Human Integration of Shape Primitives*

Gregory Dudek
Center for Intelligent Machines
McGill University
Montreal, Canada H3A 2A7

Martin Arguin
Dept. de Psychologie
Univ. de Montrèal
Montreal, Canada H3A 2B4

Daniel Bub
Montreal Neurological Inst.
McGill University
Montreal, Canada H3A 2A7

Abstract

This paper deals with human shape recognition. In particular, we present results relating the role of human inferior temporal cortex to the description and recognition of shapes using a parametric three-dimensional shape space.

In experiments with a patient with damaged IT cortex, we show that his inability to recognize or distinguish between members of a large family of simple shapes can be traced to an inability to simultaneously combine information related to multiple global shape dimensions. We have discovered that the relevant dimensions describing global object shape, in this case, are related to simple geometric deformations. When object recognition tasks involve multi-dimensional comparisons between potential targets our subjects have difficulty. This deficit is not related to early visual processing, but to an apparently higher-level processing stage.

We suggest that the role of IT cortex in combining shape dimensions may be related to its acknowledged role in the control of visual attention. This has implications for the construction of computer vision systems that recognize large ensembles of objects.

1 Introduction

This may be, in part, because there are multiple mechanisms that contribute to the concept of shape [5, 15]. The very power of the human shape recognition system and the multiple computations apparently involved have made it difficult to decompose functionally [13, 7, 8, are but a few examples]. In this paper we describe an approach to uncovering the mechanisms underlying human shape recognition by looking at the precise deficits that arise in the presence of localized damage to the human visual

*The paper was presented at the 2nd International Workshop on Visual Form, Capri, Italy, May 1994 and appears in the book "Visual Form" published by World Scientific Press, 1994.

system. By observing the shape-specific deficits that occur as a result of localized brain damage, we believe important characteristics of the involved computations can be elucidated. In particular, although the specific deficit that we have examined occurs in only a small number of patients, the fact that similar functional deficits may be associated with different types of brain lesion suggests that the presence of the shape-processing characteristics we have discovered may have far-reaching implications.

Brain-damaged subjects with different types of high-level shape processing deficit are shown to be impaired in the integration of multiple shape primitives that appear to be used in object identification. One of these (who we will focus on in this paper) is a visual agnosic (suffers from an inability to recognize certain objects) while the other is an optic aphasic (suffers from a deficit in labelling visual objects with spared recognition if tested non-verbally). The first patient's impairment (subject ELM) is independent of either subject's ability to perform a wide range or visual tasks (including discrimination between the specified objects) at a perceptual level – that is, under simultaneous presentation. It is only when the object shape must be remembered and recalled that the deficit is apparent, and this deficit appears to be intimately associated with specific shape attributes of the objects involved. This patient's ability to only partially recover object shape information from memory following brain-damage implies that the representation of shape knowledge about objects in neurologically intact individuals itself based on a decomposition into discrete parameters (i.e. it is not an integral representation). Normal object recognition therefore requires these discretely stored shape primitives to be re-associated together via an integration mechanism.

2 Global Parametric Shape

The first patient, ELM, suffered bilateral posterior and inferior temporal lobe lesions. He shows no significant language impairment, no visual field deficit, no hemispatial neglect, no evidence for a perceptual encoding deficit, and he drives a car. Naming and recognition of line drawings of certain objects (for example many classes of biological objects) is severely impaired whereas performance is markedly higher for other groups of objects, for example man-made artefacts (Table 1 a and b). While the recovery of encyclopedic information about biological objects from their names is largely spared, recovery of visual properties is impaired (Table 1c). Stored knowledge of visual properties of objects that cannot be named (eg. biological objects) is not entirely inaccessible or compromised however, since word-to-picture matching errors for fruits and vegetables are shown to be determined by visual similarity (Table 1d).

We attempted to characterize the shape properties that specify the relationships between the objects that ELM fails to correctly identify. The morphologically simplest class of real objects for which ELM exhibits impaired recognition is fruits and vegetables (59 per cent errors in naming, as opposed to near perfect scores for normal

	Stimulus category	
	Biological objects	Artefacts
Controls	6%	7%
ELM	61%	12%

a) Naming

	Stimulus category	
	Biological objects	Artefacts
ELM	41%	7%

b) Object decision (ELM)

	Category of probe question	
	Encyclopedic	Visual
Controls	2%	3%
ELM	15%	45%

c) Knowledge recovery from animal names

Word	Picture					
	Banana	Carrot	Eggplant	Apple	Orange	Pumpkin
Cucumber	80%	60%	40%	0%	0%	0%
Tomato	0%	0%	0%	80%	60%	40%

d) Sample of performance on word-to-picture matching (ELM)

