A Biomimetic Vision Architecture

Bruce A. Draper
Department of Computer Science
Colorado State University
Fort Collins, CO, 80523, U.S.A.
draper@cs.colostate.edu

Abstract. The goal of biomimetic vision is to build artificial vision systems that are analogous to the human visual system. This paper presents a software architecture for biomimetic vision in which every major component is clearly defined in terms of its function and interface, and where every component has a analog in the regional functional anatomy of the human brain. We also present an end-to-end vision system implemented within this framework that learns to recognize objects without human supervision.

Keywords: Biomimetic vision, biologically-inspired vision, object recognition, computer vision architecture

Introduction

Biomimetic vision research hypothesizes that in the long run artificial vision systems will be more robust, more adaptable and easier to work with if they mimic human vision. After all, the design is proven, and it will be easier to interact with robots and other artificial agents if they see the world more or less the same way we do. Unfortunately, in the short term biomimetic vision can seem unnecessarily difficult. Mimicking human vision is often an indirect route to solving a computer vision task.

This paper presents a software architecture for biomimetic vision. The goal of the architecture is to describe the major components of human vision and the interactions among them. The components are at the level of the regional functional anatomy of the human brain, and of complex subsystems in computer vision. The goal to build a computer vision system whose major components are functionally analogous to anatomical brain centers, so that its macro-level design is similar to human vision.

The emphasis here is on large-scale components. We are less concerned with how the component modules are implemented. After all, computer hardware is very different from neural "wetware", and software components correspond to massive networks of heterogeneous neurons. At this level of abstraction, we allow the software components to be implemented by standard algorithms, and do not restrict ourselves to neural networks.

The software architecture is described in terms of modules and interfaces. Many systems could be implemented within this framework that match the top-level architecture of human vision, although some will be better than other in terms of performance and/or biological fidelity. We describe and demonstrate a system

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de

developed within the architecture, and discuss alternative implementations for many components. We conclude with comments about aspects of human vision that are not yet reflected in the architecture, and how the architecture and system will be expanded to include them in the future.

The Regional Functional Anatomy of Human Vision

The gross regional functional anatomy of the human visual system is well-known. The early vision system includes the retina, the dorsal lateral geniculate nucleus of the thalamus (LGNd), the superior colliculus of the midbrain, and cortical regions V1 through V4. Beyond early vision the system splits into the ventral and dorsal streams. The ventral stream includes the lateral occipital complex (LOC) and posterior inferotemporal cortex (pIT). It processes object properties for tasks such as object recognition and landmark-based navigation. The dorsal stream includes region V3a, the mediotemporal cortex (MT), and structures in the posterior parietal cortex. It processes spatial and movement properties for tasks such as tracking, ego-motion estimation, and hand-eye coordination. The two streams converge on associative memories in the anterior inferior temporal cortex (aIT), the angular gyrus and area 19. The associative memories in turn communicate with the dorsolateral prefrontal cortex, which closes the loop by providing feedback to LGNd and superior colliculus through pathways that include the frontal eye field. For accessible overviews of the anatomy of human vision, see Milner & Goodale [1], Kosslyn [2] or Palmer [3].

Regional functional anatomy does not by itself define an architecture. Architectures specify both components and interfaces. This paper defines an architecture with interfaces inferred from behavioral studies, lesion studies, brain imaging techniques and electro-physical recordings. The architecture is limited to the task of object recognition. It does not consider visual tasks such as tracking or ego-motion estimation that are computed in the dorsal visual stream, allowing us to concentrate on the early vision system, the ventral stream and associative memories. In the past, we have looked at the even more limited task of recognizing highly familiar objects, a.k.a. expert object recognition [4]. This paper extends that work to more general cases of object recognition. The result is a new model of the inferotemporal cortex and its relation to associative memory, as well as refined models of early vision and the lateral occipital complex.

A Biomimetic Software Architecture

The biomimetic architecture we propose formalizes the major components of the human visual system and adds well-defined interfaces, and is shown in Figure 1. Readers may already be familiar with rough functional characterizations of many of the modules. In particular, Figure 1 shows four modules outlined in black: the early visual system (attention and retinotopic processing), lateral occipital complex (feature extraction), posterior inferotemporal cortex (object recognition) and associative memories (object identification). As discussed below, the most distinctive part of this

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de

architecture lies in the definition of the object recognition module, which models pIT, and its interface to the associative memories. Object recognition is defined as an unsupervised clustering task, not a supervised (or even unsupervised) labeling problem. Labeling and other forms of cross-modal associations are modeled in the associative memories, which operate over clusters, not samples.

