NEUCOM 141
pp: 1-6 (col.fig.: Nil)
PROD. TYPE: COM
ED: Chan
PAGN: Murthy.N – SCAN: Padma

ARTICLE IN PRESS

ELSEVIER

Neurocomputing ▮▮▮ (▮▮▮▮) ▮▮▮–▮▮▮

NEUROCOMPUTING

www.elsevier.com/locate/neucom

Recognizing the gist of a visual scene: possible perceptual and neural mechanisms

Christoph Rasche*, Christof Koch

Caltech, Division of Biology 139-74, Pasadena CA91125, USA

Abstract

We try to understand the basics of human image processing from a gist recognition perspective. Because the gist is only a subset of the image's information, we think that it is extracted with help of interpretation (feedback). In a perceptual section we list possible mechanisms that the interpretation process uses to determine the gist: in addition to the commonly known local-to-global perception evolvement, there is likely to be also a global-to-local evolvement direction, a coarse/fine scale, as well as a foreground/background scale. In a neural section we first summarize feedback connections that can possibly be involved in gist recognition. Second, we propose that the perceptual mechanisms are spread all over the cortex and that cortical visual computation occurs distributively rather than hierarchical. © 2002 Published by Elsevier Science B.V.

Keywords: Vision; Gist; Scales; Feedback

1. Introduction

We humans are able to understand the gist of a suddenly presented image in very short time, presumably before the first saccade (around 200 ms) is launched [15,20]. How can the visual system extract this subset of the image information in such short time? Scene perception is generally understood as an interaction of a bottom-up and top-down component in information flow. But does this scheme also hold for gist recognition? Recent models propose that only a pure bottom-up takes place during rapid visual processing [22]. We list here evidence that gist recognition occurs distributively and quite possibly using feedback.

* Corresponding author.
E-mail address: rasche@klab.caltech.edu (C. Rasche).

PII: S0925-2312(02)00500-3

Models of scene perception try to explain how the visual system performs a local-to-global perception along a hierarchy of neocortical, visual areas. In the first step, the entire image is preprocessed in lower (early) cortical areas by dissecting it into its local features. Then, perception is built up by fusing the local features together to a global percept in higher areas [14,6]. Attentional feedback from higher cortical areas would help fusing the corresponding features to form object perceptions in scenes [21]. We see two major problems with this local-to-global feedback approach: (1) The cortical feedback connections, motivating the attentional feedback, are high in the hierarchy and might be too slow for gist recognition. (2) Building up the percept from only local features, even with help of feedback, is a tricky task. We therefore search for earlier feedback connections and for additional perceptual mechanisms (other than the local-to-global) supporting a fast gist recognition.

2. Perceptual mechanisms

Because we can recognize the gist so rapidly, some psychologists have proposed that we understand a scene by decomposing it in a global-to-local manner, meaning that we first perceive the global characteristics in an image and then later proceed to details ('forest before trees', [11]). Navon's study has triggered a debate about the local/global preference and presently it is discussed, whether local and global processing occurs concurrently [8].

An image can be processed on different resolutional scales [3]. The original input image represents the fine scale, the low-pass filtered (blurred) image represents a coarse scale. Oliva and Schyns [13] argue that depending on the task, a coarse-to-fine or a fine-to-coarse scale processing can take place. Furthermore, they argue that on the coarse/fine scale a local/global perception can take place and thus these scales are to be treated as independent scales (Fig. 1a).

A similar discussion takes place regarding foreground and background [9]. One can understand the gist by firstly perceiving a few key objects (foreground) of the scene, followed by concluding to the embedding context (background). Vice versa, the visual system could also firstly process the background followed by the objects within (Fig. 1a).

In hardly any of the studies on these above debates could it be concluded that an image is parsed in only one direction of a scale. We therefore regard a picture of a real-world scene as an individual, which needs to be processed individually. Thus, the interpretational power, using these different evolvement scales, must be highly dynamic and flexible.

3. Neural mechanisms

There are two additional main types that can act earlier than the late cortico-cortical feedback (Fig. 1b): (1) Neocortical areas often project back to the lateral geniculate nucleus (LGN) in the thalamus. Feedback from V1 to LGN seems to be 10 times as