Table 1: Summary of ELM's agnosia

a) Error rates shown by ELM and 10 matched controls in naming line drawings presented for an unlimited duration. These results represent performance on 66 pictures of biological objects (animals, birds, insects, and fruits and vegetables) and 80 pictures of artefacts. b) ELM's error rate in deciding whether pictures were of real or unreal animals (n=70) or artefacts (n=41). c) Error rates shown by ELM and 10 matched controls on two-alternative forced-choice probe questions about animals given their names. Encyclopedic questions (n=60) probed verbal knowledge about animals. Visual probe questions (n=78) concerned visual properties of the items. d) Representative subset of ELM's error rates in matching pictures of fruits and vegetables to names.

observers). These can be described as a single component (in the sense of Hoffman and Richards [6]) that can be obtained through global deformations to an ellipsoid.

Through a careful examination of the identification errors for this and other subjects (using multi-dimensional analysis) we hypothesised three simple shape parameters that are implicated in the performance of ELM. These are:

- changes of elongation (ratio of minor to major axes for a defining ellipsoid),
- tapering along the major axis of the ellipsoid,
- and global curvature (bending) of the ellipsoid perpendicular to the major axis.

These were defined using sequential global transformations of the coordinate plane $(x, y)' = \vec{p'} = \hat{\Omega}(\vec{p})\vec{p} + \hat{\mathrm{T}}(\vec{p})$. where $\vec{p}$ and $\vec{p'}$ are the initial and transformed points

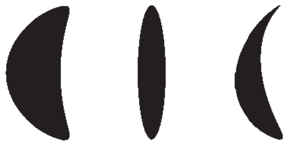
Condition	Examples of test picture(s)	Percept. matching errors	Concept. matching errors
Positive trial		9%	23%
Shares 0/3 primitives with target		0%	8%
Share 1/3 primitives with target		3%	25%
Share 2/3 primitives with target		10%	44%

Figure 1: Sample shapes.
Several of the sample shapes in the three-dimensional shape space. This figure illustrates the test pictures used when the target stimulus (picture or word) was 'banana'.

and $\hat{\Omega}()$ and $\hat{\mathrm{T}}()$ are the rotation ans translation matrices. The transformation are specified as follows (deformations are given below with respect to the y axis, with full generality). For elongation by a factor α we have

$$\hat{\Omega}(\vec{p}) = \begin{bmatrix} 1 & 0 \\ 0 & \alpha \end{bmatrix}, \; \hat{\mathrm{T}}(\vec{p}) = [0, 0]. \tag{1}$$

For tapering by a factor of γ we have:

$$\hat{\Omega}(\vec{p}) = \begin{bmatrix} \gamma/(K-y) & 0 \\ 0 & 1 \end{bmatrix}, \; \hat{\mathrm{T}}(\vec{p}) = [0, 0] \tag{2}$$

where K is a constant larger that the maximum y value of any shape of interest. For bending by a factor β, we have:

$$\hat{\Omega}(\vec{p}) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \; T = [0, \beta|y|] \tag{3}$$

Some sample shapes are shown in Figure 1.

By parametrically combining deformations to a circle using these three independent dimensions, a surprisingly wide variety of candidate shapes can be synthesised.

In particular, combinations of these three shape dimensions can define a wide variety of shapes resembling fruits and vegetables. This has allowed use to examine the shape perception deficit of ELM, as observed in terms of simple biological forms, in the context of parametrically variable synthetic shapes. It is interesting to observe that these deformations are closely related to the "intuitive" parametric deformations typically associated with synthetic solid modelling (for example using Platonic solids or superquadrics) [1, 3, 11, 2, 12] as well as to the Lie groups often associated with motion processing. The stimuli used in the present experiments were filled blobs whose shapes were generated in terms of this three dimensional parameter space. In all experiments, orientation of the items on the picture plane was constantly varied and object length was held constant (we also considered, but rejected, holding other attributes such as object area constant).

3 Perceptual and Conceptual Processing

The experimental approach used was based on the synthesis of artificial shapes from this three dimensional parameter space. The experiments focussed the the ability of a subject to recognize one of these shapes with respect to an ensemble of objects seen previously. By associating these objects with familiar biological forms, a version of the experiment could be carried out that directly related to a pre-existing recognition deficit (as well as being a readily accessible and comprehensible task for naive subjects).

