# Occlusion, Attention and Object Representations

Neill R. Taylor[1], Christo Panchev[2], Matthew Hartley[1],
Stathis Kasderidis[3], and John G. Taylor[1]

[1] King's College London, Department of Mathematics, The Strand,
London, WC2R 2LS, U.K.
{neill.taylor, john.g.taylor}@kcl.ac.uk,
mhartley@mth.kcl.ac.uk
[2] Sunderland University, School of Computing and Technology, St. Peter's Campus,
Sunderland SR6 0DD, U.K.
christo.panchev@sunderland.ac.uk
[3] Foundation for Research & Technology Hellas, Institute of Computer Science,
Vassilika Vouton, 71110 Heraklion, Greece
stathis@ics.forth.gr

**Abstract.** Occlusion is currently at the centre of analysis in machine vision. We present an approach to it that uses attention feedback to an occluded object to obtain its correct recognition. Various simulations are performed using a hierarchical visual attention feedback system, based on contrast gain (which we discuss as to its relation to possible hallucinations that could be caused by feedback). We then discuss implications of our results for object representations per se.

## 1 Introduction

In any complex visual scene there will be many occluded objects. It has proved a challenging problem for machine vision to see how occluded shapes can be segmented and identified in such scenes [1], though has been dealt with by a number of models notably the Neocognitron [2]. Here we consider the use of attention as a technique to help the recognition problem. We do that on the basis of a hierarchical set of neural modules, as is known to occur in the brain. Brain guidance of this form has been much used in the past to deduce pyramidal vision architectures. We add to that hierarchy an additional processing system in the brain, that of attention. This is also well known to act as a filter on inputs, with those stimuli succeeding to get through the filter being singled out for further high-level processing involving prefrontal cortical executive functions.

In this paper we propose to employ both hints from the brain: hierarchy and attention feedback, to help resolve the problem of occlusion. In particular we will show that attention helps to make an object representation of the occluded stimulus more separate from that of the occluder, so helping improve the recognition of the occluded object. At the same time the position of the occluded object can be better specified. We also show that the process of attending to an occluded object does not cause it to become a hallucination, so that it regains all of its components in the internal image of the stimulus in the brain. This is achieved partly by an approach using attention feedback defined as contrast gain on the inputs to a neuron from the attended stimulus.

S. Kollias et al. (Eds.): ICANN 2006, Part I, LNCS 4131, pp. 592 – 601, 2006.

In that way inputs that are zero (due to occlusion) will not become increased at all, whereas special constraints would have to be imposed to prevent other sorts of feedback from eliciting posterior activation of the whole of the encoded occluded stimulus.

The paper starts in the next section with a description of the hierarchical neural architecture. It continues in section 3 with a description of the results obtained on simulating an occluded shape by another different one and the activations in the network arising from attention to the stimulus or elsewhere. In section 4 we use these results to characterize object representations in the brain. Section 5 is a concluding discussion section.

## 2 Visual Architecture

The visual system in the brain is hierarchical in nature. We model that by a sequence of neural network modules forming two hierarchies, one dorsal for spatial representations and one ventral for object representations. Both of these are well known to occur in the brain. We take a simplified model composed of leaky-integrate-and-fire neurons, with the hierarchies as shown in fig. 1, table 1 shows the sizes of all modules.

The model is composed of the ventral 'what' stream and the dorsal 'where' stream. The ventral stream is trained from V2 upwards in a cumulative process, whilst the dorsal stream is hard-wired. The retinal input undergoes edge detection and orientation detection (4 orientations 0º, 45º, 90º, 135º) to provide the input to the ventral stream lateral geniculate nucleus (LGN). Above LGN all regions are composed of a topographically related excitatory layer and an inhibitory layer of neurons that are reciprocally connected, both layers receive excitatory input from other regions; and there are lateral excitatory connections in the excitatory layer.

