., and Bressloff, P. C. (2006). Contribution of feedforward, lateral and feedback connections to the classical receptive field center and extra-classical receptive field surround of primate V1 neurons. *Prog. Brain Res.* 154, 93–120. doi: 10.1016/S0079-6123(06)54005-1

Azzopardi, G., and Petkov, N. (2014). Ventral-stream-like shape representation: from pixel intensity values to trainable object-selective cosfire models. *Front. Comput. Neurosci.* 8:80. doi: 10.3389/fncom.2014.00080

Bertalmío, M. (2014). From image processing to computational neuroscience: a neural model based on histogram equalization. *Front. Comput. Neurosci* 8:71. doi: 10.3389/fncom.2014.00071

Booth, M., and Rolls, E. (1998). View-invariant representations of familiar objects by neurons in the inferior temporal visual cortex. *Cereb. Cortex* 8, 510–523. doi: 10.1093/cercor/8.6.510

Fukushima, K. (1980). Neocognitron: a self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol. Cybernet.* 36, 193–202. doi: 10.1007/BF00344251

Ghodrati, M., Farzmahdi, A., Rajaei, K., Ebrahimpour, R., and Khaligh-Razavi, S. M. (2014). Feedforward object-vision models only tolerate small image variations compared to human. *Front. Comput. Neurosci.* 8:74. doi: 10.3389/fncom.2014.00074

Han, S., and Vasconcelos, N. (2014). Object recognition with hierarchical discriminant saliency networks. *Front. Comput. Neurosci.* 8:109. doi: 10.3389/fncom.2014.00109

Herzog, M. H., and Clarke, A. M. (2014). Why vision is not both hierarchical and feedforward. *Front. Comput. Neurosci.* 8:135. doi: 10.3389/fncom.2014.00135

Hubel, D., and Wiesel, T. (1968). Receptive fields and functional architecture of monkey striate cortex. *J. Physiol.* 195, 215–243. doi: 10.1113/jphysiol.1968.sp008455

Kermani Kolankeh, A., Teichmann, M., and Hamker, F. H. (2015). Competition improves robustness against loss of information. *Front. Comput. Neurosci.* 9:35. doi: 10.3389/fncom.2015.00035

Khosla, D., Chen, Y., and Kyungnam, K. (2014). A neuromorphic system for video object recognition. *Front. Comput. Neurosci.* 8:147. doi: 10.3389/fncom.2014.00147

Konen, C. S., and Kastner, S. (2008). Two hierarchically organized neural systems for object information in human visual cortex. *Nat. Neurosci.* 11, 224–231. doi: 10.1038/nn2036

Krüger, N., Janssen, P., Kalkan, S., Lappe, M., Leonardis, A., Piater, J., et al. (2013). Deep hierarchies in the primate visual cortex: what can we learn for computer vision? *IEEE Trans. Pattern Anal. Mach. Intell.* 35, 1847–1871. doi: 10.1109/TPAMI.2012.272

Kubilius, J., Wagemans, J., and Op de Beeck, H. P. (2014). A conceptual framework of computations in mid-level vision. *Front. Comput. Neurosci* 8:158. doi: 10.3389/fncom.2014.00158

Leeds, D. D., Pyles, J. A., and Tarr, M. J. (2014). Exploration of complex visual feature spaces for object perception. *Front. Comput. Neurosci.* 8:106. doi: 10.3389/fncom.2014.00106

Marfil, R., Palomino, A. J., and Bandera, A. (2014). Combining segmentation and attention: a new foveal attention model. *Front. Comput. Neurosci.* 8:96. doi: 10.3389/fncom.2014.00096

Moran, J., and Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science* 229, 782–784. doi: 10.1126/science.4023713

Neumann, H., and Sepp, W. (1999). Recurrent V1–V2 interaction in early visual boundary processing. *Biol. Cybernet.* 81, 425–444. doi: 10.1007/s004220050573

Olshausen, B., and Field, D. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* 381, 607–609. doi: 10.1038/381607a0

Perry, C. J., and Fallah, M. (2012). Color improves speed of processing but not perception in a motion illusion. *Front. Psychol.* 3:92. doi: 10.3389/fpsyg.2012.00092

Perry, C. J., and Fallah, M. (2014). Feature integration and object representations along the dorsal stream visual hierarchy. *Front. Comput. Neurosci.* 8:84. doi: 10.3389/fncom.2014.00084