For unique identification of an arbitrary target shape defined within this space, the simultaneous consideration of multiple shape dimensions (in the sense described above) is required. The exposition to this point does not preclude the possibility that the actual measurement dimensions for shape used by humans are quite different from these three and hence the number of dimensions spanned by these target objects may be less than three. As it happens, there is a direct relationship between dimensionality of the ensemble in terms of the specific shape dimensions described above and the performance of our subjects.

Consider a simple shape recognition tasks with shapes labelled with respect to similar-looking fruits. Analysis of the elongation of an object may establish that it is either a banana or a cucumber but, to disambiguate between these alternatives, estimation of the global curvature of the object is also necessary (although perhaps not explicitly). One of our experiments was *word-to-picture matching*: a word and a filled blob were simultaneously presented and the subject had to simply indicate whether the two were consistent (i.e. was the word the correct label for the object). For the set of shapes we used, this task was extremely easy for normal observers.

The recognition impaired subject ELM consistently accepted pictures that only partially matched the shape primitives corresponding to the target words (Fig. 1). In particular, ELM would accept word-picture matches that were correct with respect to *at least one shape dimension* while ignoring the inconsistency in the other dimension(s). This suggested that ELM's visual inability to recognize objects may be determined by a failure to integrate or select specific shape properties, such that his responses are based on only partial shape information.

To precisely examine the distinction between perceptual and conceptual shape recognition as a function of the number of shape dimensions, we developed a more refined experiment, as follows. We compared subject's capacity for a) *perceptually encoding* shapes defined by combinations of shape dimensions (perceptual integration) *without* having to remember them except for a very brief delay, and b) recovering shapes (as a function of multiple properties) from memorized (i.e. stored) representations of objects that were previously viewed (*conceptual integration*). These tasks were based on the subject indicating whether or not a sample shape matched a target shape using either simultaneously presented stimuli or remembered stimuli. The two experiments conducted used the same design.

The target stimuli used were either filled blobs (perceptual matching) that could unambiguously be identified by normal observers as 'banana', 'carrot', 'cucumber', or 'eggplant', or the visually presented names of these items (conceptual matching). The present figure illustrates the test pictures used when the target stimulus (picture or word) was 'banana'.

On each perceptual matching trial, a target picture was presented for 2 sec., followed by a blank 1 sec. interval, and then by a test picture which remained visible until the subject responded. In conceptual matching, the target word and the test picture were presented simultaneously and remained on until the response. In both experiments, ELM indicated whether the test picture matched the target (yes/no).

On negative trials, the number of shape primitives the test-picture shared with either a previously presented target-picture (perceptual matching) or the shape corresponding to a simultaneously presented target-word (conceptual matching) was varied. If a match is performed on partial shape information, error rates on negative trials will increase with the number of shared primitives between the test-picture and the target. This effect occurred only in conceptual matching and ELM's error rates were notably greater in this task than in perceptual matching. This dissociation indicates a selective impairment for the conceptual integration of shape primitives. The results of these trials are summarized in Figure 1.

On negative trials, the test-picture had either zero, one, or two shape primitives in common with the target. If only a subset of the relevant shape primitives (elongation, curvature, and tapering) is considered to perform the match, error rates on negative trials will increase with the number of primitives the test-picture shares with the target. For instance, if only curvature is considered, the error rates on negative trials would be 0%, 33%, and 66% with zero, one, and two shared primitives, respectively. Overall, the error rate was much higher on conceptual than on perceptual matching [$c2(1) = 52.2$; $p < 0.001$]. On negative trials, the effect of number of shape primitives by which the test- picture differed from the target was highly significant in conceptual matching [$c2(2) = 18.5$; $p < 0.001$] but not in perceptual matching [$c2(2) = 5.1$; n.s.].

This integration impairment for shape primitives dissociates from both verbal and semantic processing. Thus, the experiment required no verbal mediation and did not refer to any semantic category. It therefore appears that ELM's impairment involves either

i) the acquisition of shape information from the visual image or

ii) the recovery of shape knowledge from memory.

The learning of arbitrary shape-to-location associations was tested to determine whether ELM's recognition errors reflect a deficit in the integration or combination of a shape measurements. Shape contrasts between the items to be learned were defined over a single dimension (elongation or tapering) or over a conjunction of these dimensions (Fig. 2). The subject had to memorize the positions of four target shapes located at the corners of a rectangular layout. These targets were then removed and the subject had to point to the appropriate reference location for the corresponding target when a sample shape was presented (i.e. the corresponding corner of the now-invisible rectangular layout of targets). An increased error rate on the two-dimensional conjunction cases relative to the single-dimension sets would indicate a separable analysis of shape properties. This was exactly the case for ELM.