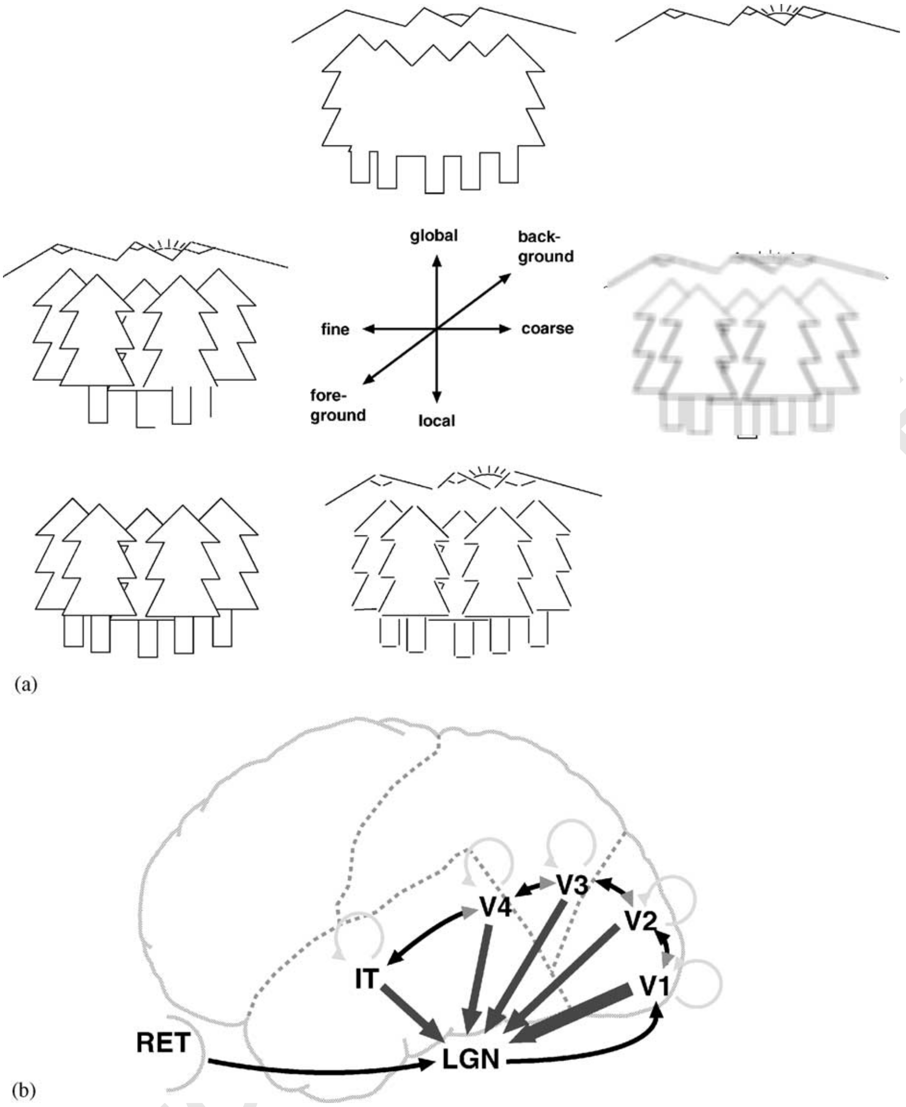

Fig. 1. (a) Perceptual evolvement directions involved in gist recognition. (b) Main recurrent connections in the visual system found so far. For simplicity, only a few cortical visual areas are denoted (V1–V4, IT). Three main types of feedback have been found in the visual brain: (1) From cortical areas to LGN. (2) Feedback connections across layers in a cortical area (recycling arrow). (3) 'Classic' feedback connections from higher cortical areas (IT) towards lower ones (V1), which have been used to explain scene recognition models.

large as feedforward connections from LGN to V1 [17]. (2) There are substantial local recurrent connections across different layers in V1 [5,2]. We posit that some of this earlier feedback must be used for gist recognition.

In the local-to-global view, distinction between foreground/background and global/local information would occur in high cortical areas due to the gradual built-up of the scene. Two lines of evidence suggest otherwise: (1) Earliest mammals had only V1 and V2 [1]. We assume that these areas did some sort of global processing (in order to distinguish predator from prey or background) and that this global perception has been principally conserved throughout evolution. (2) Many hints indicate that cells in V1 also encode global characteristics: Gestalt principles might be already implemented in V1 [18]; information for object size judgment might exist in all visual areas [4]; V1 orientation selective cells signal some sort of foreground/background distinction [19,10,23]. In conclusion, the summarized perceptual evolvement directions are not layouted in a particular order along the hierarchy, but exist locally in many cortical areas.

The latter conclusion questions the pure usage of the hierarchy. We critizise it from two further aspects: (1) Anatomical studies have revealed many hierarchical discrepancies: feedforward and feedback connections do not only project to the next higher or lower level in the hierarchy, respectively, but also jump one or more levels [7]. (2) Spike timing recordings have shown that higher areas can be activated earlier than lower ones [12,16]. Some visual information could be first processed in higher areas and then fed back to lower ones. We therefore think that visual input is almost concurrently spread across several areas and then processed by communicating with each other rapidly in a distributive manner.

4. Summary

(1) Gist recognition can occur on three different recognition evolvement scales (Fig. 1a) driven by the interpretational process.
(2) These evolvement scales are distributed all over neocortex.
(3) Scenes are thus processed very distributively using likely the many feedback loops in the visual system—even before the first saccade.