The ventral 'what' stream via hard-wired and trained connections has a progression of combining object features from oriented bars, to angles formed by 2 bars, to arc segments composed of 3 and 4 bars and finally to objects. At each level there is a loss of spatial information due to the reduction in layer size until at the modelled anterior part of the inferotemporal cortex (area TE) and above representations are spatially invariant. The ventral primary visual cortex (V1 ventral) is composed of 4 excitatory layers, one for each orientation, which are interdigitated in 2-dimensions to provide a pin-wheel structure such that neurons responding to the 4 orientations for the same spatial position are grouped together. V2 is known to preferentially respond to angles formed by pairs of oriented bars [3], so the model V2 is trained using guided spike-time dependent plasticity (STDP) to give similar responses. More specifically, V2 is trained on pairs of bars forming single angles that are present in the objects (square, triangle and circle), and for each single pattern at the input only the neurons assigned to represent that angle are allowed to fire and thereby adapt to the stimulus. V4 receives input from V1 and V2; those to the excitatory neurons are trained using STDP.

Experimental results [4] indicate that V4 responds preferentially to arc segments composed of 3-4 lines; we train V4 on a random selection of length 5 arcs from our object group (square, triangle and circle). TEO (within posterior inferotemporal cortex) receives inputs from V2 and V4; only those connections to excitatory TEO

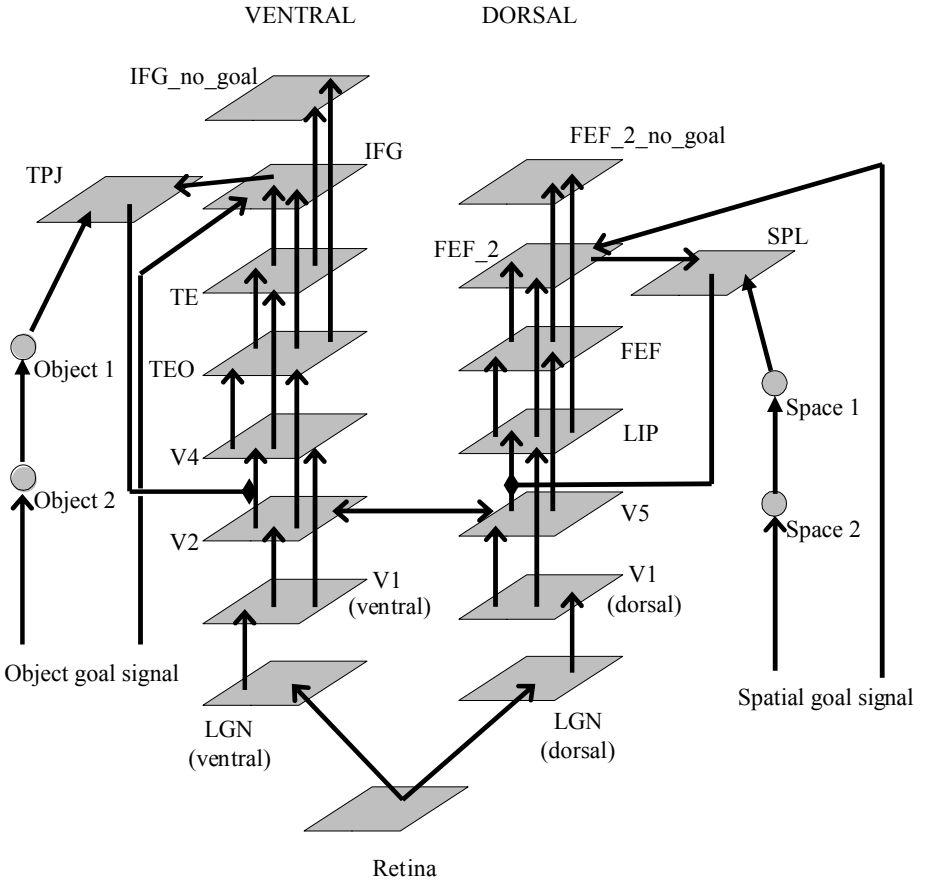



**Fig. 1.** Hierarchical Neural Model Architecture. Open arrow heads indicate excitatory connections, closed-arrow heads are inhibitory connections, and closed-diamond connections indicate sigma-pi weights. We only show the general hierarchical structure, the internal structure of each region is described in the text.