Perry, C. J., Sergio, L. E., Crawford, J. D., and Fallah, M. (2015). Hand placement near the visual stimulus improves orientation selectivity in V2 neurons. *J. Neurophysiol.* 113, 2859–2870. doi: 10.1152/jn.00919.2013

Ramakrishnan, K., Scholte, H. S., Groen, I. I. A., Smeulders, A. W., and Ghebreab, S. (2015). Visual dictionaries as intermediate features in the human brain. *Front. Comput. Neurosci.* 8:168. doi: 10.3389/fncom.2014.00168

Rezai, O., Kleinhans, A., Matallanas, E., Selby, B., and Tripp, B. P. (2014). Modeling the shape hierarchy for visually guided grasping. *Front. Comput. Neurosci.* 8:132. doi: 10.3389/fncom.2014.00132

Riesenhuber, M., and Poggio, T. (1999). Are cortical models really bound by the "binding problem"? *Neuron* 24, 87–93. doi: 10.1016/S0896-6273(00)80824-7

Rinkus, G. J. (2014). Sparsey^{TM}: event recognition via deep hierarchical sparse distributed codes. *Front. Comput. Neurosci.* 8:160. doi: 10.3389/fncom.2014.00160

Rodríguez-Sánchez, A. (2010). *Intermediate Visual Representations for Attentive Recognition Systems*. PhD thesis, York University, Department of Computer Science and Engineering.

Rodríguez-Sánchez, A., and Tsotsos, J. (2011). "The importance of intermediate representations for the modeling of 2D shape detection: endstopping and curvature tuned computations," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Colorado Springs, CO), 4321–4326. doi: 10.1109/cvpr.2011.5995671

Rodríguez-Sánchez, A. J., and Piater, J. (2014). "Models of the visual cortex for object representation: learning and wired approaches," in

Brain-Inspired Computing, Vol. 8603 of *Lecture Notes in Computer Science*, eds L. Grandinetti, T. Lippert, and N. Petkov (Springer International Publishing), 51–62.

Rodríguez-Sánchez, A. J., Simine, E., and Tsotsos, J. K. (2006). "Feature conjunctions in visual search," in *Artificial Neural Networks (ICANN)*, eds S. Kollias, A. Stafylopatis, W. Duch, and E. Oja (Athens: Springer), 498–507. doi: 10.1007/11840930_52

Rodríguez-Sánchez, A. J., and Tsotsos, J. K. (2012). The roles of endstopped and curvature tuned computations in a hierarchical representation of 2D shape. *PLoS ONE* 7:e42058. doi: 10.1371/journal.pone.0042058

Serre, T., Oliva, A., and Poggio, T. (2006). A feedforward architecture accounts for rapid categorization. *Proc. Natl. Acad. Sci. U.S.A.* 104, 6424–6429. doi: 10.1073/pnas.0700622104

Tal, A., and Bar, M. (2014). The proactive brain and the fate of dead hypotheses. *Front. Comput. Neurosci.* 8:138. doi: 10.3389/fncom.2014.00138

Theys, T., Romero, M. C., van Loon, J., and Janssen, P. (2014). Shape representations in the primate dorsal visual stream. *Front. Comput. Neurosci.* 8:43. doi: 10.3389/fncom.2015.00043

Tschechne, S., and Neumann, H. (2014). Hierarchical representation of shapes in visual cortex - from localized features to figural shape segregation. *Front. Comput. Neurosci.* 8:93. doi: 10.3389/fncom.2014.00093

Tsotsos, J. K. (1987). A complexity level analysis of immediate vision. *Int. J. Comput. Vis.* 1, 303–320.

Tsotsos, J. K., Culhane, S. M., Wai, W. Y. K., Lai, Y. H., Davis, N., and Nuflo, F. (1995). Modeling visual-attention via selective tuning. *Artif. Intell.* 78, 507–545. doi: 10.1007/BF00133569

Webb, T. J., and Rolls, E. T. (2014). Deformation-specific and deformation-invariant visual object recognition: pose vs. identity recognition of people and deforming objects. *Front. Comput. Neurosci.* 8:37. doi: 10.3389/fncom.2014.00037

Xiong, H., Rodríguez-Sánchez, A. J., Szedmak, S., and Piater, J. (2015). Diversity priors for learning early visual features. *Front. Comput. Neurosci.* 9:104. doi: 10.3389/fncom.2015.00104

Zeman, A., Obst, O., and Brooks, K. R. (2014). Complex cells decrease errors for the Müller-Lyer illusion in a model of the visual ventral stream. *Front. Comput. Neurosci.* 8:112. doi: 10.3389/fncom.2014.00112

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.