SpringerPlus
a SpringerOpen Journal

RESEARCH Open Access

Application of the SP theory of intelligence to the understanding of natural vision and the development of computer vision

J Gerard Wolff

Abstract

The *SP theory of intelligence* aims to simplify and integrate concepts in computing and cognition, with information compression as a unifying theme. This article is about how the SP theory may, with advantage, be applied to the understanding of natural vision and the development of computer vision. Potential benefits include an overall simplification of concepts in a universal framework for knowledge and seamless integration of vision with other sensory modalities and other aspects of intelligence. Low level perceptual features such as edges or corners may be identified by the extraction of redundancy in uniform areas in the manner of the run-length encoding technique for information compression. The concept of *multiple alignment* in the SP theory may be applied to the recognition of objects, and to scene analysis, with a hierarchy of parts and sub-parts, at multiple levels of abstraction, and with family-resemblance or polythetic categories. The theory has potential for the unsupervised learning of visual objects and classes of objects, and suggests how coherent concepts may be derived from fragments. As in natural vision, both recognition and learning in the SP system are robust in the face of errors of omission, commission and substitution. The theory suggests how, via vision, we may piece together a knowledge of the three-dimensional structure of objects and of our environment, it provides an account of how we may see things that are not objectively present in an image, how we may recognise something despite variations in the size of its retinal image, and how raster graphics and vector graphics may be unified. And it has things to say about the phenomena of lightness constancy and colour constancy, the role of context in recognition, ambiguities in visual perception, and the integration of vision with other senses and other aspects of intelligence.

Keywords: Vision; Information compression; Artificial intelligence; Perception; Cognition; Representation of knowledge; Learning; Pattern recognition; Natural language processing; Reasoning; Planning; Problem solving

1 Introduction

The *SP theory of intelligence*, introduced below, aims to simplify and integrate ideas across artificial intelligence, mainstream computing, and human perception and cognition, with information compression as a unifying theme. This article is about how the SP theory may, with advantage, be applied to the understanding of natural vision and the development of computer vision, and to discuss associated issues.

In these areas, potential benefits of the SP theory and its broad perspective include:

- Developing concepts that combine *simplicity* with descriptive or explanatory *power*, in accordance with Occam's Razor (Wolff 2014a, Sections 2 and 4).
- Deeper insights and better solutions to problems (Wolff 2014a, Section 6).
- With natural vision, developing concepts that are consistent with evidence that general principles apply across different parts of the brain and across different aspects of brain function (Wolff 2014b, Section III-A.7).
- The seamless integration of artificial vision with other sensory modalities, and with other aspects of intelligence such unsupervised learning, pattern recognition, reasoning, planning, problem solving,

Correspondence: jgw@cognitionresearch.org
CognitionResearch.org, Menai Bridge, UK

and more (Wolff 2014a, Section 7)—with consequent benefits in terms of the versatility and adaptability of artificial systems.
- Contributing to the development of a *universal framework for the representation and processing of diverse kinds of knowledge* (UFK) and thus helping to overcome the problem of variety in big data (Wolff 2014b, Section III).

The central idea is that, in accordance long-established principles in science, we should aim for theories with broad scope and avoid micro-theories that only work in restricted areas. If one's view is too narrow, it may be difficult to grasp the big picture—like the blind men trying to understand an elephant.

The next section describes the SP theory in outline, and subsequent sections discuss how it may be applied to the understanding of natural vision and the development of artificial vision. These two aspects of vision are discussed together throughout the article, since each one may illuminate the other.

2 Outline of the SP theory

The SP theory is described fairly fully in an extended overview (Wolff 2013), with enough detail to ensure that the rest of this article makes sense. The most comprehensive description of the theory and its applications is in the book, *Unifying Computing and Cognition* (Wolff 2006a). Applications of the theory are described in Wolff (2006b, 2007, 2014a,b) and other articles, details of which may be found via http://www.cognitionresearch.org/sp.htm.

In broad terms, the SP theory has three main elements:

- All kinds of knowledge are represented with *patterns*: arrays of atomic symbols in one or two dimensions.
- At the heart of the system is compression of information via the matching and unification (merging) of patterns, and the building of *multiple alignments* like the one shown in Figure 1.
- The system learns by compressing *New* patterns to create *Old* patterns like those shown in columns 1 to 11 in the figure.

Because of the intimate connection between information compression and concepts of prediction and probability (Li and Vitányi 2009), the SP system is intrinsically probabilistic. Each SP pattern has an associated frequency of occurrence, and probabilities may be calculated for each multiple alignment and for associated inferences (Wolff 2013, Section 4.4; Wolff 2006a, Section 3.7).

The system is realised in the form of a computer model which provides all the examples of multiple alignment in this article. At present, the model works only with one-dimensional patterns (Wolff 2013, Section 3.3) but it is envisaged that, at some stage, it will be generalised to work with patterns in two dimensions.

It is intended the SP computer model will be the basis for the development of a high-parallel *SP machine*, an expression of the SP theory, a vehicle for research, and a means for the theory to be applied (Wolff 2013, Section 3.2).

Although the main emphasis in the SP programme has been on the development of abstract concepts as outlined above, several of those concepts may be realised in terms of neurons and their inter-connections. This aspect of the SP theory—called *SP-neural*—is described in Wolff (2013, Section 14) and Wolff (2006a, Chapter 11).

The theory has things to say about several aspects of computing and cognition, including unsupervised learning, concepts of computing, aspects of mathematics and logic, the representation of knowledge, natural language processing, pattern recognition, several kinds of reasoning, information storage and retrieval, planning and problem solving, and aspects of neuroscience and of human perception and cognition.

3 Low-level perceptual features

It is now widely accepted that, at 'low' levels in vertebrate and invertebrate visual systems, there are processes that recognise perceptual features such as edges and corners. Some relevant evidence is outlined in subsections below.

In this section, the main focus is on features that may be regarded as 'explicit' because they derive directly from visual input. But it is well known that we may 'see' things that have little or no counterpart in the visual input, such as the 'subjective contours' in Marr (2010, Figure 2-6) or the edge of one leaf where it overlaps another in Marr (2010, Figure 4-1(a)). These kinds of 'implicit' features will be considered in Section 7.1.

In two respects, explicit perceptual features sit comfortably with the SP theory:

- They may be seen to provide a means of encoding perceptual information in an economical manner. For example, Attneave (1954) writes that "Common objects may be represented with great economy, and fairly striking fidelity, by copying the points at which their contours change direction maximally, and then connecting these points appropriately with a straight edge" (p. 185). He illustrates this with the now-famous picture of a sleeping cat, reproduced in Figure 2.
- At lowish levels, perceptual features may function as if they were the atomic symbols that provide the foundation for all higher-level structures, even though they themselves have been constructed from lower-level components.

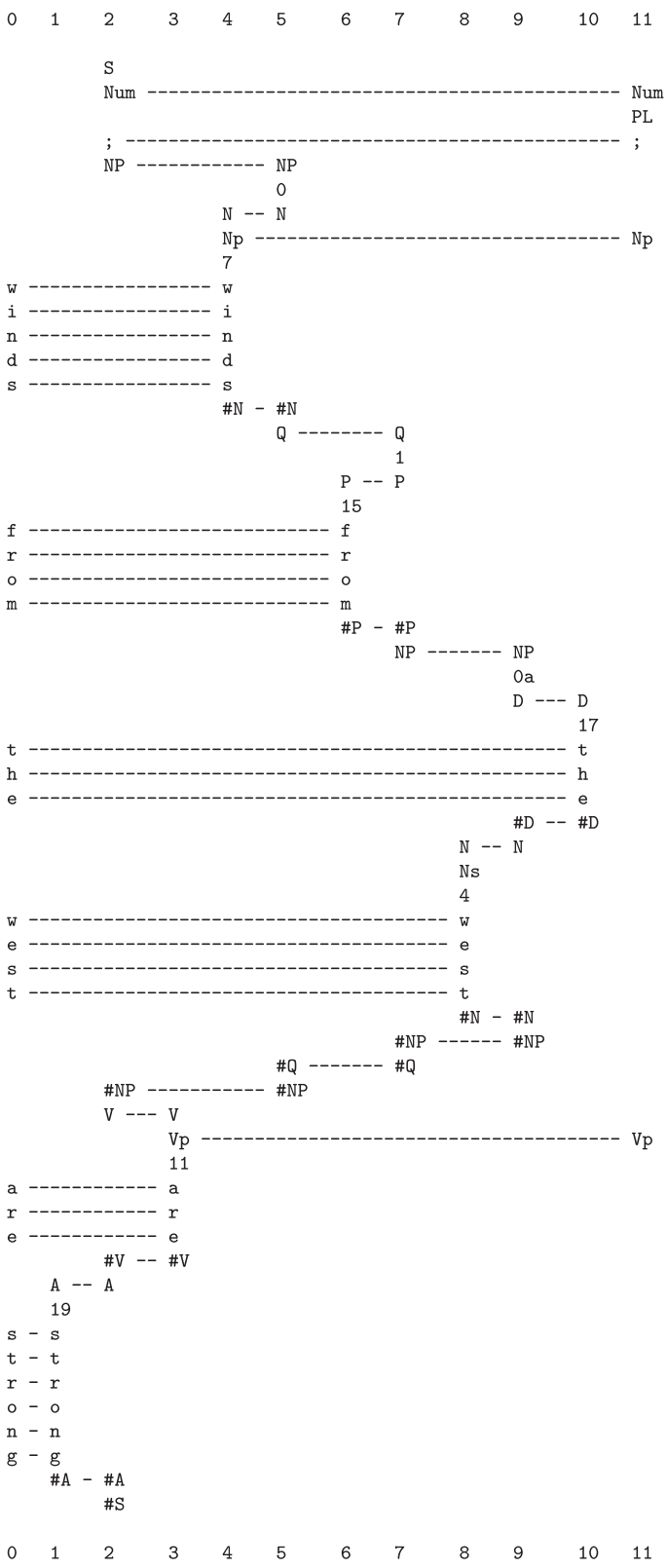

Figure 1 The best multiple alignment created by the SP computer model (the multiple alignment that achieves the highest level of compression) with a store of Old patterns like those in columns 1 to 11 (representing grammatical structures, including words) and a New pattern (representing a sentence to be parsed), shown in column 0.