neurons undergo learning. Inputs are complete objects (square, triangle and circle) and use the previously trained V4. TE is also trained on complete objects, with inputs from V4 and TEO. The inferior frontal gyrus (IFG) modules, IFG and IFG_no_goal, are exactly the same and composed of 3 neurons, each one of which represents one of our objects and is hardwired to the TE neurons that preferentially respond to that object. The IFG module can receive inputs from an object goal site such that the system can have its attention directed to a particular object group, where the goal is currently externally determined. This attentional process occurs through sigma-pi weights from temporal–parietal junction (TPJ), the generator of attention movement control signals, onto the V2→V4 (excitatory nodes) connections. Normally when there is no object goal TPJ is kept inhibited by high spontaneous firing of a node termed Object 2; when a goal is setup, this node is itself inhibited by the Object 1 node, which allows TPJ nodes to become active. The sigma-pi weights connect TPJ

object nodes to V2→V4 connections that have been identified as being important for developing representations for that object at the higher levels (TE, IFG modules). Hence when the goal 'square' is set-up, say, we see an increase in the firing rate of the neurons that represent arc segments of squares in V4 (only for neurons that have V2 'square' inputs since this attentional process is modulatory not additive), and via this increased activity to TEO, TE and IFG_no_goal site. The IFG_no_goal module indicates the results of the attentional process; the IFG module does not show the results of attention since it is where the object goal is formed and hence the majority of its activation is goal related.

The dorsal 'where' stream is hard-wired to refine spatial representations as information is passed upwards via V1 dorsal, V5, and frontal eye field (FEF) modules: FEF_1 and FEF_2. Spatial goals can be set-up in the dorsal stream with a particular location being excited in FEF_2. Superior parietal lobule (SPL) is the dorsal equivalent of the ventral TPJ as the generator of the movement of spatial attention signals to posterior sites; SPL receives input from FEF_2 but can only fire if a spatial goal is set allowing for disinhibition via the Space 1 and Space 2 nodes. The spatial goal position is currently determined externally. Sigma-pi weights connect SPL to the V5→lateral intraparietal area (LIP) connections, only the weights between excitatory nodes are affected. A spatial goal can, via the sigma-pi connections, modulate the inputs to LIP nodes from V5; an increased firing rate for nodes in this region results for LIP and higher dorsal modules if an input exists at the same spatial location as the goal. We use the FEF_2_no_goal module to view the affects of spatial attention. We have previously shown, using a similar dorsal route architecture, that attention can highlight a particular spatial location [5].

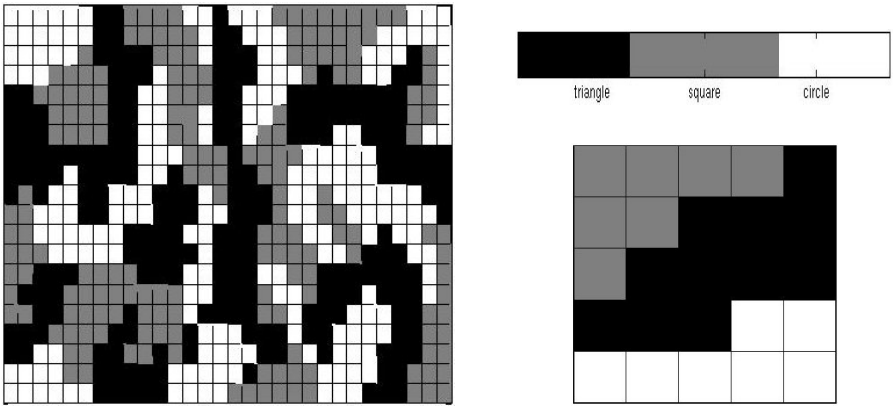
**Table 1.** The sizes of modules

| Module | Size |
|---|---|
| LGN | 38*28 |
| V1 ventral | 76*56 |
| V2 | 76*56 |
| V4 | 30*20 |
| TEO | 15*10 |
| TE | 5*5 |
| IFG, IFG_no_goal, TPJ | 3*1 |
| V1 dorsal, V5. LIP, FEF_1, FEF_2, FEF_2_no_goal, SPL | 19*14 |

Lateral connections between V4 and LIP allow for the passage of information between the 2 streams [6, 7]. When an object is attended to (object goal set) these connections lead to increased firing in the location of the attended object in LIP which then through processing at higher dorsal levels indicates the position of the object. Alternatively spatial attention in the dorsal stream increases firing rates of V4 nodes at that location; via TEO and TE the activations in IFG_no_goal indicate which object is located at that spatial location.

