Gestalt theory and computational image processing.

Mark Lusted (22084992)
November 2005
Goldsmiths College

In this essay I briefly examine the origins of Gestalt theory, its implications for computational image processing and finally investigate examples of programmers using it's laws for task dependent image processing and general models.

Introduction

Gestalt theory played an important role in sparking sustained scientific research into human visual perception in the last century, and was ultimately successful in displacing the notion that we see things for the simple reason that that is 'how they are.' Since computers became increasingly important to data analysis and modelling in the latter half of the 20^{th} century, the relationship between Gestalt theory and the emerging computer science has been multi-faceted, with some seeking to utilise it's 'laws' for specific tasks in computational image processing while others sought to build accurate computer models of how our own neural networks process visual data.

Section 1: Origins

The three key thinkers central to the emergence of Gestalt theory at the beginning of the twentieth century were Kurt Koffka, Max Wertheimer and Wolfgang Köhler. Gestalt theory was positioned to challenge the notion that we 'learn' how to see things simply through past experiences, or because that was 'how they are.' They instead argued that we have an innate ability to instantly understand complicated visual scenes by perceiving what 'belongs together,' and pointed to ways in which this could be tested and 'tricked'. This ability was, for these theorists, an example of the holistic, self-organising nature of the human body.

In addition to this more abstract aspect of their theory, they usefully outlined what they perceived as being the central factors that influence what we see when presented with visual stimuli. According to this theory we 'see' objects according to the following rules:

- Similarity: Elements with similar physical properties (size, colour ,etc) are naturally grouped together as likely being part of the same object.

- Proximity/Contiguity: The smaller the distance between two elements, the more likely we are to view them as part of the same object. This 'law' can be outweighed in importance by other Gestalt 'laws' (see point below.)

- Symmetry: Objects in symmetrical alignment are perceived as belonging to the same object. This often occurs despite elements not having close proximity to one another.

- Good continuation: Our mind often continues to see a pattern beyond where the visual stimuli had ended.

- Common Fate: Elements with the same moving direction are seen as a unit.

- Closure: Our mind adds missing elements to complete a figure.

Coupled with the above 'laws', Gestalt theorists stressed the importance of how we naturally perceive the *whole* when presented with a scene, as opposed to focussing individually on the fractured elements that often exist in our scope of vision. Wertheimer, for instance, argued that 'when we are presented with a number of stimuli we do not as a rule experience "a number" of individual things… instead larger wholes separated from and related to one another are given in experience.'[1]

In a philosophical sense they viewed the sum of all the parts as containing unique properties not existing unless grouped together. At a very basic level it is evident that a dot has 'just three perceptual properties: colour, size and position.' When aligned together in a line, an array of dots has further emergent properties not evident in the original, 'such as length, orientation and curvature.'[2] Observations such as these emphasised the importance of understanding the 'gestalt' qualities of objects and scenes that existed, and further underlined their philosophical positions on the holistic nature of the world as whole.

It was such views that lead Kofka to argue for a 'new' kind of science. Instead of focussing all their attention on gaining knowledge applicable to their particular field of study, he argued that scientists should seek 'knowledge of the rational system, the interdependence of all facts.'[3] Science, therefore, 'gains in value and significance not by the number of individual facts it collects but by the generality and power of its theories.'[4]

The influence of Gestalt theory was initially hindered by circumstances external to academia; the rise of National Socialism in Germany in the 1930s. Coupled with this many scientists, while accepting many of the specific insights that Gestalt-inspired experiments into visual perception had produced, were unwilling to accept the theory and philosophy as a whole. As one author noted, 'the Gestaltists were largely successful in arguing against their predecessors' ideas but were less successful in promoting their own.'[5]

---

[1] M. Wertheimer, *Laws of Organisation in Perceptual Forms* (http://psy.ed.asu.edu/%7Eclassics/Wertheimer/Forms/forms.htm).
[2] S. E. Palmer, *Vision Science* (Massachusetts Institute of Technology, 1999), p.50.
[3] K. Koffka, *Perception: An introduction to the Gestalt-theorie* (http://psy.ed.asu.edu/%7Eclassics/Koffka/Perception/perception.htm).
[4] K. Koffka, *Principles of Gestalt Psychology* (http://www.marxists.org/reference/subject/philosophy/works/ge/koffka.htm)
[5] S. E. Palmer, *Vision Science*, p.50.