Figure 2 Drawing made by abstracting 38 points of maximum curvature from the contours of a sleeping cat, and connecting these points appropriately with a straight edge. Reproduced from Figure 3 in Attneave (1954), with permission.

As just indicated, vision begins with images, not perceptual features. The latter must be somehow discovered or detected within the images. The following subsections consider how the SP theory may be applied in this area, starting with a consideration of options for the encoding of light intensities.

3.1 The encoding of light intensities

In the design of artificial systems for vision, it seems natural and obvious that light intensities in images should be expressed as numbers. But, in itself, the SP system recognises only atomic symbols, each one of which can be matched in an all-or-nothing manner with another atomic symbol. It is true that, in principle, it may be supplied with patterns that express Peano's axioms or similar information, and it may then interpret numbers correctly (Wolff 2006a, Chapter 10). But this has not yet been explored in any depth and, in any case, numbers are probably a distraction in understanding how SP principles may be applied to vision.

To simplify the discussion here, we shall assume that we are processing monochrome images with just two categories of pixel: black and white. With that kind of representation, the lightness in any given small area may be encoded via the *densities* of black and white pixels in that area, without using explicit numbers, somewhat like the encoding of dark and light in monochrome newspaper photographs, at least as they used to be (see also (Wolff 2006a, Section 2.2.3)). It is true that such pixels may be represented with the symbols '1' and '0' but these are simply atomic symbols (as required by the SP system), without numerical meanings.

3.2 Edge detection with neurons

It is relevant to this discussion to consider briefly how edges may be detected with neurons. Figure 3 shows two sets of recordings from a single visual receptor ('ommatidium') of the horseshoe crab, *Limulus polyphemus*. In both sets of recordings, the eye of the crab was illuminated in a rectangular area bordered by a dark rectangle of the same size (producing a step function as shown at the top right of the figure). In both cases, successive recordings were taken with the pair of rectangles in successive positions across the eye along a line which is at right angles to the boundary between light and dark areas. This achieves the same effect as—but is easier to implement than—keeping the two rectangles in one position and taking recordings from a range of receptors across the light and dark areas.

In the top set of recordings (triangles) all the ommatidia except the one from which recordings were being taken were masked from receiving any light. In this case, the target receptor responds with frequent impulses when the light is bright and at a sharply lower rate in the dark. In the bottom set of recordings (circles) the mask was removed so that all the ommatidia were exposed to the pattern of light and dark rectangles. In this case, positive and negative responses are exaggerated near the border between light and dark areas but the target receptor fires at or near a background rate in areas which are evenly illuminated (either light or dark). This kind of effect—which is seen elsewhere in the animal kingdom—appears to be due to lateral inhibition between neurons in the visual system (von Békésy 1967, pp 172–174).

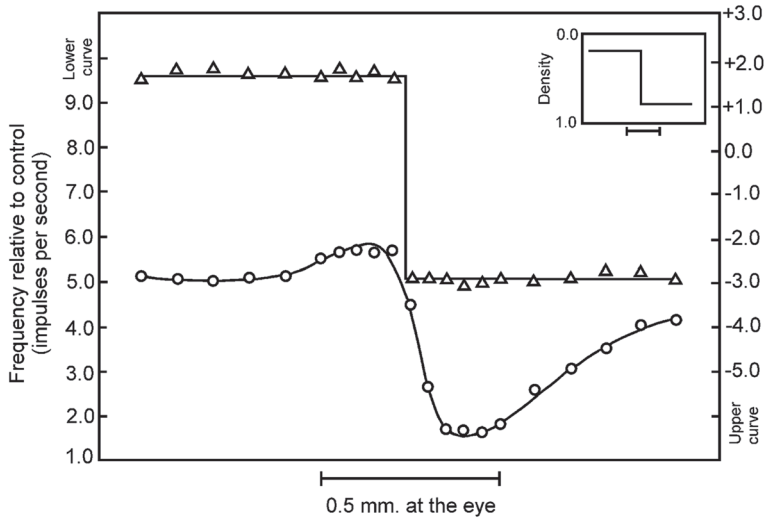

Figure 3 Two sets of recordings from a single ommatidium of *Limulus* (Ratliff and Hartline 1959, p. 1248). Reproduced from Figure 4, *The Journal of General Physiology, 42*, p. 1248, by copyright permission of The Rockefeller University Press.

It has been recognised for some time that the dampening of the response in regions of uniform illumination (light or dark) may be seen to achieve the effect of compressing visual information by extracting redundancy from it (Barlow 1959). It is somewhat like the 'run-length coding' technique for compression of information: a symbol or group of symbols that repeats in a contiguous sequence may be reduced to a single instance, perhaps marked for repetition[a].

A boundary between one uniform area and another may be represented economically by two such compressed representations, side-by-side. In the neural case, the upswing near the light/dark boundary may be seen as an economical representation of the idea that the whole of the preceding area is light, the downswing on the other side may be seen as a succinct marking of the fact that the following area is dark, while the two together may be seen to serve as a compressed representation of the boundary.

Although it is less directly relevant to the present discussion, it is pertinent to mention that there are 'complex' cells in mammalian visual systems that respond selectively to edges, and also to 'lines' and 'slits' (see, for example, Frisby and Stone (2010), pp 215–219).

3.3 Edge detection with the SP system

In the SP framework, the effect of run-length coding may be achieved via recursion, as illustrated in Figure 4[b].

Here, each instance of '`a b c`' in the New pattern in row 0 is matched to an appearance[c] of the self-referential Old pattern '`X 1 a b c X #X #X`' in each of rows 1 to 4. It is self-referential because '`X #X`' in the body of the pattern may be matched and unified with '`X ... #X`' at the start and end of the pattern.

The encoding of the New pattern that we may derive from this multiple alignment is the relatively short sequence '`X 1 1 1 1 #X`'[d]. It achieves lossless compression of the original sequence by recording that the run contains 4 instances of the pattern '`a b c`', in the manner of unary arithmetic. With lossy compression, the encoding may be reduced to '`X #X`', which simply records a sequence of instances of '`a b c`' without specifying the length of the sequence.

As before, two such encodings, side-by-side, would be an economical representation of the boundary between one uniform region and another.

Of course, this does not look much like lateral inhibition with neurons, as outlined in Section 3.2. But at an abstract level, the two things may be seen to produce the same result: the extraction of redundancy from uniform regions, leaving information about the boundaries between such regions as an economical representation of the raw data, like David Marr's (2010) 'primal sketch'.

With other developments—such as the generalisation of the SP concepts to two dimensions (Section 2)—this kind of technique may be applied in computer vision.

3.4 Orientations, lengths, and corners

So far, we have said nothing about the orientations of edges or their lengths. In principle, those things may be encoded mathematically, and very economically, in the manner of vector graphics. But that does not seem very likely in a biological system and it is not necessarily the

```
0     a b c     a b c     a b c     a b c                   0
      | | |     | | |     | | |     | | |
1     | | | X 1 a b c X   | | |     | | |         #X #X    1
      | | | |             |   |     | | |         |  |
2 X 1 a b c X             |   | | |     | | |         |  #X #X 2
                          |   | | |     | | |         |
3                         X 1 a b c X   | | |       #X #X       3
                                        |   | | |     |
4                                       X 1 a b c X #X #X         4
```

Figure 4 A multiple alignment produced by the SP model with the New pattern 'a b c a b c a b c a b c' in row 0 and an appearance of the Old pattern, 'X 1 a b c X #X #X' in each of rows 1 to 4.

best option for any artificial system that aspires to human-like capabilities in vision (Section 7.2.2).

As mentioned above, the visual cortex in mammals is populated by large numbers of 'complex' neurons, each one of which responds to an 'edge', 'slit', or 'line', at a particular orientation. There is a good coverage of different angles within each small area (see, for example, Frisby and Stone (2010), Chapter 9). These observations suggests that, in natural vision, the orientation of any edge may be encoded quite simply and directly in terms of the corresponding type of neuron, and likewise in an artificial system.