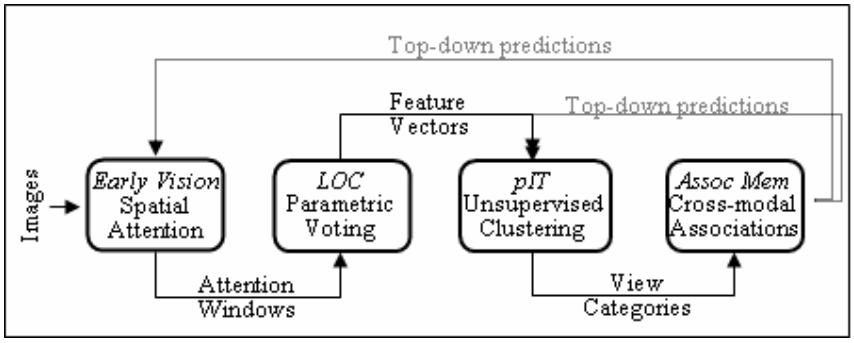

Fig. 1. The biomimetic architecture. *LOC* refers to the lateral occipital complex, *pIT* to the posterior inferotemporal cortex, and *Assoc Mem* to associative memories. Arrows in gray are not yet implemented.

Figure 1 also shows arrows in light gray which are part of top-down rather than bottom-up object recognition and which have not yet been added to the model. Some of these top-down connections pass through the dorsolateral prefrontal cortex and frontal eye field. Without these connections, the architecture models object recognition in the absence of context. In fact, most recognition is highly predictive, and we intend to add top-down recognition in the near future.

Early Vision

Architectural Description

The early vision system is modeled as a spatial selective attention function. It consumes raw images and top-down predictions, and produces image windows defined in terms of image positions and scales. The function should optimize stability in the sense that if the same object appears in two images at different positions and scales but from the same 3D viewpoint (and under similar illumination), the system should center attention windows at the same positions and relative sizes on the object.

Biological Justification

The early vision system is perhaps the most thoroughly studied part of human neuro-anatomy. Decades of study have produced detailed models of ganglion cell responses in the retina and the parvocellular, magnocellular and interlaminar layers of LGNd. Types of known orientation-selective cells in V1 include simple cells, complex cells,

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de

end-stopped cells and grating cells, to name just a few. Other cells are sensitive to colors, disparities or motions.

For all the discussion of edge sensitivity and feature maps, however, the products of early vision are spatial attention windows. The early vision system is retinotopic, which is to say that every cell has a fixed receptive field in the retina (although they also receive efferent inputs), and neighboring cells generally have neighboring receptive fields. Features in the early vision system are therefore kept in a 2D spatial format. Feature maps in the early vision system also cover the entire retinal image, creating essentially a series of image buffers. Moreover, the early vision system is almost the only part of the brain with this organization. As a result, it is a valuable resource: mental imagery recruits image buffers top-down to reconstitute images from memory [5], and tactile input triggers V1 when subjects read Braille [6].

Why would the brain compute any feature across the entire retinal image? It requires far fewer neurons to compute features downstream in LOC, where the computation is restricted to attention windows. If we assume that vision is efficient, the only features computed in the early vision system should be those needed for selective attention. This is why we model early vision as a spatial attention engine, with one caveat: some dorsal pathway tasks such as ego-motion estimation rely on extra-attentional features computed over the full field of view. Motion features are also needed for spatial attention, however, so the general rule still holds: only features needed for selective attention are computed in early vision.

We should be careful to distinguish among types of attention, particularly overt from covert attention, and spatial attention from feature-based or object-based attention. Overt attention refers to movements of the eyes and head to fixate gaze on points in 3D space. This paper models covert spatial attention, which is the selection of (not necessarily foveal) windows within the retinal image for further processing. Covert attention cannot be externally observed, but it can be measured at the neural level throughout the early vision system [7]. Unfortunately, because covert attention cannot be externally observed, we do not know its average dwell time or whether it is sequential or coarsely parallel. As a result, we do not know how many spatial attention windows can be selected per second. Covert spatial attention is also different from feature-based or object-based attention, which selects or discards data further downstream.