Section 2: Computational age

Despite this, many academic disciplines have embraced elements of Gestalt theory in some form or another since its inception nearly a century ago. David Marr, for instance, argued that the Gestalt 'laws of organisation' are useful to scientists seeking to understand visual cognition for the simple reason that these 'laws' were built upon observations and assumptions about the world of objects that were both 'sensible,' and tangible. For instance, as different parts of the surface of an object will often reflect light in a similar way, 'similar elements in an image are quite likely to belong together; as smooth shapes are common in nature, smoothly changing contours should be preferred; as objects have closed boundaries, their contours in an image are likely to be closed,' and so on.[6] Or, put another way, 'the structure of our world is such that neighbouring entities will with a greater likelihood belong together than those that are not adjacent to each other.'[7]

It was Marr who revolutionised the way in which many scientists attempt to understand human vision today with his seminal work; *'Vision: a computational investigation into the human representation and processing of visual information.'* (1982). While many of his specific predictions have proven to be inaccurate in the two decades since his work was published, his central thesis of utilising mathematical (and by implication, computational) models and systems as a means of furthering knowledge of our own organic methods of visual perception has proved far more resilient. The application of Gestalt theory to computer science has resulted in two broad categories of outcomes:

1. The use Gestalt theory to perform specific, task-dependent image processing.

2. The use Gestalt theory to build computational models mimicking human neural processes, both to build successful 'artificial' visual cognition systems, and also to help further our understanding of 'organic' visual cognition.

One practical example of point 1 can be seen in the *Speed-FX* plug-in, which was created for detecting and mending a common affliction in old film reels; line scratching.

---

[6] Cited in: G. W. Humphreys and V. Bruce, *Visual Cognition* (Hove: Lawrence Erlbaum Associates Ltd, 1991) p.29.
[7] Florentin Worgotter et al, *Early cognitive vision: Using Gestalt-laws for task-dependent, active image-processing*, (http://www.pspc.dibe.unige.it/ecovision/pubs/papers/woe_etal2004_naturalcomputing.pdf).

Fig.1. Frames of a film to be modified by SpeedFX

One strategy used by the creators of the system was to focus on a specific aspect of Gestalt theory, the importance of similarity in visual scenes, and build a system that seeks to identify common characteristics ('e.g., alignment, colour constancy, similarity of shape or texture, ...') within particular frame. The computer is programmed to respond to anomalies in the 'Gestalt', and amend them according to set rules.[8] In this example we can see a practical application of specific insights of Gestalt theory to accomplish a precise task, while not attempting to mimic human visual cognition in its entirety.

An example of point 2 can be seen in the work of a team at the University of Bielefeld in Germany who utilised a number of the Gestalt 'laws' to build an 'artificial neural network' aimed at processing image data.[9] Their project, titled 'Mechanisms of perceptual grouping', built a model for computational image processing which applied the gestalt laws of 'proximity, good continuation, symmetry, and closure' to a systematic computer model.

Fig. 2 (below) demonstrates the hierarchical approach that the Bielefeld team adopted; with the bottom level examining only 'one-dimensional primitives that are grouped according to collinearity, curvilinearity, and proximity.' The middle level 'consists of symmetric and parallel grouping hypotheses' while the top 'level encloses hypotheses of closed contours.'[10]

[8] '*Speed-FX*' - http://www.tecn.upf.es/gti/speedfx/htmls/plug.htm.
[9] *'Perceptual Grouping'* - http://www.techfak.uni-bielefeld.de/~posch/PG/grouping.html.
[10] ibid.