A sequence of such codes would describe both the orientation and length of a straight line but it would contain the same kind of redundancy as is discussed in Section 3.3 because the orientation is repeated in successive parts of the line. So we may guess that, in natural vision, some kind of run-length coding may operate, reducing the redundancy within the body of the line and preserving information where the repetition stops—at the points where the line begins and where it ends (see also (Wolff 2006a, Section 13.2.1.4)).

This kind of technique may be applied not only to straight lines but also to lines with a uniform curviture. Any such line may be encoded as repeated instances of a short segment that expresses the curvature of the whole line.

With regard to straight lines, some relevant evidence comes from studies showing the existence of 'end stopped' hypercomplex cells that respond selectively to a bar of a defined length, or a corner (see, for example, Frisby and Stone (2010), pp 216–217). In keeping with Attneave's (1954) remarks quoted earlier, we may guess that, in mammalian vision, the orientation and length of an edge, slit or line, is to a large extent encoded via neurons that record the beginning and end of the line and any associated corners. Orientation-sensitive neurons would provide the input for this 'higher' level of encoding.

In artificial systems, this kind of coding may in principle be done within the multiple alignment framework, as outlined in Section 3.3.

3.5 Noisy data and low-level features

Readers may, with some justice, object that real visual data is rarely as clean as the example in Figure 4 may suggest. With monochrome images, it is likely that most areas will be some shade of grey, not purely black or purely white, and there are likely to be blots and smudges of various kinds.

What appears to be a promising answer to this kind of problem is that the SP system is designed to search for optimal solutions and is not unduly disturbed by errors of omission, commission and substitution. There is more on this topic in Sections 4.1 and 5.6.

4 Object recognition and scene analysis

In some respects, object recognition is like parsing in natural language processing (see, for example, Farabet et al. (2013) and Han (2005)). Since the SP system works well in parsing, as outlined in Wolff (2013, Section 4), it may also prove useful in computer vision. Naturally, it would be necessary for the SP machine to have been generalised to work with patterns in two dimensions (Section 2). And in this discussion we shall assume that low-level perceptual features have been identified, and that they may be treated as atomic symbols, in accordance with the SP theory (Section 3).

Figure 5 shows schematically how someone's face, with their ears, may be parsed within the multiple alignment framework. Row 0 in the figure contains a New pattern representing incoming information. Each part has been aligned with an Old pattern representing stored knowledge of the structure of an ear, an eye, etc. And these are aligned with a pattern in row 2 representing the higher-level structure of someone's head.

Although this is schematic, I believe the approach has potential, as described in the following subsections.

4.1 Noisy data in parsing and recognition

Contrary to the impression one might gain from Figure 5, the SP system is quite robust in the face of errors in such tasks as parsing natural language or pattern recognition. This is illustrated in the multiple alignment in Figure 6

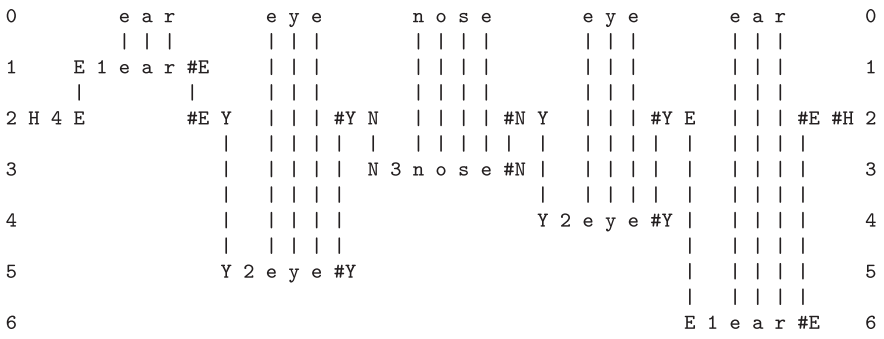

Figure 5 A multiple alignment showing schematically how a person's face, with their ears, may be recognised.

where the New pattern in column 0 is the same sentence as in Figure 1 but with the omission of the 'n' in 'w i n d s', the addition of 'x q' within the word 'w e s t', and the substitution of 'h' for 't' in 's t r o n g'. Despite these errors, the best multiple alignment created by the SP model is, as shown in the figure, the one that we judge intuitively to be 'correct'.

This kind of ability to cope gracefully with noisy data is really essential in any system which aspires to explain or emulate our ability to recognise things despite fog, snow, falling leaves, or other things that may obstruct our view.

4.2 Family-resemblance or polythetic concepts

A related idea is that the SP system can accommodate 'family-resemblance' or *polythetic* concepts, meaning that recognition does not depend on the presence or absence of any particular feature or combination of features (Wolff 2013, Section 9; Wolff 2006a, Section 6.4.3). This is partly because the system can accommodate errors via its search for optimal solutions (Sections 4.1 and 5.6), and partly because it allows for the specification of knowledge structures that may have alternatives at any or all points in any given structure.

Most of our concepts are polythetic. Although the possession of four legs seems to be a defining feature of the concept 'dog', we would still recognise Fido as a dog, even if he had had the misfortune to lose one of his legs. Likewise for most attributes of most concepts.

As with noisy data, the ability to accommodate polythetic concepts is really essential in any system that aspires to human-like vision.

4.3 Part-whole hierarchies, class hierarchies, and their integration

A strength of the multiple alignment concept is that it provides a simple but effective vehicle for the representation and processing of part-whole hierarchies, class hierarchies, and their integration, as described and illustrated in Wolff (2013, Section 9.1) and Wolff (2006a, Section 6.4.1).

The example in Wolff (2013, Figure 16)—a multiple alignment that integrates botanical categories with the parts and sub-parts of a plant—does not describe the visual appearance of an object, but it should be apparent that this system, when it has been generalised to work with patterns in two dimensions, has potential as a means of representing and processing both the parts and sub-parts of an object's image, and how that information relates to any hierarchy of classes to which that object belongs. Each of those two types of hierarchy is an effective means of expressing visual information in a compressed form.

4.4 Scene analysis

Scene analysis may also be viewed as a kind of parsing (see, for example, Shi (1983)). For the analysis of a seascape, for example, there may be a high-level structure recording the kinds of things that one sees in a typical seascape (sea, beach, sky, rocks, boats, and so on), with a more detailed description for each one of those things.

There seem to be two main complications in scene analysis:

- Any one thing may be partially obscured by another. In our seascape, a boat may be partially obscured by, for example, waves, sea birds, or members of the crew.
- The locations of things may be quite variable. A boat may be in the sea or on the beach; people can appear almost anywhere; and so on.

Of course, people cope easily with both those things, but there may be a problem with 'naive' kinds of parsing system. The SP framework may accommodate these aspects of scene analysis in three main ways:

- As we saw in Section 4.1, parsing can be done successfully despite errors or omission, commission, or substitution. Thus there is reason to believe that, when the SP models have been generalised to work with patterns in two dimensions, an object may be recognised even if it is partially obscured.