Parameter searches are performed at each level such that useful partial object representations and object representations are found at TEO and TE, respectively.

A more detailed look at the way that the three figures are encoded into the various modules is shown in fig. 2, where the preference maps are shown in a) for V4 and in b) for TE. These maps assign the colours: grey, black and white for square, triangle or circle, respectively, to each of the neurons in the relevant module according to the stimulus shape to which it is most responsive. As we see in fig. 2, in V4 the preference map has a semi-topographic form, with clusters of nodes preferring a given stimulus in their nearby area; in TE there is no topography and the representations are spatially invariant.




a) V4 preference map b) TE preference map

**Fig. 2.** Preference maps in V4 and TE for the shapes: triangle, square and circle

We now turn to consider in the next section what happens when we present an occluded stimulus.

## 3 Occlusion with Attention

The total image we investigate here is shown in fig. 3. It consists of a square and a triangle. Of course it is problematic which of the two figures is occluding the other in fig. 3, since it is two dimensional with no hint of three-dimensionality (to which the resolution of occlusion can contribute). We will discuss that question later, but consider here that we are looking at either figure as occluding the other, and ask the question as to how one might extract a clearer recognition of each figure in the presence of the distorting second figure. This task is thus more general and difficult, we suspect, than occlusion, where one figure is to be taken as the occluding figure in the foreground and the other (the occluded figure) as background or ground (as detected by the three-dimensional clues).

The activities in several of the modules are shown in figures 4, 5 and 6. Of these some entries have negative values, these being where the change between two different conditions has been calculated. In particular the figure 4a, denoted 'V4 Square-Away' shows the firing rates for V4 neurons where the results for attend triangle have had the attend-away results subtracted from them to show the overall difference that attending the square causes. A similar result holds for fig. 4b.

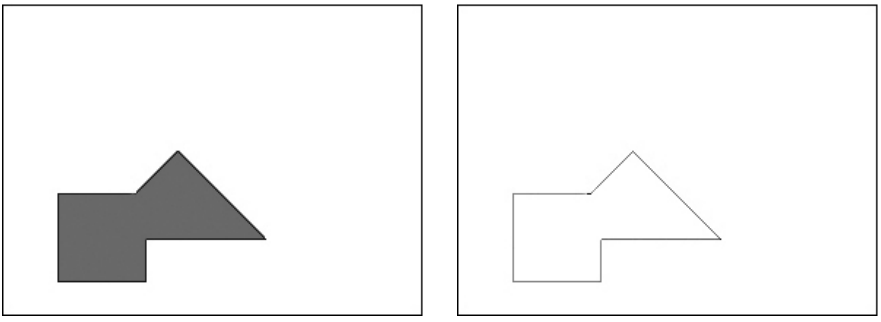
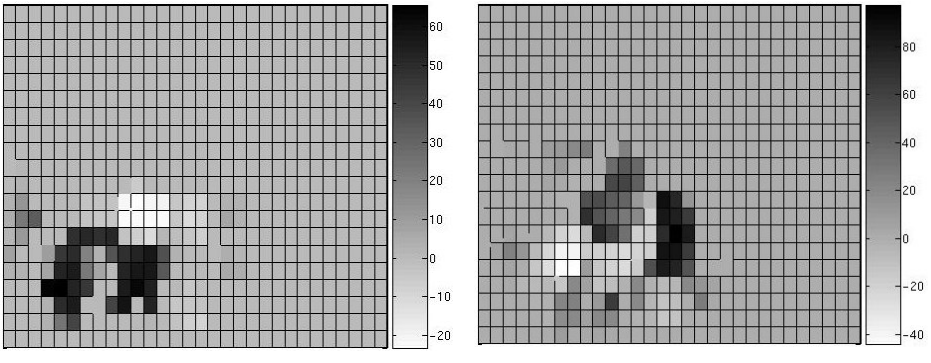


**Fig. 3.** Occluded input, composed of a square and a triangle. Left figure shows the complete input, right the edge detected input.