Three sets of items differing from one another on a single dimension were constructed. The conjunction sets required the simultaneous consideration of two shape dimensions for unique identification. For example, for any item in the conjunction set there was another that had the same degree of elongation and yet another that had the same degree of tapering. Differences in elongation or tapering between items were *greater* in the conjunction condition than in the single-dimension conditions. Each stimulus set was tested in 10 consecutive blocks. Each block started with a learning trial where all items in a set were displayed simultaneously, each at its assigned location (corners of an imaginary square). Each shape was then shown twice, individually, and in a random order and the task was to point to its assigned location.

The problem in combining individually solvable dimensions in a stimulus suggests that the computational process involved in recognizing these shapes is sensitive specifically to these dimensions. Further, it implies that because increased difficulty arises only when such stimulus dimensions must be extracted *simultaneously*, there must be some degree of separation in the way these individual shape parameters are extracted. We believe that the increased severity of ELM's visual agnosia when the contrast between items was defined over a *conjunction* of multiple shape dimensions indicates a deficit in the *integration* or *selection* of independent shape measurements. Perhaps most striking was the nature of the errors on the multi-dimensional stimuli: errors corresponded to an incorrect estimate of one of the shape parameters involved in the decision, but almost never two parameters simultaneously. In the conjunction condition only 4 per cent of his errors involved confusions between items that differed on *both* elongation and tapering.

Figure 2: Results for shape identification task by pointing.
A) Positions in two dimensions of the shape shape described above of the stimuli used for testing the learning of shape-to-location association. Each axis represents values on a particular shape dimension. Each point represents a specific stimulus used in testing. For any item in the conjunction set, there was another that had the same degree of elongation and yet another that had the same degree of tapering.
B) Error rates shown by ELM and by a brain-damaged (left occipital lesion) non-agnosic control in the learning of arbitrary shape/location associations with single-dimension and conjunction sets. The control subject made fewer errors with the conjunction stimuli than with the single-dimension sets. The opposite result was found in ELM.

A non-agnosic brain-damaged control with a left-hemisphere lesion that includes area V4 showed *more* errors on *single-dimension* sets. Normals also find these single-dimension cases *more difficult* although their error rates are very low. In general, having additional shape characteristics available makes the task easier. In contrast, subject ELM was dramatically impaired on the conjunction set (Fig. 2). (Note that the dimensional parameterization of the shapes was not made explicit to the patient – this is an implicit characteristics of the ensemble of objects.)

In recent experiments, a second patient with with a left-hemisphere lesion that includes area V4 as well as lower visual areas unilaterally who suffers from a deficit related to naming objects has also shown a performance deficit that is directly related to the dimensionality of the task in sense described above. This this case preliminary data indicates that the patient can correctly localize shapes independent of their shape dimensions (unlike ELM) but his reaction time for uttering their names is reduced with the task involves multiple shape dimensions. The apparent significance of these shape dimensions in this rather different task suggest they may play a significant role of human shape processing.

Acquisition of structural information from the image was tested in a third experiment to rule out a deficit in the early vision system. It used the same stimuli as experiment 2 (i.e.blob/learning/pointing experiment) in a paradigm that emphasized a perceptual load as opposed to the memory load emphasized in the previous experiment, i.e. a perceptual matching task. A single shape (target) was presented for 1 sec. and the subject was asked to remember it. After a delay of 1 sec., the four shapes tested in a particular condition (conditions were tested in a single block and were presented in a random order) were shown simultaneously and the subject was asked to point to the target. In this experiment, both the brain -damaged control and ELM showed a better performance with the conjunction set of stimuli than with the single-dimension sets. Thus, with the conjunction set of stimuli, neither of the patients made a single error. In contrast, with the single-dimension sets, the control subject made 6.9% errors and ELM made 11.5% errors. In the case of ELM, the results indicate a cross-over interaction between experiments 2 and 3. In experiment 2, his error rate was markedly higher with conjunction stimuli whereas, in experiment 3, he showed more errors with the single-dimension sets. This, despite the fact that the perceptual discriminations that had to be performed in each experiment were identical. The perceptual matching results of experiment 3 therefore indicate that ELM's impairment with conjunction stimuli in a recognition task (experiment 2) cannot be due to a perceptual (early vision) failure. It is argued that ELM's visual recognition deficit can be attributed to an impairment in maintaining and/or retrieving shape knowledge about objects in memory. Furthermore, given that his impairment in shape-knowledge retrieval is specific to a task involving a conjunction set of stimuli, it can be concluded that this impairment is characterized by partial retrieval of shape information from memory.