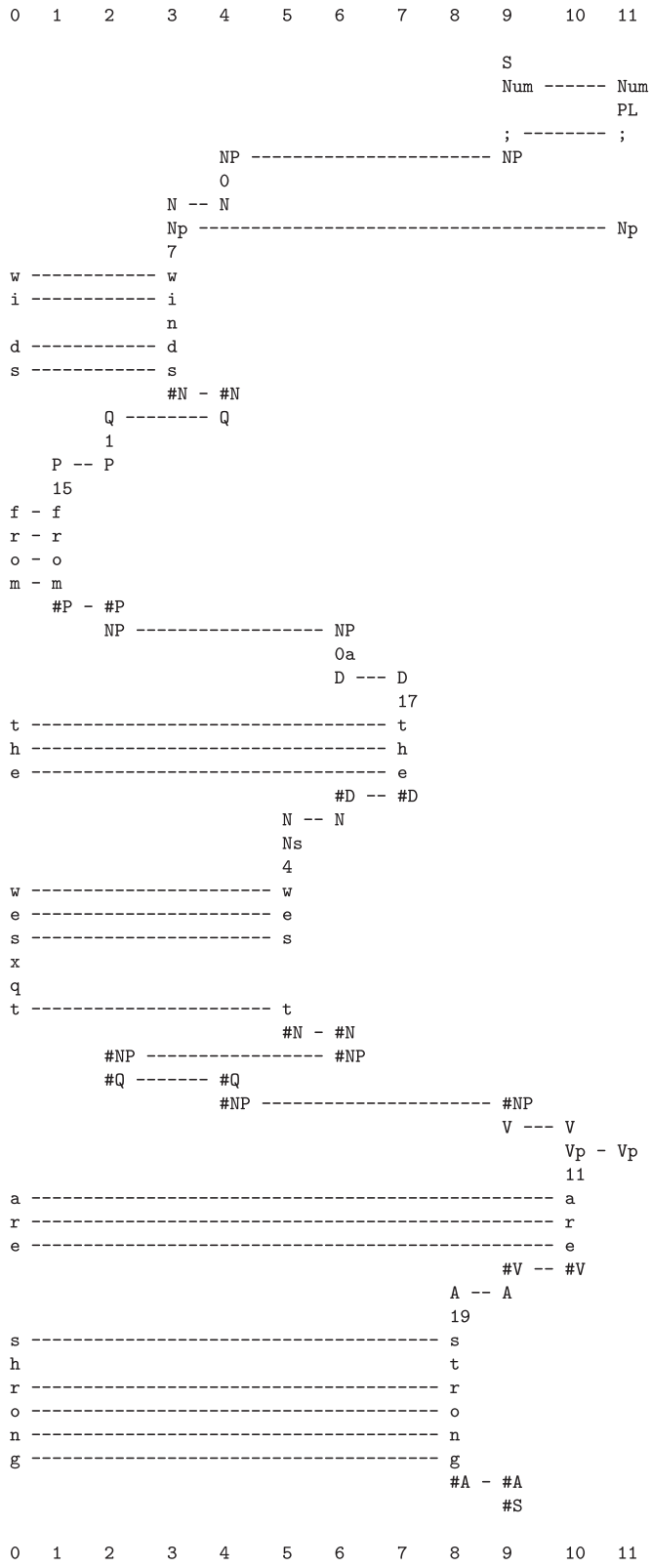

Figure 6 A multiple alignment created like the one shown in Figure 1 but with errors of omission, commission and substitution in the New pattern in column 0 (representing a sentence to be parsed).

- The variability of scenes is broadly similar to the variability of sentences in natural language. Artificial parsing systems, including the SP system, can cope with that variability by providing information about a wide variety of types of sentences and phrases, including recursive forms such as *This is the man all tattered and torn that kissed the maiden all forlorn that milked the cow with the crumpled horn ....* The same principles may be applied to vision.
- Where existing knowledge can't cope, the system may learn—as discussed in Section 5, next.

5 Unsupervised learning and the discovery of objects and classes

There is clearly a close relationship between learning and vision since vision is an important means of gaining new information about the world. In general, we learn via vision in a manner that is 'unsupervised' in the sense that it does not require the intervention of a 'teacher', or the provision of 'negative' samples, or the grading of samples from simple to complex (*cf.* Gold (1967)). We take in information through our eyes (and other senses) and try to make sense of it as best we can.

In this section, we consider unsupervised learning as it has been developed in the SP framework, and how it may be applied in vision. As a preliminary to this discussion, readers may like to consult (Wolff 2013, Section 5), which outlines how unsupervised learning is done in the SP model. In particular, the 'DONSVIC' principle—how 'natural' structures may be discovered via information compression—is described in Wolff (2013, Section 5.2).

In brief, the product of learning from a body of information, **I**, may be seen to comprise: a *grammar* (**G**) which captures *redundancies* in **I**—information that is repeated within **I**—and may be seen as a distillation of the 'essence' of **I**; and an *encoding* (**E**) which expresses the non-redundant features of **I**. In accordance with the principle of *minimum length encoding* (Solomonoff 1964), the SP system aims to minimise the overall size of **G** and **E**.

5.1 The discovery of objects via stereo matching

As with the structures of natural language, it is clear that we have to learn the structures that are significant in vision, including objects[e]. Some insights into how this may be done may be gained from a consideration of random-dot stereograms like the one shown in Figure 7.

Here, each of the two images is a random array of black and white pixels, with no discernable structure. But there is a relationship between them, as shown in Figure 8: both images are the same except that a square area near the middle of the left image is further to the left in the right image.

When these images are viewed in a stereoscope (so that the left image is viewed by the left eye and the right image by the right eye), the central square appears as a discrete object suspended above the background[f]. The focus of interest here will be on how we come to see that discrete object, while possible implications for our understanding of depth perception are discussed in Section 6.3.

A little analysis shows that seeing the central square means finding an alignment between pixels in the left image and pixels in the right image, that there are many alternative such alignments, and that some are better than others. One solution is the algorithm developed by Marr and Poggio (1979). Another potential solution is the kind of processing that builds multiple alignments in the SP models, but generalised for two dimensions. As noted in Wolff (2013, Section 4.3), the complexity of the matching problem can, in general, be reduced by applying constraints to the process of searching and thus reducing the size of the search space.

Figure 9 shows how the SP model can solve a one-dimensional analogue of the stereo matching problem. Here, the Old pattern (row 1) may be seen as an analogue of the left image and the New pattern (row 0) may be seen to stand in for the right image. Both patterns have been prepared from a random sequence of digits[g], with a displacement of the middle section, much as in Figure 8. This multiple alignment is the best of several different multiple alignments created by the SP model with those two patterns.

In the figure, one can see how the central sequence of 10 integers (analogous to the central square in Figure 8) has been isolated from the 'background' sequences to the left and right, and this despite repetitions of integers in both patterns and the formation of plenty of 'wrong' alignments on the route to the 'correct' result. It seems likely that the processes can be generalised to work with patterns in two dimensions.

5.2 Structure from motion

The kinds of processing just described may also be applied to objects in motion.

Consider, for example, a flatfish with a sandy, speckled colouration, lying on a sandy and speckled area on the bed of the sea. Such a creature would be very well camouflaged but with one proviso: it must stay still. As soon as it moves, it will become very much easier to see. Why? Apart from the motion itself, an important reason seems to be that movement creates at least two images (normally more), rather like the two images in a random-dot stereogram. And by a process of matching, much as described above, a predator or other observer will be able to see the fish standing out as a distinct entity with distinct boundaries—like the square that can

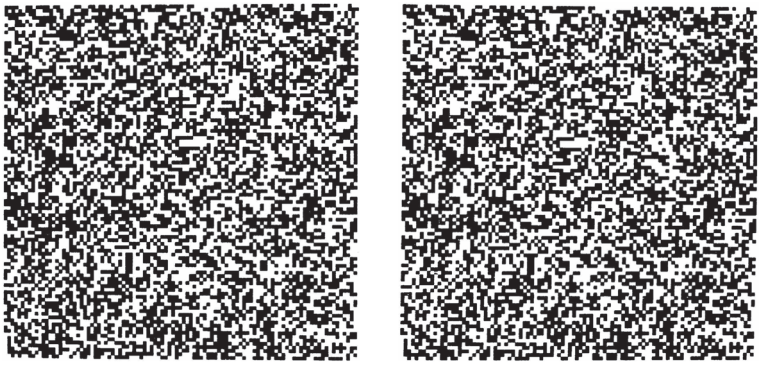

Figure 7 A random-dot stereogram from (Julesz 1971, Figure 2.4-1), reproduced with permission of Alcatel-Lucent/Bell Labs.

be seen when the two images in Figure 7 are viewed in a stereoscope.

More generally, we see any object in motion—such as a car travelling along a road—as a single entity, not a multitude of images like the frames in a video or film. In all such cases, we merge the many instances into one, and likewise for the background. The process of merging those many instances, which is likely to yield high levels of compression, requires a process of matching and unification, much as before. And those processes serve to define the boundaries of the entity and to distinguish it from the background[h].

5.3 Motion and speed

What about the motion itself, and the associated concept of speed? Figure 10 is intended to suggest how these things may be encoded from successive images of a ball moving horizontally in front of a wall.

The raw data is shown in the top four frames with the time for each frame (**T1** to **T4**) marked above them, and with the position of the ball in each frame marked with **P1** to **P4**.

In the first of the lower four frames, the four images of the ball have been merged (unified) into a single image, with '**Bg**' (short for 'ball (green)') as its label or code, in accordance with the *schema-plus-correction* technique for information compression (Wolff 2014c, Section 3; Wolff 2006a, Chapter 2)[i]. In the following 3 frames, the code for the ball ('**Bg**') replaces the image of the ball itself[j].

In a similar way, in that first frame, the 4 images of the background have been unified and marked with '**Wr**' (short for 'wall (red)'), and in the following 3 frames, **Wr** serves to represent the wall[k].

Further compression is possible because there is repetition of the time interval between one frame and the next (marked in the figure with blue arrows), and likewise for the distance that the ball has travelled in each time interval (marked in the figure with red arrows). In each case, the repeated patterns may be merged to create a single instance.

The overall result would be something like the following:

- The unified images of the ball and the wall, each with their associated codes, '**Bg**' and '**Wr**', respectively.