In particular we see from fig. 4a that when attention is turned to the square, in comparison with the case when no attention feedback is used at all, then there is considerable increase in the firing rates of V4 neurons preferring the square, as seen from the V4 preference map of fig. 3a. Symmetrically, the triangle-representing neurons increase their firing considerably when attention is directed towards the triangle, to the detriment to those neurons preferring the square.

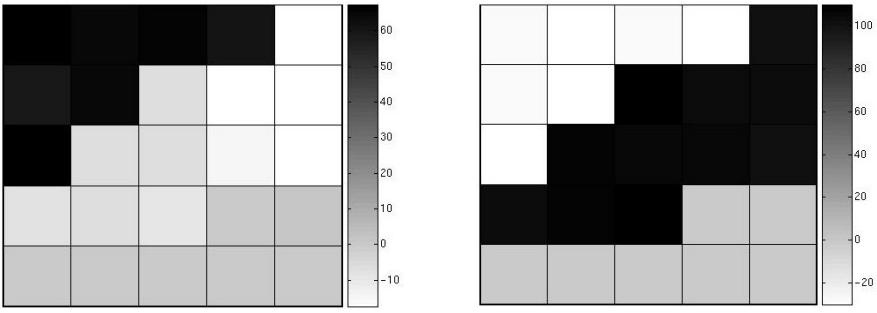



a) V4 square-away b) V4 triangle-square

**Fig. 4.** V4 responses. Scale in Hz. We plot a) V4 response to attend square of composite object minus response to composite image without attention.; b) is the response when the triangle of the composite object is attended to minus the attend away firing rate.

A similar situation occurs for the neurons in TE as seen in fig. 5. In fig. 5a it is exactly those neurons preferring the square that are boosted, with an average firing rate increase of about 64 Hz. In the same manner, the attention feedback to the triangle-causes the triangle-preferring nodes in TE to increase their firing rates by above 100Hz on average.

These results indicate that either figure becomes more strongly represented in V4 and TE when attention is turned to it. This is in general to the detriment of the other figure, as is to be expected.




a) TE Square-Away b) TE Triangle Away

**Fig. 5.** TE responses. Scale in Hz. Where a) is the TE firing rate response when the square is attended to minus the response in the no attention case; b) is the TE response to attend triangle minus the attend away response.

The IFG_no_goal module shows the affects of object attention. When there is no attention the square representation is the most highly activated having a stable firing rate of 49Hz, whilst the triangle representation has a firing rate of only 9Hz. With attention directed to the square, the firing rate for the square increases to 118Hz, and the triangle reduces to 1Hz, but when attention is focused on the triangle the firing rates become 12Hz for the square representing node and 123Hz for the triangle. These results agree with the previous ones described for activities residing in lower-level modules under the affect of attention to one or other of the shapes - square or triangle (as shown in figures 4 and 5).