Direct evidence that spatial attention selects windows in terms of position and scale comes from Grill-Specter [8], who used repetition suppression effects in fMRI to show that the input to LOC from the early vision system was unchanged when the stimulus was translated or scaled within a factor of 2. Oddly, the same study showed that human spatial attention does not impart rotational invariance, despite evidence from computational systems such as SIFT [9] that attention windows can compensate for image rotations as well.

Implementation of Early Vision

We implemented early vision as finding local maxima in multi-scale DoG responses. This approach was first proposed by Koch and Ullman [10], and has been refined over the years to form the basis of both NVS [11] and SIFT [9]. Our implementation is based on NVS, but was modified to select scales as well as positions and to be less sensitive to image transformations [12].

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de

Whether DoG responses are good biological models of bottom-up spatial attention in humans is debatable. Parkhurst et al [13] and Ouerhani et al [14] show better-than-random correspondence between DoG responses and human eye tracking data. Eye tracking, however, measures overt rather than covert attention, and Privetera and Stark [15] show that almost any high-frequency feature has a better-than-random correspondence to eye tracking data. Kadir and Brady [16] have proposed an alternative model of bottom-up salience based on local entropy.

Feature Extraction in LOC

Architectural Description

The lateral occipital complex is modeled as a feature extraction mechanism that converts spatial attention windows into feature vectors. The feature vectors are sparse and high-dimensional, and should capture the local geometric structure and to a lesser extent the color information in attention windows. The goal is to project the contents of attention windows into a high-dimensional feature space such that structurally similar windows will cluster.

Biological Justification

The term *lateral occipital complex* denotes a large cortical region that spatially connects parts of the early vision system to the inferotemporal cortex. Although it has been studied for years, its exact boundaries in people and monkeys remain open to debate, as does the question of whether it is a single functional unit, two units, or possibly more. A general discussion of LOC can be found in Grill-Spector et al [17].

Although the anatomy of LOC is unclear, its significance is not. A subject with bilateral lesions to LOC developed visual form agnosia, a condition which left her unable to recognize even the simplest objects and shapes [18]. By measuring repetition suppression in fMRI, Kourtzi and Kanwisher showed that parts of LOC respond identically to an image of an object or its edge image [19], even if its profile is interrupted [20]. Using a similar technique, Lerner et al [21] showed that LOC responses are able to "fill in" gaps created by projecting bars over images.

These studies provide converging evidence for a view of LOC as computing structural features of attention windows, even in the face of geometrically structured noise. More recently, Kourtzi et al [22] have shown that LOC is involved with learning shape descriptions for later use, and that it becomes even more active if the shapes being learned are partially disguised by complex backgrounds, possibly because it has to work harder. A study by Altmann et al [23] suggests that LOC combines edge information with motion and disparity data and/or top-down predictions.

Confusing this picture somewhat is a study that suggests that at least part of LOC also responds to colors [24], although this may depend partly on the disputed boundaries of LOC. A study by Delorme et al [25] suggests that feature vectors may include both structural and color information, but that the two are kept separate and that some subjects take advantage of color features while others do not. Also, the size

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de

of LOC and the fact that it becomes only diffusely active in fMRI studies of object recognition suggests that the feature vectors are high-dimensional but sparse.

Implementation

We implement LOC as a collection of parametric voting spaces, in the style of a Hough transform. The studies above suggest that LOC aggregates structural information, and behavioral studies by Biederman [26] suggest that collinearity, co-termination, symmetry, anti-symmetry and constant curvature are particularly important structural features. We therefore created parametric representations of collinearity (defined over edges), axes of symmetry and anti-symmetry (defined over edge pairs), and of centers of curvature and termination (also defined over edge pairs). Edges and edge pairs from attention windows vote in these spaces, and the vote tallies form feature vectors. A single color histogram is used as a color feature vector. The final feature space representation is the concatenation of its structural and color feature vectors.

Object Recognition in Inferotemporal Cortex

Architectural Description

The inferotemporal cortex is modeled as unsupervised clustering. It consumes feature vectors and produces *view categories*, which are groups of feature vectors that are similar in structure and color. View categories do not correspond to semantic object labels; semantic object classes may be divided across many view categories. Black cats, for example, do not look like calico cats, and the front view of a cat doesn't look like its side view. View categories are viewpoint and illumination dependent, and semantic object classes may be further divided because of differences among instances (e.g. black cats vs. calico). Also, view categories typically correspond to parts of objects, since attention windows do not presuppose image segmentation.