1	0	1	0	1	0	0	1	0	1
1	0	0	1	0	1	0	1	0	0
0	0	1	1	0	1	1	0	1	0
0	1	0	Y	*A*	*A*	*B*	*B*	0	0
1	1	1	X	*B*	*A*	*B*	*A*	0	1
0	0	1	X	*A*	*A*	*B*	*A*	1	0
1	1	1	Y	*B*	*B*	*A*	*B*	0	1
1	0	0	1	1	0	1	1	0	1
1	1	0	0	1	1	0	1	1	1
0	1	0	0	0	1	1	1	1	0

1	0	1	0	1	0	0	1	0	1
1	0	0	1	0	1	0	1	0	0
0	0	1	1	0	1	1	0	1	0
0	1	0	*A*	*A*	*B*	*B*	X	0	0
1	1	1	*B*	*A*	*B*	*A*	Y	0	1
0	0	1	*A*	*A*	*B*	*A*	Y	1	0
1	1	1	*B*	*B*	*A*	*B*	X	0	1
1	0	0	1	1	0	1	1	0	1
1	1	0	0	1	1	0	1	1	1
0	1	0	0	0	1	1	1	1	0

Figure 8 Diagram to show the relationship between the left and right images in Figure 7. Reproduced from (Julesz 1971, Figure 2.4-3), with permission of Alcatel-Lucent/Bell Labs.

```
0     4 7 4 6 4 1 3 7 5     8 5 2 4 0 2 9 1 9 3 8 0 1 4 1 1 2 9 7 1 2    0
      | | | | | | | | |     | | | | | | | | | |     | | | | | | | | |
1 J a 4 7 4 6 4 1 3 7 5 9 4 8 5 2 4 0 2 9 1 9 3     1 4 1 1 2 9 7 1 2 #J 1
```

Figure 9 The best multiple alignment created by the SP computer model with an Old pattern (row 1) and a New pattern (row 0) as one-dimensional analogues of the left and right images in a random-dot stereogram.

- An encoding of the original four frames and their associated information, something like this: '(**Bg**, **Wr**, blue →, red →)*', with the superscript '*' to mark iteration[l].

In accordance with the distinction between 'grammar' and 'encoding' (Section 5, above; see also Wolff (2014a, Section 6.8); Wolff (2014b, Section IV-A)), the first item may be regarded as a grammar for the original data, while the second item may be regarded as an encoding of those data in terms of the grammar. The grammar and the encoding is a compressed representation of the ball, the wall, and the left-to-right motion of the ball at a uniform speed. Further refinement would be needed if the speed of the ball was changing or if the path of the ball was curved.

The waterfall illusion[m] suggests that we may have more than one system for assessing motion and speed: a quick-and-dirty system for when fast responses are needed and a slower-but-more-accurate system which may serve us better when there is less time pressure.

5.4 Deriving concepts from fragments

If we only ever see parts of an object—perhaps a rare creature in its natural habitat that we have only seen in fleeting glimpses—we can nevertheless develop a coherent concept of the whole object via alignments amongst the fragmentary views: 'A B' may be aligned with 'B C' and unified to create 'A B C'; 'C D' may be aligned with 'D E' to create 'C D E'; 'A B C' may be aligned with 'C D E' ..., and so on. This is like the 'sequence assembly' technique in bioinformatics[n], or the stitching together of overlapping photos to create a panorama.

Similar things may be said about how we develop a coherent view of what's in front of us despite frequent saccadic movements of our eyes from one focus of attention to another. The several views may be stitched together to create a single view.

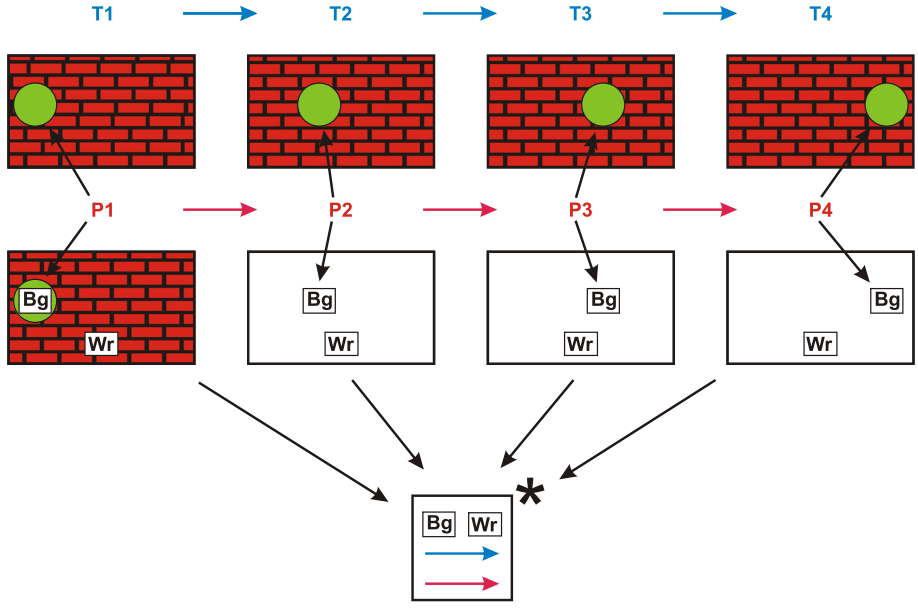

Figure 10 A schematic representation of how successive images of a moving ball, with a wall in the background, may be compressed, with an encoding of the ball's motion and its speed. *Key*: **T1** to **T4**—times for the corresponding images; **P1** to **P4**—successive positions of the ball; **Blue arrows**—each one represents a time interval between one frame and the next, with a past-to-future direction; **Red arrows**—each one represents the distance travelled by the ball between one frame and the next, with the direction of travel; **Bg**—a relatively short name or code for the ball, with an associated colour (green); **Wr**—a relatively short name or code for the wall, with an associated colour (red); *—iteration of the encoding in the envelope.

In both cases, the matching may be achieved via multiple alignment, as developed in the SP theory.

5.5 The discovery of classes of entity, hierarchies of classes, and part-whole hierarchies

Similar things may be said about the learning of categories, classes or concepts like 'person' or 'house', and hierarchies of categories like 'species', 'genus', 'family', 'order', and so on (Section 4.3).

Any such category is a powerful aid to information compression because it saves the need to repeat information in individual instances. If we know that something is a house, we may assume that it has a roof, walls, windows, and at least one door—and that it is likely to have a kitchen, bathroom, bedrooms, and so on. As in object-oriented design, lower-level classes may 'inherit' attributes from higher-level categories.

There is clear potential with the SP system to find the commonalities amongst collections of entities, to create classes and hierarchies of classes like those that have been mentioned, and to learn the relationship between an entity and its parts. As mentioned earlier, the multiple alignment framework can accommodate the seamless integration of class hierarchies with part-whole hierarchies (Wolff 2013, Section 9.1).

5.6 Noisy data and learning

As with such things as the parsing of natural language and pattern recognition (Section 4.1), unsupervised learning via information compression can yield 'correct' results despite errors of omission, commission or substitution in the data which is the input for learning (Wolff 2014b, Section X-B; Wolff 2013, Section 5.3; Wolff 2006a, Section 12.6.1).

As mentioned above, the product of learning from a body of information, **I**, may be seen to comprise: a *grammar* (**G**) and an *encoding* (**E**). As a general rule, errors in **I** are likely to be excluded from **G** for two main reasons:

- Although errors of addition or substitution can, collectively, be quite common, any particular kind of error is likely to be rare. Since **G** captures the redundancies in **I** and excludes non-redundant information, it is likely exclude most of the errors of addition or substitution in **I**.
- A consequence of aiming to minimise the overall size of **G** and **E** (the principle of minimum length encoding) is that **G** may generalise beyond the data in **I** without over-generalising. This means that errors of omission may be corrected in **G** (via generalisation) but, normally, the system would not introduce new errors via over-generalisation.

6 Space and depth

In the SP theory, all kinds of knowledge are represented with patterns in one or two dimensions. Superficially, this seems to rule out anything with more dimensions, and suggests that there might be a need to introduce patterns with three dimensions and possibly more. However, this has been rejected, at least for the time being, for these main reasons:

- Although the multiple alignment concept may in principle be generalised to patterns in three or more dimensions, it is difficult to see how it could be made to work in practice and it looks implausible as a model for any kind of structure or process in the brain.
- A central idea in SP-neural (Section 2) is that the cerebral cortex—which is, topologically, a two-dimensional sheet—may be, in some respects, like a sheet of paper on which *pattern assemblies* (neural analogues of SP patterns) may be written (Wolff 2006a, Chapter 11). This is shown schematically in Wolff (2013, Figure 31).
- If we exclude processes of interpretation in terms of harmonics, colours, or the like, raw sensory data may be seen to come in either one dimension (eg sound) or two (eg visual images).
- Three-dimensional structures may be represented with patterns in two dimensions, somewhat in the manner of architects' drawings (Wolff 2006a, Section 13.2.2). With the development of mathematical concepts within the SP framework (Wolff 2006a, Chapter 10), four or more dimensions may be represented in much the same way as is done now with mathematical techniques.

The following three subsections consider some aspects of the visual perception of space and depth, and how the SP theory may be applied.

6.1 Three-dimensional objects

If an object is viewed from several different angles, with overlap between one view and the next (as illustrated in Figure 11), the several views may be stitched together to create what is at least a partial and approximate 3D model of the object, much as in widely-available systems for creating 3D models from sets of photographs[o].

This is similar to the piecing together of fragments to create a coherent concept, as outlined in Section 5.4. As before, it may be achieved via multiple alignment as that concept has been developed in the SP theory.