4 Discussion and Conclusions

To summarize, an impairment specifically affecting the capacity of a subject to consider multiple shape primitives for object recognition was described. This impairment dissociates from verbal and semantic processes and from the perceptual integration of shape primitives in the image. The recognition deficit can therefore be attributed to a failure in the recovery of multiple properties from stored representations of structural knowledge, a process we called conceptual integration.

The effect of inferior temporal cortical brain-damage on a process of conceptual integration of shape features indicates that this operation contributes to object recognition in neurologically intact individuals. The requirement of conceptual integration in visual recognition has two related implications. First, it implies that visual knowledge about object shapes is decomposed into discrete features rather than as integral, picture-like representations. More fundamentally, it implies that visual recognition cannot merely be defined as a process by which perceptual information is mapped to stored structural representations. The present observations rather demonstrate that visual recognition also involves an assembly operation by which discretely stored primitives are conjoined.

The link between the role of inferior temporal cortex and spatial visual attention has been previously established[10, 4]. Likewise, it is well acknowledged that attention refers to a control of processing that can be applied to domain other than simply the spatial one (for example specific colors or sound pitches). It may be that the impairment in object recognition we have observed as a result of IT damage relates to a form of attentional control in the *feature* or *shape dimension* domain: that is, IT serves to direct a non-spatial aspect of attention to the correct shape attributes to be used in recognition and access to memory. This idea has interesting implications for the design of computational vision systems [14].

Although ELM's agnosic symptoms are category-specific (see above), the underlying impairment extends to the visual classification of arbitrary shapes that have no referential content. This raises the possibility that biological and man-made objects differ in the requirements they impose in terms of shape-knowledge recovery for unique identification.

These results suggest that human object representation involves the analysis of object shape with respect of distinct, separable shape attributes. More surprisingly, it also implies that shape recognition involves the use of an *incomplete subset* of the available shape dimensions. Further, the subset of shape dimensions used for a given recognition task appears to be highly context and task dependent. This may suggest that computational vision systems can also profit from such strategies. In fact, they have tantalizing similarities to existing (but far more restrictive) recognition approaches based on using "precompiled" strategies that are context sensetive.

Acknowledgements

This research was supported by a fellowship from the Medical Research Council of Canada (MRC) to Martin Arguin, a grant from the MRC to Daniel Bub, and a grant from the Natural Sciences and Engineering Research Council of Canada to Gregory Dudek.

References

[1] A. H. Barr. Superquadrics and angle-preserving transformations. *IEEE Comp. Graph-*

ics and Appl., 1:1–120 1–20, 1981.

[2] I. Biederman. Recognition by components. *Psychological Review*, 94:115–147, 1987.

[3] Rodney A. Brooks. Symbolic reasoning among 3-d models and 2-d images. *AI*, 17:285–348, 1981.

[4] Leonardo Chelazzi, Earl K. Miller, John Duncan, and Robert Desimone. A neural basis for visual search in inferior temporal cortex. *Nature*, 363:345–347, 27 May 1993.

[5] M. A. Fischler and R. C. Bolles. Perceptual organization and the curve partitioning problem. In *Proc. of the International Joint Conf. on Artificial Intel.*, pages 1014–1018, Karlsruhe, Germany, Aug. 1983.

[6] D. D. Hoffman and W. A. Richards. Parts of recognition. *Cognition*, 18:65–96, 1984.

[7] J. J. Koenderink and A. J. van Doorn. Dynamic shape. *Biological Cybernetics*, 53:383–396, 1986.

[8] M. Leyton. A process grammar for shape. *Artificial Intelligence*, 34(2):213–247, March 1988.

[9] S. Marshall. Review of shape coding techniques. *Image and Vision Computing*, 7(4):281–294, November 1989.

[10] Jeffrey Moran and Robert Desimone. Selective attention gates visual processing in the extrastriate cortex. *Science*, 229:782–784, August 1985.

[11] A. P. Pentland. On describing complex surface shapes. *Image and Vision Computing*, 3(4):153–162, 1985.

[12] Alex Pentland. Automatic extraction of deformable part models. Vision Sciences TR-104, MIT Media Lab, July 1988.

[13] R. N. Shepard. *J Exp. Psychology*, 3:287, 1966.

[14] John K. Tsotsos. Analysing vision at the complexity level. *Behavioral and Brain Sciences*, 13(3):423–496, 1990.

[15] Elizabeth K. Warrington and Angela M. Taylor. Two categorical stages of object recognition. *Perception*, 7:695–705, 1978.