## 4 Object Representation

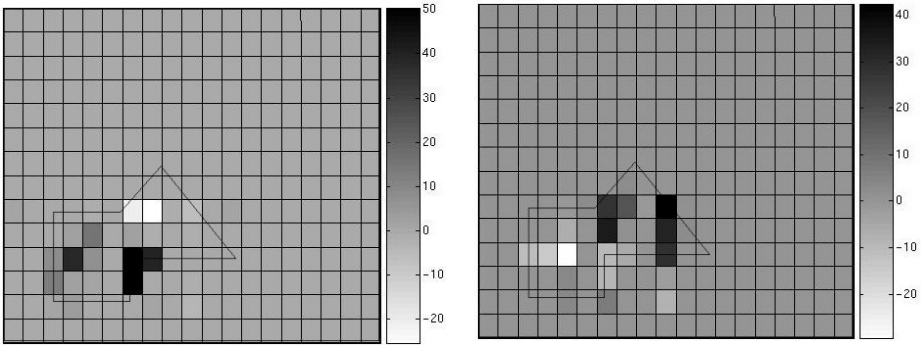
As previously mentioned the lateral connections between the ventral and dorsal streams allow for activations modulated by attention to be transferred from stream to stream. We see from fig. 6 the change in FEF responses when attention is directed to the square or triangle (within the ventral stream) versus the no attention condition. In both cases firing rates increase by up to 40-50Hz, highlighting the centres of the objects as well as parts of the overlap region, there are also decreases near the centre of the non-attended object in range of 20-30Hz. The highest firing rates occur near the overlap regions, but overall the firing rates are higher for the spatial location of the attended object.

Such results indicate that a neural 'object representation' in our architecture is dependent on where attention is directed. If it is directed at the stimulus providing activation of the learnt object representation in the brain then there are two aspects of the representation which are different from the set of neurons activated when there is no attention to the object or attention is directed to another object in the visual field. These effects are

1) An extension of the representation into the other stream (from ventral to dorsal stream activations if objects are being attended to, or from spatial position to object, so from dorsal to ventral streams, if positions are being attended to. This extension is

not as strong if there is no attention to one or other of space or object, with attention in the ventral stream there are increases in LIP firing rates at the location of the attended object of up to 100Hz, in FEF_no_goal module the increases are up to 50Hz in attention cases as against the no attention cases.

2) A lateral extension of the representation in each module when attention is involved. Such a lateral extension, for example in V4, is due to the lateral excitatory connections in the module itself, which allow for the spread of activity due to the increased firing of central neurons to the object representation when attention is being used. Whilst there is little change in the numbers and positions of active V4 excitatory neurons, an increase of ~8% in the number of neurons with most of these having low firing rates (<10Hz), there is a large increase in the firing rates of neurons that are active and showing a preference to the attended object (up to 80Hz), and a decrease in firing rates of active nodes showing a preference for the unattended object caused by increased inhibition. The spreading of activity at the excitatory layer is prevented by an increase in numbers and firing rates of the V4 inhibitory neurons, from 3 neurons in the non-attentive case to 16 for the attentive case and firing rates increasing by up to 20Hz.




a) FEF Square-Away b) FEF Triangle-Away

**Fig. 6.** FEF responses, with the perimeter of the input scaled and superimposed on the neuron firing rate changes. Scale in Hz. The plots are a) FEF response to attend square minus response to attend away; and b) is the response to attend triangle minus the attend away result.

We conclude that an object representation can only be specified when the attention state of the system is itself known. This corresponds to the statement that when feedback is present then not only do the afferent synaptic weights to the neurons of a module specify the unattended representations it carries, but the feedback weights are also needed to specify how these representations are modified by the attention state.

It can be commented that this feature of attention-dependent object representations is well known. However we are here indicating, beyond the feature itself of attention-dependence, that this feature is important to help resolve the question of occlusion, and to separate out components of an occluded/occluding set of objects which the objects are. Such a process needs, as a base to start from, the object representations of the unattended figures. The resolution of the occluded figures into its component

objects can then be achieved by using the slightly activated classifier neurons in IFG (without attention) so as to return to their component attended object representations sequentially to check that these objects are indeed present. Without attention the firing rates in the IFG_no_goal module are: triangle 4Hz, square 64Hz, circle 0Hz for our occluded input (fig. 2).