The model will be partial if, for example, it excludes views from above or below. And it is likely to be approximate because a given set of views may not be sufficient for an unambiguous definition of the object's geometry: there

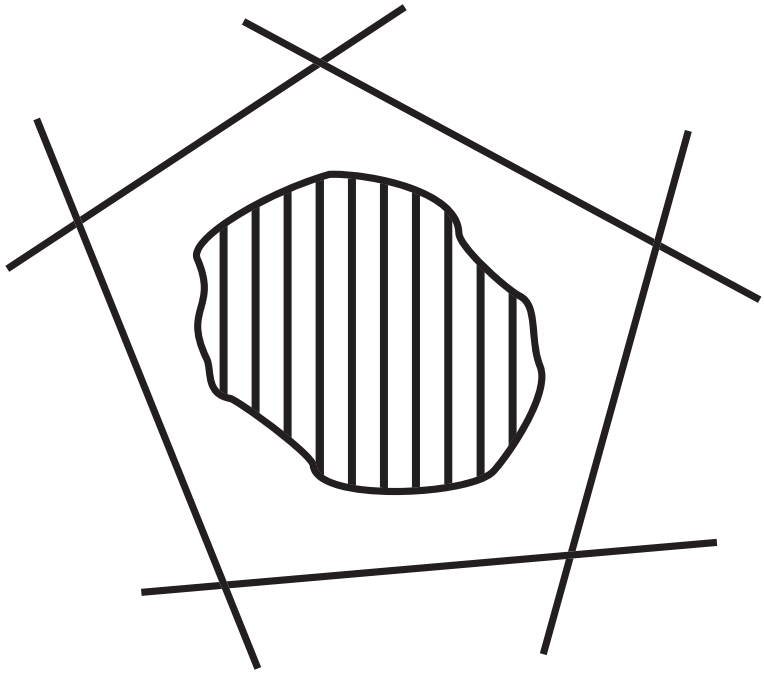

Figure 11 Plan view of a 3D object, with each of the five lines around it representing a view of the object, as seen from the side.

may be variations in the shape that would be compatible with the given set of views.

Do these deficiencies matter? For many practical purposes, the answer is likely to be "no". If we want a rock to put in a rockery, or a stick to throw for a dog, the exact shape is not important. And if we want more accurate information, we can inspect the object more closely, or supplement vision with touch.

Evidence that people do something like what has been described is our ordinary experience that things can be harder to recognised from unfamiliar viewpoints than from familiar ones—the basis of some trick photos. That observation is confirmed in experimental studies showing that people are both slower at recognising things, and less accurate, when the viewpoint is unfamiliar (Bülthoff and Edelman 1992; Tarr 1995; Tarr and Pinker 1989).

Evidence that our vision can be a rather poor guide to 3D structure is the failure of most people to see the distorted nature of the Ames' distorted room (see also Section 6.3.1)[p]. Somehow, this deficiency in our vision does not prevent us from dealing very effectively with cups, plates, tables, and other everyday objects, or from from finding our way around. It seems likely that, in a similar way, robots may be effective in many situations without a geometrically-accurate knowledge of objects and surroundings.

Although what has been described is like the stitching together of overlapping photos to create a panorama, the SP theory suggests that, with people, the visual information would be compressed via the encoding, within the SP system, of part-whole relations, class-inclusion relations, and other kinds of regularities[q].

6.2 Building a model of one's environment and finding one's way around

Similar processes may be at work when we move around in our environment and learn about it. Successive views that overlap each other may be stitched together, as before, to create a model of the streets or other places where we have been. This is essentially what is done with Google's "Street View"[r] and appears to be the basis for the creation of 3D maps using smartphones in the Google-funded "Project Tango"[s].

The main difference between what has been achieved with Street View and what is envisaged for the SP system is that, in the latter case, as noted in Section 6.1, visual information would be compressed via the mechanisms in the SP system.

As with objects (Section 6.1), a model of our environment that is created via overlapping views may not be geometrically precise[t]. But, as before, some ambiguity may not matter very much for many practical purposes. Topological maps, such as the classic map of the London underground, can be quite good enough for finding one's way around. However, if greater geometric accuracy is needed, it may be increased by gathering more

information, especially information about areas between roads, paths or other routes.

In connection with finding one's way around, the SP system may be relevant in two ways:

- If a robot has stored representations of one or more places, perhaps compressed via recurrent patterns as indicated in Wolff (2013, Section 2.1), then, via the building of multiple alignments (as in Section 4), it should be able to recognise when it has reached one of those places, using incoming visual information as New patterns and stored knowledge as Old patterns. If it has stored information about an entire route or network of routes, then, within that environment, it should be able to identify where it is at any time. Similar things may be true of people.
- With an appropriate set of Old patterns, each one of which represents a direct connection between two places, the SP system, via the building of multiple alignments, can work out one or more routes between any two of the relevant places, including routes via two or more of the direct connections (Wolff 2006a, Chapter 8). The example in Figure 12 shows one such flying route between Beijing and New York.

These points about how we may build a model of our environment and find our way around relate to the topic in robotics of 'simultaneous localization and mapping' (SLAM)[u].

6.3 Depth perception and stereoscopic vision

Without attempting a comprehensive discussion of the complex subject of depth perception, this section offers some thoughts about stereoscopic vision, and the possible relevance of the SP theory.

6.3.1 Triangulation

For any given object that we are looking at, we can in principle work out its distance by a process of triangulation like that which has been widely used in cartography, at least as it used to be. But there appear to be snags:

- For this mechanism to work with reasonable accuracy, it would be necessary for one to have a rather accurate sense of the direction of gaze for each eye and the angle between that direction of gaze and the line between the two eyes. It seems unlikely that we can sense the positions of our eyes with the necessary accuracy.
- There is evidence that, with the previously-mentioned Ames' distorted room illusion, the illusion persists when people view the room with two eyes (Glennerster et al. 2003), although, in that case, the effect may be reduced (Gehringer and Engel 1986). This suggests that any information about distance that may be gained via triangulation, or any other clue such as the focussing of our eyes, is not sufficiently clear or precise to overcome viewers' preconceptions that the room has the conventional rectangular form.
- Triangulation cannot work with a stereoscope or a 3D film because what we are looking at is all at one distance, with nothing to differentiate one part of the picture from another. The spear which makes us jump as we see it coming out of the cinema screen towards us is no closer to us than anything else in the film.

We cannot rule out triangulation altogether—it may have a role in some situations—but some other mechanism is needed to explain how we see depth with a stereoscope or a 3D film.

6.3.2 Possible alternatives

With random-dot stereograms, it is clear that our brains are capable of forming an alignment between the left and right images that is good enough to identify the displaced area in the middle as a discrete entity (Section 5.1).

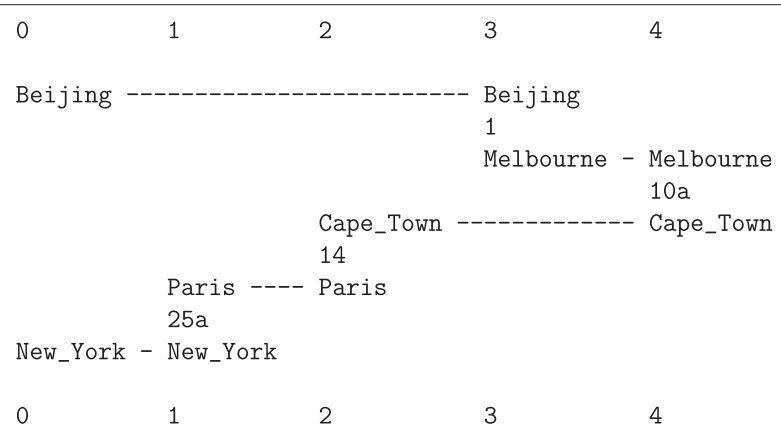

Figure 12 A multiple alignment showing a flying route between Beijing and New York, one of several produced by the SP computer model with a set of Old patterns, one for each leg of this and other possible journeys. Reproduced from Figure 8.5 (e) in Wolff (2006a), with permission.

By identifying the displaced area and distinguishing it from the surrounding area, we may also gain an accurate knowledge of the size of the displacement.

What can the size of the displacement (D) tell us about depth? With a knowledge of the distance between the eyes but without other information, it can tell us the *ratio* of the distances between the observer and the object (A), and between the object and the background (B). With some other information about the first of these distances, we can infer the second distance, and *vice versa*.

But the foregoing presupposes a knowledge of geometry. Since geometry is a relatively recent invention in the millions of years that there have been people on earth, and since a knowledge of geometry would normally require schooling and would in any case not be available to animals, it seems likely that some other mechanism is needed to explain how, in most cases, people and other animals make judgements about depth using observations about displacements.

The tentative answer suggested here is that:

- From experience, we build a table of associations between the distances A, B, and D.
- In later situations, a knowledge of any two of those three values would enable us look up the third.

Both of these processes are well within the scope of the SP system. The system is designed to discover recurrent patterns in data, including the kinds of association considered here (Wolff 2013, Section 5, Wolff 2006a, Chapter 9). And table look-up can be modelled quite simply via multiple alignment (Wolff 2007, Section 3.1, Wolff 2006a, Section 6.3).

With those kinds of capability, we have a means of making inferences about depth that may not always be geometrically accurate but may be good enough for many practical purposes.

7 Some other aspects of vision

The SP theory has things to say about some other aspects of vision, as discussed in the following subsections.