Acknowledgements

This work was supported in part by the Engineering Research Centers Program of the National Science Foundation under Award Number EEC-9402726 and in part by the Swiss National Science Foundation.

References

[1] J.M. Allman, Evolution of the visual system in the early primates, in: J.M. Sprague, A.M. Epstein (Eds.), Progress in Psychology and Physiological Psychology, Vol. 7, 1977, pp. 1–53.
[2] E.M. Callaway, Local circuits in primary visual cortex of the macaque monkey, Annu. Rev. Neurosci. 21 (1998) 47–74.

[3] R.L. DeValois, K.K. DeValois, Spatial Vision, Oxford University Press, New York, 1988.
[4] A.C. Dobbins, R.M. Jeo, J.Fiser, J.M. Allman, Distance modulation of neural activity in the visual cortex [see comments]. Science 281(5376) (1998) 552–555. Comment in: Science 281(5376) (1998) 500–501.
[5] R.J. Douglas, C. Koch, M.A. Mahowald, K.A.C. Martin, H. Suarez, Recurrent excitation in neocortical circuits, Science 269 (1995) 981–985.
[6] M.J. Farah, The Cognitive Neuroscience of Vision, Blackwell Publishers, Malden, Mass, 2000.
[7] D.J. Felleman, D.C. Van Essen, Distributed hierarchical processing in the primate cerebral cortex, Cereb. Cortex 1 (1991) 123–200.
[8] H.J. Heinze, H. Hinrichs, M. Scholz, W. Burchert, G.R. Mangun, Neural mechanisms of global and local processing. A combined pet and erp study [see comments]. J. Cognition Neurosci. 10(4) (1998) 485–498. Comment in: J. Cognition Neurosci. 12(2) (2000) 355–356; discussion 357–359.
[9] J.M. Henderson, A. Hollingworth, High-level scene perception, Annu. Rev. Psychol. 50 (1999) 243–271.
[10] V.A. Lamme, V. Rodriguez-Rodriguez, H. Spekreijse, Separate processing dynamics for texture elements, boundaries and surfaces in primary visual cortex of the macaque monkey, Cereb. Cortex 9 (4) (1999) 406–413.
[11] D. Navon, Forest before trees: the precedence of global features in visual perception, Cognitive Psychol. 9 (1977) 353–383.
[12] L.G. Nowak, J. Bullier, The timing of information transfer in the visual system, in: K. Rockland, J.H. Kaas, A. Peters (Eds.), Cerebral Cortex, Plenum press, New York, 1997, pp. 205–241. (Chapter 5)
[13] A. Oliva, P.G. Schyns, Coarse blobs or fine edges? evidence that information diagnosticity changes the perception of complex visual stimuli, Cognitive Psychol. 34 (1) (1997) 72–107.
[14] S.E. Palmer, Vision Science: Photons to Phenomenology, MIT Press, Cambridge, MA, 1999.
[15] M.C. Potter, Meaning in visual search, Science 187 (4180) (1975) 965–966.
[16] M.T. Schmolesky, Y. Wang, D.P. Hanes, K.G. Thompson, S. Leutgeb, J.D. Schall, A.G. Leventhal, Signal timing across the macaque visual system, J. Neurophysiol. 79 (6) (1998) 3272–3278.
[17] M.S. Sherman, C. Koch, Thalamus, in: The Synaptic Organization of the Brain, 4th Edition, Oxford University Press, Oxford, 1998, pp. 289–328.
[18] L. Spillmann, From elements to perception: local and global processing in visual neurons, Perception 28 (12) (1999) 1461–1492.
[19] Y. Sugita, Grouping of image fragments in primary visual cortex, Nature 401 (6750) (1999) 269–272.
[20] S. Thorpe, D. Fize, C. Marlot, Speed of processing in the human visual system, Nature 381 (1996) 520–522.
[21] A. Treisman, Features and objects: the fourteenth bartlett memorial lecture, Quart. J. Exp. Psychol. A 40 (2) (1988) 201–237.
[22] R. VanRullen, J. Gautrais, A. Delorme, S. Thorpe, Face processing using one spike per neurone, Biosystems 48 (1–3) (1998) 229–239.
[23] H. Zhou, H.S. Friedman, R. von der Heydt, Coding of border ownership in monkey visual cortex, J. Neurosci. 20 (17) (2000) 6594–6611.

Christoph Rasche received a diploma in Zoology and a Ph.D. degree in computational neuroscience at the University of Zürich. He is currently in the Klab at Caltech, studying object and scene recognition, and neural coding.

Christof Koch runs a multi-disciplinary lab at Caltech focusing on consciousness, attention, neurons and visual recognition. He is author of ‘Biophysics of Computation: Information Processing in Single Neurons’, coeditor of ‘Methods in Neuronal Modeling’.