Biological Justification

The psychological literature makes a distinction between unimodal *recognition* and multi-modal *identification*. As defined by Kosslyn [2], recognition occurs when input matches a perceptual memory, creating a feeling of familiarity. Identification, on the other hand, occurs when input accesses representations in multi-modal memory. Thus we might visually recognize an object as being familiar before we identify it as a cat, at which point we know what it looks like, sounds like, feels like, etc.

Recognition and identification can become disassociated in patients with brain damage. Farah [27] summarizes a collection of patients with associative visual agnosia. These patients cannot recognize objects, even though they can accurately copy drawings and describe the features of an object, suggesting that the early vision system and lateral occipital cortex are intact. These patients also show no deficits in identifying objects by other modalities; their ability to identify objects from language, sound and touch is unimpaired. They therefore demonstrate behaviors that are

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de

consistent with damage to a visual recognition module while the multi-modal identification module remains intact.

The opposite scenario is seen in patients with semantic dementia [27]. These patients retain basic recognition abilities in all of their senses, but loose the ability to form cross-modal associations, for example to associate visual percepts with auditory percepts or abstract concepts. The simultaneous loss of identification abilities across senses is consistent with a damaged identification system but intact sensory recognition modules. There are also cases of selective semantic dementia, in which patients are unable to identify specific classes of objects, for example living things. This is probably the result of damage to part but not all of the identification system, as may be suggestive of how the multi-modal identification system is organized.

Evidence that the inferotemporal cortex learns highly specific view categories comes from several sources. An fMRI study by Haxby et al [28] suggests that IT responds differently to views of standard and inverted faces, while a study by Troje and Kersten goes further [29]: people are expert at recognizing other people's faces head-on or in profile, but are only expert at recognizing themselves head-on, because that is how they see themselves in mirrors. Behavioral studies of face recognition suggest that we are faster and more accurate at recognizing faces illuminated from above than below [30]. Single-cell recordings from the inferotemporal cortices of monkeys suggests different responses to images of faces based on expression [31]. Perhaps most tellingly, Tsunoda et al [32] combined fMRI and single-cell recordings in macaques to probe IT responses to stimulus changes, for example removing part of a target or removing its color. Every significant change resulted in different cellular-level responses in IT. Tanaka et al showed that changes in orientation triggered different cells in macaque IT [33].

The evidence for highly-specific and appearance-based view categories combined with the separation of recognition from identification suggests that IT should be modeled as unsupervised clustering, while associative memories combine collections of category views with training signals to create cross-model object categories. This contradicts some other recent biologically-inspired models (e.g. [34]), which learn to map from stimuli to labels at the level of the lateral occipital complex.

Implementation

We implement IT as a single layer of neurons trained by repetition suppression. Every neuron individually learns to divide feature space in two without dividing any densely populated portions of feature space (i.e. clusters). As a group, neurons produce binary codes that identify view categories. An alternative biologically-inspired unsupervised clustering model of IT has been proposed by Granger et al [35]. We are currently implementing Granger's algorithm in order to compare the two approaches.

Qualitative System Performance

The purpose of this paper is to describe a biomimetic architecture, not to promote a specific system. Nonetheless, a minimal requirement for an architecture is that working systems can be built in it. In the sections above, we described an

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de

implementation for every component. Here we describe how the resulting system performs.

We applied the system to a sequence of 591 images of a toy artillery piece on a turntable; one of the images is shown in Figure 2. The system selected approximately 10 attention windows per image, converted the attention windows to parametric feature vectors and then clustered the resulting feature vectors into view categories. The average image windows for the eight most commonly occurring view categories are shown in Figure 3.

Fig. 2. One of 591 images of a toy artillery piece on a turntable. The average rotation between images is a little less than 1.5°.

In all eight cases, we can easily identify what part of the target or background the view category represents, and in all cases the categories are "pure" in the sense that every feature vector assigned to a category comes from the same target or background location. Different views of the an object part generate different categories; for example, there are two view categories for wheels: one for nearly parallel projections, and another for wheels at more oblique angles (although the latter was not one of eight shown in Figure 3).

Not all of the view categories in Figure 3 are equally meaningful. The first category, in fact, corresponds to the end of the shelf in the background behind the target. This was the most common category, because it never changed viewpoint and was visible in almost all the images. We need the semantic reasoning capabilities of the dorsolateral prefrontal cortex to infer that this category is uninteresting, and top-down control to suppress it from being attended to in the future.