## 5 Discussion

We have shown using a hierarchical visual model how attention, modelled as contrast gain, can be used to recognise occluded objects. By activating object goals in the ventral visual stream the components of an occluded object can be recognised, in this case a square and a triangle. Additionally the increased activation due to this attentional process can be transferred into the dorsal stream to cause increased activations in the spatial location of the attended object. Indeed, though not shown here, results have shown that with lateral connectivity at the V4 – LIP level spatial attention can lead in the ventral to the identification of the object at the attended location. Contrast gain attention, as modelled here, does not cause hallucination of the occluded part of an object, since the occluded parts have zero activation and cannot be increased by the attentive modulation. Additive attention could lead to the occluded parts of the object becoming active, as the feedback from higher levels travels down the visual stream. We are not necessarily talking here about 'filling-in' since it is clear that looking at figure 2 we can attend to the part of the object that resembles a square without 'seeing' the occluded vertex. We have yet to include in a comprehensive manner the known feedback connections in the visual cortex. Limited additions of these weights have shown that small weight values help attention as contrast gain to distinguish between attended and unattended objects, as well as helping to refine representations at higher levels in the attend away cases.

Certain attention results [8] have so far only successfully been modelled using attention as multiplicative for graded and spiking neurons in a variety of architectures [8, 9, 10, 11]. Visual models using additive attention include [12] and recent models [13] have suggested that a multiplicative component of attention can be the result of additive attention, though they did not investigate whether the model gave similar results to the experimental studies [8].

There is the question of what is occlusion in a 2-dimensional image in a monocular system. Here, we have both objects the same colour and perhaps this is not occlusion but a composite object that attention can be moved around via goals to find which parts resemble the system's learnt objects (square, triangle). This could be a harder problem than dealing with occlusion in 3-dimensions with a binocular system and the aid of extra information such as depth, and our results on the 2-dimensional problem give good grounds for claiming that attention will be crucial in resolving 3-dimensional occlusion.

Finally, attention-dependent object representations are important in resolving occlusion, by the separation of the composite object into component objects, using the unattended object representations.

## Acknowledgments

One of us (NRT) would like to thank EPSRC, others (JGT, CP, & SK) would like to thank the EC, under the GNOSYS (FP6-003835) project, and another (MH) would like to thank the EC, under the MATHESIS project for support to carry out this work.

## References

1. Zitnick C. L. & Kanade T.: 'A Cooperative Algorithm for Stereo Matching and Occlusion Detection', IEEE Trans. Patt. Anal. Mach. Intel. (2000) 22: 675-684
2. Fukushima K.: "Recognition of partly occluded patterns: a neural network model", Biol Cybern (2001) 84: 251-259.
3. Ito M. & Komatsu H.: Representation of Angles Embedded within Contour Stimuli in Area V2 of Macaque Monkeys. J Neuroscience (2004) 24: 3313-3324.
4. Pasupathy A. & Connor C.E.: Shape Representation in Area V4: Position-Specific Tuning for Boundary Configuration. J Neurophysiol (2001) 86:2505-2519.
5. Taylor J.G., Hartley M., & Taylor N.R. 'Attention as Sigma-Pi controlled ACh-based feedback', Proc. of IJCNN'05 (2005).
6. Lanyon L. J. & Denham S.L.: 'A model of active visual search with object-based attention guiding scan paths', Neural Netw. (2004) 17: 873-97.
7. van der Velde F. & de Kamps M.: 'From Knowing What to Knowing Where: Modeling Object-Based Attention with Feedback Disinhibition of Activation', J. Cog. Neurosci. (2001) 13: 479-491.
8. Reynolds J.H., Chelazzi L. & Desimone R.: 'Competitive mechanisms subserve attention in Macaque areas V2 and V4', J. Neurosci. (1999) 19: 1736-53.
9. Taylor J.G. & Rogers M.: 'A control model of the movement of attention' Neural Netw. (2002) 15:309-326.
10. Taylor N.R., Hartley M. & Taylor J.G. 'The Micro-Structure of Attention', accepted for CNS'06 (2006).
11. Taylor N.R., Hartley M. & Taylor J.G. 'Analysing Attention at Neuron level' accepted for BICS'06 (2006).
12. Grossberg S. & Raizada R. D. 'Contrast-sensitive perceptual grouping and object-based attention in the laminar circuits of primary visual cortex', Vision Res. (2000) 40: 1413-32.
13. Deco G. & Rolls E.T.: 'Neurodynamics of biased competition and cooperation for attention: a model with spiking neurons', J. Neurophysiol. (2005) 94: 295-313.