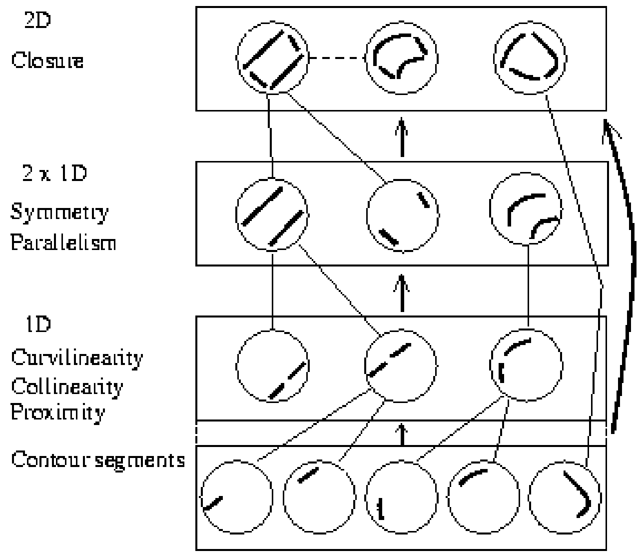

Fig.2. The 'artificial neural network' model created at the University of Bielefeld

The 'artificial neural network' that they constructed is notable for its 'bottom-up,' or sequential approach. While theoretically successful as a template for building computer systems such as the *Speed-FX* plug in above, such models have proved problematic for those seeking to use Gestalt theory to further understand human visual processing. Recent studies into human visual neural networks suggest that the layers of material in our eyes do not, as once argued by David Marr, simply pass 'up' information**,** with each part of the cycle achieving more sophisticated perception than the one directly below.[11]

The true method in which visual data is processed in our neural networks is likely to be far more dynamic, with different layers in the process passing information 'up' and 'down'. The precise nature of the flow of data is, however, still unknown, and is contested amongst those working in the field of visual cognition. Computational models that seek to apply Gestalt theory in a linear, sequential approach therefore seem hopelessly out of sync with the data emerging from studies of human visual neural networks.

However, one author, Steven Lehar, has sought to reflect the complexity of Gestalt theory, in particular the concept of emergence, within a *perceptual* model rather than one seeking to mimic neural networks. In doing so he argues that he has been able to resolve 'the fragmented and hierarchical processing architecture revealed in neurophysiological studies with the global unity and coherence of the visual world revealed by Gestalt theory.'[12]

---

[11] Florentin Worgotter, Norbert Kruger, Nicolas Pugeault, Dirk Calow, Markus Lappe, Karl Pauwels, Marc Van Hulle, Soriva Tan and Alan Johnston, *Early cognitive vision: Using Gestalt-laws for task-dependent, active image-processing*, p.294.
[12] S. Lehar, *Computational Implications of Gestalt Theory: The Role of Feedback in Visual Process* (http://cns-alumni.bu.edu/pub/slehar/webstuff/orivar/orivar3.html).

His model, titled Multi-Level Reciprocal Feedback (MLRF), freed from the restraints of attempting to accurately depict human neural networks due its perceptual nature, lays out a system of inter-relational dependence between various layers of processing. Crucially, the model allows for information to be passed from layer to layer in 'simultaneous forward and inverse transformations between every pair of levels in order to couple the various representations at the different levels to define a single coherent perceptual state.'[13]

Just as Koffka had argued that soap bubbles arise out of the simultaneous acts of an undefined number of local forces, with the final shape emerging as a result of dynamic pressure between these forces, Lehar argued that the 'visual hierarchy defines a coupled dynamic system whose equilibrium state represents a balance or dynamic compromise between constraints experienced at all levels simultaneously, as suggested by Gestalt theory.'[14]

The originality of Lehar's work has been questioned, with similarities being pointed out to Grossberg & Mingolla's (1985) neural network model.[15] Despite this, the importance of Lehar's work is two-fold. Firstly, it demonstrates the contested nature of Gestalt theory itself today. More importantly, it presents us with an example of creating complicated abstractions seeking to process visual data while being freed from the constraints of constantly trying to work within the parameters that our (limited) knowledge of our own visual neural network system sets.