Although view categories correspond to particular points and viewpoints, not all images in which a specific view is visible get included in a category. For example, there are more side-views of wheels than were found and assigned to the 7th category in Figure 3. Often this occurs because the wheel was not attended to; sometimes it was assigned to its own singleton view category. We believe that top-down reasoning will improve the detection rate for most view categories. For example, contexts that imply wheels will generate top-down predictions that increase the frequency with which that view category is found.

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de

Conclusion and Future Work

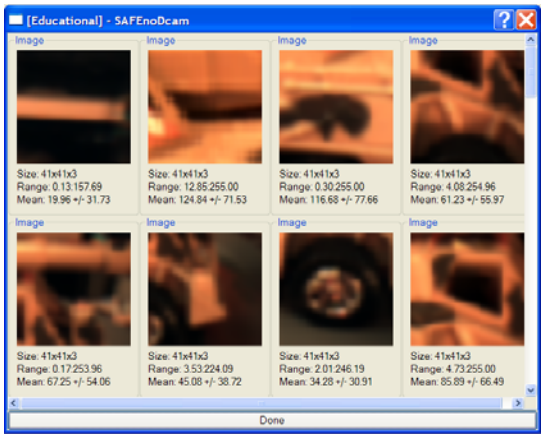

Fig. 3. The 8 most frequently occurring view categories, represented by the average attention window.

We presented a biomimetic architecture that copies the high-level design of human object recognition, and demonstrated a system built in that architecture. We make no claims of optimality for any component; indeed, we believe they all can be improved. Even with the current implementation, however, we were able to apply the system to a sequence of 580 images, and learn meaningful view categories without training data. (The same system has been applied to 2,000 table-top images from a Lego robot and 3,500 images from a floor-level ER1 robot.) Our evaluation is qualitative rather than quantitative because (1) by definition we do not have ground truth labels for view categories, and (2) we know of no other system that categorizes attention window into view categories without supervision features to directly compare it to.

Although we are encouraged by these early results, we would not field the current system as an application in its current form. First we need to close the predictive loop, by implementing biomimetic models of strategic and reflexive top-down processing. We are also interested in adding modules for the dorsal visual stream. Tracking, in particular, provides a significant unsupervised relation between view categories; if a tracked attention window shifts from view category 'A' to category 'B', then those two categories correspond to different views or illuminations of the same object.

References

1. Milner, A.D. and M.A. Goodale, *The Visual Brain in Action*. Oxford Psychology Series. 1995, Oxford: Oxford University Press. 248.
2. Kosslyn, S.M., *Image and Brain: The Resolution of the Imagery Debate*. 1994, Cambridge, MA: MIT Press. 516.
3. Palmer, S.E., *Vision Science: Photons to Phenomenology*. 1999, Cambridge, MA: MIT Press. 810.
4. Draper, B.A., K. Baek, and J. Boody, *Implementing the Expert Object Recognition Pathway*. Machine Vision and Applications, 2004. **16**(1): p. 115-137.
5. Kosslyn, S.M. *Visual Mental Images and Re-Presentations of the World: A Cognitive Neuroscience Approach*. in *Visual and Spatial Reasoning in Design*. 1999. Cambridge, MA: MIT Press.
6. Burton, H., et al., *Adaptive Changes in Early and Late Blind: A fMRI Study of Braille Reading.* Journal of Neuorphysiology, 2001. **87**: p. 589-607.