7.1 Seeing things that are not there

As noted in Section 3, we often 'see' things that are not objectively present in what we are looking at. We may see 'subjective contours' in certain kinds of images, or we may see the edge of a leaf where it overlaps another leaf despite there being little or nothing to mark the boundary.

The multiple alignment in Figure 1 provides an example of how the SP system may accommodate these kinds of things. Here, the New pattern is the sentence '`w i n d s f r o m t h e w e s t a r e s t r o n g`' with nothing to mark the boundary between one word and the next. But those boundaries are clearly marked in the figure via the parsing of the sentence into its constituent parts.

More generally, we infer things that are not immediately visible: when we see the unbroken shell of a nut, we expect to find an edible kernel inside; when we see a horse partially obscured by a tree, we expect to see the whole animal when it moves into full view; and so on.

This kind of inference is an integral part of how the SP system works. In Figure 6, the word '`w i n d s`' appears in the New pattern as '`w i d s`', but the parsing interpolates the missing '`n`'. In an example of pattern recognition (Wolff 2013, Section 9, Figure 16), the SP system identifies an unknown plant as, probably, an example of the Meadow Buttercup (*Ranunculus acris*) from a few features in the New information with which it is supplied: that the stem is hairy, that the petals are yellow, that there are numerous stamens, and more. From its stored knowledge about that species, the system can infer several other things, not in the New information: that the plant photosynthesises, that it has five petals, that it is poisonous, and more.

7.2 Variations in image size

Variations in the sizes of images are significant in both natural vision (Section 7.2.1) and in the processing of images by computers (Section 7.2.2).

7.2.1 Recognition despite variations in image size

A prominent feature of natural vision is that we can recognise something despite wide variations in viewing distance and corresponding variations in the size of the retinal image[v]. Although this phenomenon is not consistent with any simple pattern-matching model of vision, it appears that it can be accommodated within the SP framework.

Let us suppose that, as described in Section 3.3, the image to be processed is reduced to a 'primal sketch', showing boundaries between uniform areas but without the redundancy within those areas. And let us suppose that the redundancy in any line with a uniform direction or curve has been reduced or eliminated as described in Section 3.4.

For any given image, the effect of those kinds of processing will be to reduce or eliminate variations in the size of the original image. The encoding that is derived from a large version of the scene will be much the same as the encoding that is derived from a small version.

This kind of capability in the SP system, together with the system's flexibility in matching patterns (Wolff 2013, Section 4.2), should mean human-like capabilities in recognising things despite wide variations in the sizes of images of any given entity.

7.2.2 Unifying raster graphics and vector graphics

With image processing, it is generally understood that raster graphics are good for photographs or paintings but go fuzzy if images are enlarged too much, while vector graphics have the advantage that everything stays sharp when images are enlarged but are really only suitable for things like text and diagrams.

The SP system may provide a means of overcoming this image-processing apartheid. Images of any kind may be encoded economically as described in this article. But unlike raster graphics, images encoded via the SP system should be expandable without loss of definition, much as in vector graphics. The key to this scalability of SP-encoded images appears to be recursive encoding of regions of uniformity—in the manner of run-length coding—applied to areas (Section 3.3) and also to lines (Section 3.4). The recursive encoding of areas and lines should allow the whole image to be expanded indefinitely without loss of definition.

Any such encoding would be a useful step forward in the development of a UFK (Wolff 2014b, Section III).

7.3 Lightness constancy and colour constancy

Another prominent feature of natural vision is 'lightness constancy': the fact that, normally, we perceive the lightness of an object to be fixed, despite wide variations in the intensity of the incident light and corresponding variations in the amount of light that is reflected from the object (its 'luminence'). We would normally see a lump of coal as black and snow as white, even though the coal in bright sunlight may be reflecting more light per unit area than snow in shadow.

In order to account for this phenomenon, it seems necessary to suppose that, for each kind of object, we maintain some kind of table of associations between levels of illumination and corresponding values for luminance. Since we are unlikely to have an inborn knowledge of coal, snow, and the like, we must suppose that those tables are learned. As noted in Section 6.3.2, learning associations of that kind is part of what the SP system is designed to achieve.

Notice that any given table can only be applied if we have some idea of what kind of object we are looking at, otherwise we might see coal as if it was snow, or *vice versa*. There is some evidence that our perception of the lightness of an object does indeed depend on what we think the object is (Frisby and Stone 2010, Chapter 16). In a similar way, our judgements of lightness seem to depend on our perceptions of how a given object is illuminated (Stone 2012, Figure 1.10).

It seems likely that much of what has been said in this section about lightness constancy would also apply to colour constancy: the way we see the colour of an object to be fixed, despite wide variations in the colour of the incident light and corresponding variations in the colour of the light that is reflected from the object.

In terms of information compression as a unifying principle in computing and cognition, it is pertinent to mention that lightness constancy and colour constancy may each be seen as a means of encoding information economically. It is simpler to remember that a particular object is 'black' or 'red' than all the complexity of how its appearance changes in different lighting conditions.

7.4 The role of context in recognition

It is often remarked that we recognise things more easily in their familiar contexts than in unfamiliar ones. This is confirmed in formal studies (see, for example, Bar and Ullman 1993; Oliva and Torralba 2007).

This observation makes sense in terms of the SP framework because any part of a multiple alignment may be a context for any other, and because of the way the system searches for a global optimum which embraces any given entity and its context. If, in our seascape example (Section 4.4), we see a beach and the sea then, in effect, we are primed to see boats—because, in that context, boats are likely to yield multiple alignments with better scores than, say, office furniture.

Context also has a role in resolving ambiguities, as outlined at the end of the next subsection.

7.5 Ambiguity in perception

A prominent feature of perception is that, with some kinds of sensory input, there may be more than one plausible interpretation. Examples include the 'young woman/old woman' picture and the 'duck/rabbit' picture of psychology text books. An example with natural language is the ambiguous sentence *Fruit flies like a banana*, the second part of *Time flies like an arrow. Fruit flies like a banana*, attributed to Groucho Marx.

In the SP framework, this kind of ambiguity is accommodated in the way that, with appropriate data, the system may create two or more multiple alignments that have good scores. With *Fruit flies like a banana*, the two parsings can be seen in Wolff (2006a, Figure 5.1). Another example is the ambiguity in the phoneme sequence '`ae i s k r ee m`'[w], which can be read as *ice cream* or *I scream*.

The way in which the SP system may use context to disambiguate '`ae i s k r ee m`' is described in Wolff (2006a), Section 5.2.2 with examples—multiple alignments with phonetic versions of *I scream loudly* and *Ice cream is cold*—in Wolff (2006a, Figure 5.4).

7.6 Integration of vision with other senses and other aspects of intelligence

It is clear that in people and other animals, vision does not stand alone but works in close association with other

senses. Our concept of a ship, for example, is an amalgam of images, sounds, smells, the flavour of food on board, textures of different surfaces, and so on. In a similar way, natural vision works closely with other aspects of intelligence: unsupervised learning, several kinds of reasoning, understanding and producing natural language, recalling information, and non-visual kinds of recognition.

Achieving these kinds of integration without undue complexity has been a central aim in the development of the theory. And in that development, many candidate ideas have been rejected because they did not help to promote the simplification and integration of concepts. Now, the main planks of the theory are: representing all kinds of knowledge with patterns in one or two dimensions; the multiple alignment concept as it has been developed in the SP theory; and the overarching role of information compression via the matching and unification of patterns, in both the representation and processing of knowledge.

In artificial systems, the integration of vision with other sensory modalities and other aspects of intelligence is necessary for the development of versatility and adaptability in such systems.

8 Conclusion

As described in the Introduction, the main aim of this paper has been to explore how the SP theory may be applied to the understanding of natural vision and the development of computer vision, and to discuss associated issues.

The SP theory has things to say about several aspects of vision:

- Low level perceptual features such as edges or corners may be identified by the extraction of redundancy in uniform areas in a manner that is analogous to the run-length encoding technique for information compression, and comparable with the effect of lateral inhibition in the visual systems of animals.
- The concept of *multiple alignment* in the SP theory may be applied to the recognition of objects, and to scene analysis, with a hierarchy of parts and sub-parts, and at multiple levels of abstraction.
- The theory has potential for the unsupervised learning of visual objects and classes of objects, and suggests how coherent concepts may be derived from fragments. It provides an account of how we may discover objects via stereo matching and via motion.
- As in natural vision, both recognition and learning in the SP system is robust in the face of errors of omission, commission and substitution.
- The theory suggests how, via vision, we may piece together a knowledge of the three-dimensional structure of objects and of our environment that is good enough for many practical purposes.
- The theory provides an account of how we may see things that are not objectively present in an image.
- The theory suggests how we may recognise something despite variations in the size of its retinal image. It also suggests how images may be represented in a way that combines the advantages of raster graphics and vector graphics.
- The theory has things to say about the phenomena of lightness constancy and colour constancy, about the role of context in recognition, and about ambiguities in visual perception.