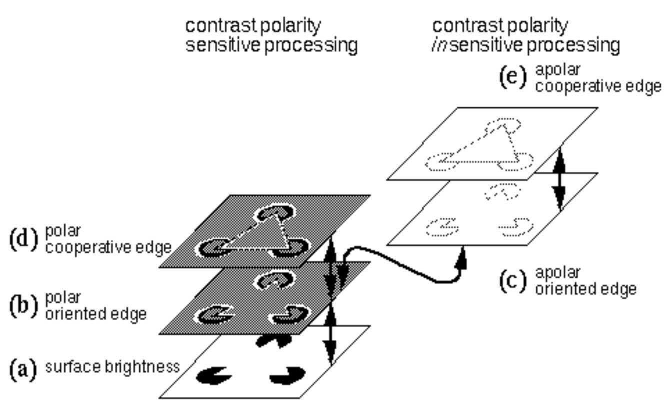

Fig.3. A visual representation of the two way processes with MLRF

[13] ibid.
[14] ibid.
[15] http://cns-alumni.bu.edu/pub/slehar/webstuff/orivar/reject3-A.html

Conclusion

The degree to which Lehar's model more accurately reflects the internal workings of organic neural networks than other more linear models is in one sense unimportant. The question remains as to what purpose computer scientists should use Gestalt theory? It has arguably become clear that creating a computer system capable of comprehensively mimicking our own visual cognition is technically extremely challenging, and perhaps ultimately unattainable due to the gaps in our knowledge of how our own eyes work and on the other hand due to limits in computer processing power.

As one author has argued, the success of our own neural networks at processing visual data would at first suggest that we should follow a 'neuronal approach' when seeking to build successful computer visual processing systems. However, as they bluntly state, this would be possible *'if only we knew how the brain does it.'*[16]

It could therefore be argued that attempting to recreate our own neural networks within computer systems is perhaps a misdirection of our efforts. As Florentin Worgotter *et al* have argued, 'one must realize that it does not make sense to try to create a carbon-copy of the existing natural visual systems.' It would be more fruitful for academics to 'arrive at abstractions which are better suited for implementation in a technical system.'[17] The *Speed-FX* plug in and Lehar's Multi-Level Reciprocal Feedback model demonstrate this spirit in very different ways.

[16] Florentin Worgotter et al, *Early cognitive vision*, p.294.
[17] ibid, p.316.

Bibliography

- J. Atkinson, *The Developing Visual Brain* (Oxford: Oxford University Press, 2000)
- D. W. Hamlyn, *Psychological Explanation and the Gestalt Hypothesis*, Mind, Vol. 60, No.240 (Oct. 1951). p.506-520
- G. W. Humphreys and V. Bruce, *Visual Cognition* (Hove: Lawrence Erlbaum Associates Ltd, 1991)
- K. Koffka, *Perception: An introduction to the Gestalt-theorie* [sic] (http://psy.ed.asu.edu/%7Eclassics/Koffka/Perception/perception.htm).
- K. Koffka, *Principles of Gestalt Psychology* (http://www.marxists.org/reference/subject/philosophy/works/ge/koffka.htm)
- A. Mack and Irvin Rock, *Inattentional Blindness* (Massachusetts Institute of Technology, 1998)
- S. Lehar, *Computational Implications of Gestalt Theory: The Role of Feedback in Visual Process* (http://cns-alumni.bu.edu/pub/slehar/webstuff/orivar/orivar3.html)
- S. E. Palmer, *Vision Science* (Massachusetts Institute of Technology, 1999).
- M. Wertheimer, *Laws of Organisation in Perceptual Forms* (http://psy.ed.asu.edu/%7Eclassics/Wertheimer/Forms/forms.htm)
- Florentin Worgotter, Norbert Kruger, Nicolas Pugeault, Dirk Calow, Markus Lappe, Karl Pauwels, Marc Van Hulle, Soriva Tan and Alan Johnston, *Early cognitive vision: Using Gestalt-laws for task-dependent, active image-processing*, (http://www.pspc.dibe.unige.it/ecovision/pubs/papers/woe_etal2004_naturalcomputing.pdf).

No Author Specified

- *'Perceptual Grouping'* - http://www.techfak.uni-bielefeld.de/~posch/PG/grouping.html
- '*Speed-FX'* - http://www.tecn.upf.es/gti/speedfx/htmls/plug.htm.