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de

7. Pessoa, L., S. Kastner, and L.G. Ungerleider, *Neuroimaging Studies of Attention: From Modulation of Sensory Processing to Top-Down Control.* The Journal of Neuroscience, 2003. **23**(10): p. 3990-3998.
8. Grill-Spector, K., et al., *Differential Processing of Objects under Various Viewing Conditions in the Human Lateral Occipital Complex.* Neuron, 1999. **24**: p. 187-203.
9. Lowe, D.G., *Distinctive Image Features from Scale-Invariant Keypoints.* International Journal of Computer Vision, 2004. **60**(2): p. 91-110.
10. Koch, C. and S. Ullman, *Shifts is selective visual attention: Towards the underlying neural circuitry.* Human Neurobiology, 1985. **4**: p. 219-227.
11. Itti, L. and C. Koch, *A Saliency-based Search Mechanisms for Overt and Covert Shifts of Visual Attention.* Vision Research, 2000. **40**(10-12): p. 1489-1506.
12. Draper, B.A. and A. Lionelle, *Evaluation of Selective Attention under Similarity Transformations.* Image Understanding 2005. **100**: p. 152-171.
13. Parkhurst, D., K. Law, and E. Neibur, *Modeling the role of salience in the allocation of overt visual attention.* Vision Research, 2002. **42**(1): p. 107-123.
14. Ouerhani, N., et al., *Empirical Validation of the Saliency-based Model of Visual Attention.* Electronic Letters on Computer Vision and Image Analysis, 2004. **3**(1): p. 13-24.
15. Privitera, C.M. and L.W. Stark, *Algorithms for Defining Visual Regions-of-Interest: Comparison with Eye Fixations.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000. **22**(9): p. 970-982.
16. Kadir, T. and M. Brady, *Scale, Saliency and Image Description.* International Journal of Computer Vision, 2001. **45**(2): p. 83-105.
17. Grill-Spector, K., Z. Kourtzi, and N. Kanwisher, *The lateral occipital complex and its role in object recognition.* Vision Research, 2001. **41**: p. 1409-1422.
18. James, T.W., et al., *Ventral occipital lesions impair object recognition but not object-directed grasping: an fMRI study.* Brain, 2003. **126**: p. 2463-2475.
19. Kourtzi, Z. and N. Kanwisher, *Cortical Regions Involved in Perceiving Object Shape.* The Journal of Neuroscience, 2000. **20**(9): p. 3310-3318.
20. Kourtzi, Z. and N. Kanwisher, *Representation of Perceived Object Shape by the Human Lateral Occipital Complex.* Science, 2001. **293**: p. 1506-1509.
21. Lerner, Y., T. Hendler, and R. Malach, *Object-completion Effects in the Human Lateral Occipital Complex.* Cerebral Cortex, 2002. **12**: p. 163-177.
22. Kourtzi, Z., et al., *Distributed Nueral Plasticity for Shape Learning in the Human Visual Cortex.* PLos Biology, 2005. **3**(7): p. 1317-1327.
23. Altmann, C.F., A. Deubelius, and Z. Kourtzi, *Shape Saliency Modulates Contextual Processin in the Human Lateral Occipital Complex.* Journal of Cognitive Neuroscience, 2004. **16**(5): p. 794-804.
24. Hadjikhani, N., et al., *Retinotopy and color sensitivity in human visual cortical area V8.* Nature Neuroscience, 1998. **1**(3): p. 235-241.
25. Delorme, A., G. Richard, and M. Fabre-Thorpe, *Ultra-Rapid Catgeorization of natural scenes does not rely on colour cues: A study in monkeys and humans* Vision Research, 2000. **40**: p. 2187-220.
26. Biederman, I., *Recognition-by-Components: A Theory of Human Image Understanding.* Psychological Review, 1987. **94**(2): p. 115-147.
27. Farah, M.J., *Visual Agnosia*. 2nd ed. 2004, Cambridge, MA: MIT Press. 192.
28. Haxby, J.V., et al., *The Effect of Face Inversion on Activity in Human Neural Systems for Face and Object Recognition.* Neuron, 1999. **22**: p. 189-199.
29. Troje, N.F. and D. Kersten, *Viewpoint dependent recognition of familiar faces.* Perception, 1999. **28**: p. 483-487.
30. Bruce, V. and A. Young, *In the Eye of the Beholder: The Science of Face Perception*. 1998, New York: Oxford University Press. 280.
31. Sugase, Y., et al., *Global and fine information coded by single neurons in the temporal visual cortex.* Nature, 1999. **400**: p. 869-873.
32. Tsunoda, K., et al., *Complex objects are represented in macaque inferotemporal cortex by the combination of feature columns.* Nature Neuroscience, 2001. **4**(8): p. 832-838.
33. Tanaka, K., *Columns for Complex Visual Objects Features in the Inferotemporal Cortex: Clustering of Cells with Similar but Slightly Different Stimulus Selectivities.* Cerebral Cortex, 2003. **13**: p. 90-99.
34. Serre, T., L. Wolf, and T. Poggio. *Object Recognition with Features Inspired by Visual Cortex.* in *IEEE Conference on Computer Vision and Pattern Recognition*. 2005. San Diego, CA: IEEE CS Press.
35. Granger, R., *Engines of the brain: The computational instruction set of human cognition.* AI Magazine, 2006. **27**(2): p. 15-32.