A strength of the SP theory is that it is not simply a theory of vision. It is designed to achieve an overall simplification and integration of concepts in computing and cognition. There is clear potential for the integration of vision with other sensory modalities and with other aspects of intelligence such as unsupervised learning, processing of natural languages, reasoning, planning, and problem solving, with consequent benefits in terms of versatility and adaptability.

A high-parallel, open-source version of the SP machine, as outlined in Wolff (2013, Section 3.2), would provide a means for researchers to explore what can be done with the system and to create new versions of it.

Endnotes

[a] See, for example, 'Run-length encoding', Wikipedia, bit.ly/eyxlY, retrieved 2013-02-04.

[b] Compared with the multiple alignment in Figure 1, this one has been rotated by 90°, replacing columns with rows. The two styles are equivalent. The choice between them depends largely on what fits best on the page.

[c] In the SP framework, any Old pattern may appear more than once in a multiple alignment. Here, an *appearance* of a pattern is *not* the same as an *instance* of a pattern, as explained in Wolff (2006a, Section 3.4.6).

[d] For a description of the method of deriving an encoding from a multiple alignment, see Wolff (2013, Section 4.1) or Wolff (2006a, Section 3.5).

[e] The Chomskian doctrine that children 'acquire' their native language or languages via an inborn knowledge of 'universal grammar' depends on the still-unproven idea that such a grammar can be defined for all the world's languages, and it is still not clear how the acquisition process might work.

[f] Some people are able to see the square by viewing the images directly, with some defocussing to help merge them into one.

[g] The random sequence of digits, with values between 0 and 9, inclusive, was generated by the Random Integer Generator from Random.org (www.random.org). The results are, they say, better than with pseudo-random

number algorithms because "atmospheric noise" is the source of randomness.

[h]Of course, there may be additional factors, such as changing perspectives as the moving object passes the observer, but the principles that have been outlined would still apply.

[i]In this discussion, we shall ignore any compression that may be achieved by taking advantage of uniformity and corresponding redundancy within the image of the ball itself, as discussed in Section 3.

[j]The relatively large size of '**Bg**' in the figure, with its surrounding envelope, may suggest that it does not yield much compression of the image of the ball. But it has been shown quite large simply to ensure that it is readable. Likewise for '**Wr**' and its envelope. In practice, it is likely that labels like these would be encoded with far fewer bits than would be required for the things they represent.

[k]As with the ball, we shall ignore any compression that may be achieved via the detection and extraction of redundancy within uniform areas within the image of the wall.

[l]Some additional information may be needed to show inter-relationships amongst the symbols.

[m]See "Motion aftereffect", Wikipedia, bit.ly/1l6nX5g, retrieved 2014-03-20.

[n]See "Sequence assembly", Wikipedia, bit.ly/1CsZOvi, retrieved 2013-02-21.

[o]See, for example, "Big Object Base" (bit.ly/1gwuIfa.bit.ly/1gwuIfa), "Camera 3D" (bit.ly/1iSEqZu), or "PhotoModeler" (bit.ly/MDj70X.)

[p]For readers who are not familiar with this illusion, a person looks into one end of a room that appears to have a conventional rectangular form but is actually constructed so that one of the two corners opposite the viewer is stretched away and is relatively high, while the other corner is nearer to the viewer and is relatively low. Anyone standing in the near corner appears to be large, and they appear to shrink if they walk to the far corner. A description of the Ames' room and the illusion may be found in "Ames room", Wikipedia, bit.ly/1Ct0mkV, retrieved 2013-03-30.

[q]It is true that, in digital systems, photos are normally compressed via JPEG or similar technique. But, as indicated in Wolff (2013, Section 5.2), there is potential for the SP system to yield higher levels of compression and more natural structures.

[r]See bit.ly/1rprOMJ.

[s]See "Software would make 3-D maps using smartphones", The Columbus Dispatch, 2014-03-24, bit.ly/1gqjZBe.

[t]From my own experience of exploring caves, I know that, while one can build up a good knowledge of how different passages connect with each other, one's understanding of their 3D geometry can be hazy, and can lead to some surprises if one has the opportunity to see a 3D model that is based on a proper survey, with measurements of distances and angles.

[u]See, for example, "Simultaneous localization and mapping", Wikipedia, bit.ly/1ikQTRR, retrieved 2014-04-04.

[v]This is related to but different from the phenomenon of 'size constancy': that, within limits, our perception of an object's size remains the same, regardless of viewing distance or the size of the retinal image.

[w]The symbols used are an alphabetic adaptation of phoneme symbols in the International Phonetic Alphabet.

Competing interests
The author declares that he has no competing interests.

Received: 4 May 2014 Accepted: 3 September 2014
Published: 21 October 2014

References

Attneave F (1954) Some informational aspects of visual perception. Psychol Rev 61:183–193
Barlow HB (1959) Sensory mechanisms, the reduction of redundancy, and intelligence. In: HMSO (ed) The mechanisation of thought processes. Her Majesty's Stationery Office, London, pp 535–559
Bar M, Ullman S (1993) Spatial context in recognition. Perception 25:324–352
Bülthoff HH, Edelman S (1992) Psychophysical support for a two-dimensional view interpolation theory of object recognition. Proc Natl Acad Sci 89(1):60–64
Farabet C, Couprie C, Najman L, LeCun Y (2013) Learning hierarchical features for scene labeling. IEEE Trans Pattern Anal Mach Intell 35(8):1915–1929. doi:10.1109/TPAMI.2012.231
Frisby JP, Stone JV (2010) Seeing: the computational approach to biological vision. The MIT Press, London
Gehringer WL, Engel E (1986) Effect of ecological viewing conditions on the Ames' distorted room illusion. J Exp Psychol Hum Percept Perform 12(2):181–185
Glennerster A, Gilson SJ, Tcheang L, Parker AJ (2003) Perception of size in a 'dynamic Ames room'. J Exp Psychol Hum Percept Perform 3(9):490. doi:10.1167/3.9.490
Gold M (1967) Language identification in the limit. Inform Contr 10:447–474
Han F (2005) Bottom-up/top-down image parsing by attribute graph grammar. In: Proceedings of the tenth IEEE international conference on computer Vision (ICCV 2005), 17–21 Oct. 2005. IEEE Computer Society, Los Alamitos, pp 1778–1785
Julesz B (1971) Foundations of cyclopean perception. Chicago University Press, Chicago
Li M, Vitányi P (2009) An introduction to Kolmogorov complexity and its applications. Springer, New York
Marr D (2010) Vision: a computational investigation into the human representation and processing of visual information. The MIT Press, London, England. This book was originally published in 1982 by W. H. Freeman and Company
Marr D, Poggio T (1979) A computational theory of human stereo vision. Proc Roy Soc Lond B 204(1156):301–328
Oliva A, Torralba A (2007) The role of context in object recognition. Trends Cogn Sci 11(12):520–527
Ratliff F, Hartline HK (1959) The response of limulus optic nerve fibres to patterns of illumination on the receptor mosaic. J Gen Physiol 42:1241–1255
Shi QY (1983) Parsing and translation of (attributed) expansive graph languages for scene analysis. IEEE Trans Pattern Anal Mach Intell PAMI-5(5):472–485
Solomonoff RJ (1964) A formal theory of inductive inference. Parts I and II Inform Contr 7:1–22 and 224–254

Stone JV (2012) Vision and brain: how we perceive the world. The MIT Press, London
Tarr MJ (1995) Rotating objects to recognize them: a case study of the role of viewpoint dependency in the recognition of three-dimensional objects. Psychonomic Bull Rev 2(1):55–82
Tarr MJ, Pinker S (1989) Mental rotation and orientation-dependence in shape recognition. Cogn Psychol 21(2):233–282
von Békésy G (1967) Sensory inhibition. Princeton University Press, Princeton
Wolff JG (2006a) Unifying computing and cognition: the SP theory and its applications. CognitionResearch.org, Menai Bridge. ISBNs: 0-9550726-0-3 (ebook edition), 0-9550726-1-1 (print edition). Distributors, including Amazon.com, are detailed on bit.ly/WmB1rs
Wolff JG (2006b) Medical diagnosis as pattern recognition in a framework of information compression by multiple alignment, unification and search. Decis Support Syst 42:608–625. See bit.ly/XE7pRG
Wolff JG (2007) Towards an intelligent database system founded on the SP theory of computing and cognition. Data Knowl Eng 60:596–624. See bit.ly/Yg2onp
Wolff JG (2013) The SP theory of intelligence: an overview. Information 4(3):283–341. doi:10.3390/info4030283. See bit.ly/1hz0lFE
Wolff JG (2014a) The SP theory of intelligence: benefits and applications. Information 5(1):1–27. doi:10.3390/info5010001. See bit.ly/1lcquWF
Wolff JG (2014b) Big data and the SP theory of intelligence. IEEE Access 2:301–315. doi:10.1109/ACCESS.2014.2315297
Wolff JG (2014c) Information compression, intelligence, computing, and mathematics. Technical report, CognitionResearch.org. In preparation. See bit.ly/1jEoECH

doi:10.1186/2193-1801-3-552
Cite this article as: Wolff: Application of the SP theory of intelligence to the understanding of natural vision and the development of computer vision. *SpringerPlus* 2014 **3**:552.