Agnès Desolneux, Lionel Moisan & Jean-Michel Morel (2008): From Gestalt Theory to Image Analysis. A Probabilistic Approach. Interdisciplinary Applied Mathematics, Vol. 34, 276 pages, 130 illustrations. Price: 51,65 Euro. ISBN 978-0-387-72635-9.

For more than 50 years now, computer image processing has existed. A big part of that activity is dedicated to coding, storing, and transforming images. The important point is that the transformed image will be presented to the human eye and interpreted by the human brain. In this field, the theory of information and psychophysics are the leading forces. Another part of image processing deals with problems of imitating the human ability to interpret images – with visual perception. Surprisingly, that line of image processing did not use the most powerful tool of psychology of perception – Gestalt psychology - and (not surprisingly) it did not succeed. The best evidence of that failure is Google's image search: it is based on analysis of captions and surrounding text. So, the book under the title "From Gestalt theory to Image Analysis" by A. Desolneux, L. Moisan and J-M. Morel was in big demand.

To my taste, the authors' focus on crucial works in Gestalt psychology is absolutely adequate: they concentrate on Wertheimer and Metzger. What I appreciate most is the authors' position with regard to the fundamental question of perception: "What is the nature of visual stimuli?" I quote:

> "Common sense tells us that a figure could not arise just by chance: we are "sure" that this organization corresponds to an intention; **somebody drew a square** there and this is why we see it."

In other words, a human being intentionally created the image, sending a meaningful message to another human being. That point of view is in agreement with Bongard's *imitation principle* (Bongard 1970, Guberman 2007), which I completely accept: our perception treats a drawing as it was created by an intellectual agent ("somebody drew a square") and restores (reconstructs) the process of creation of the visual stimuli. The authors repeat that concept: "perception must obey general laws and principles, which we shall call principles of visual **reconstruction**".

GESTALT THEORY
© 2013 (ISSN 0170-057 X)
Vol. 35, No.2, 183-206

I. When I first open a book on visual perception or computer image processing, I usually start by looking at the figures: and I did just that. The number of images was less than I expected. But what really surprised me was the number of mathematical formulas – they occur in their hundreds! Integrals, differential equations and sophisticated mathematical signs occupy the pages of the book. Then I found the explicit proclamation:

> "Phenomenology-styled Gestalt principles have no direct mathematical translation. Yet, a mathematical analysis cannot leave psychologists in the dark. Visual perception can receive - up to a certain limit - a fully mathematical treatment".

Because of my own experience, I pricked up my ears. For 25 years I worked as a chief scientist in the Institute of Applied Mathematics – a top-level institute in the Russian Academy of Sciences. I was co-working on Artificial Intelligence problems (including image processing) with one of the most outstanding mathematicians of the XX century – Prof. I. Gelfand. I can witness that we never used in our work or in our papers a single mathematical formula. I can also witness that during my career I never read a paper in which mathematics helped resolve any problem in image interpretation. I can say the same about psychology in general.

It is an old and widespread idea that „in jeder besonderen Naturlehre nur so viel eigentliche Wissenschaft angetroffen werden könne, als darin Mathematik anzutreffen ist." (Immanuel Kant, *Metaphysische Anfangsgründe der Naturwissenschaft, Vorrede*). From that point of view, all sciences were divided into precise sciences and descriptive sciences. The first were declared "first grade" sciences (mathematics, physics, chemistry), the rest were declared "second grade" sciences (linguistics, biology, geology, psychology). As a result, a false demand appeared for decorating PhD theses and papers with needless mathematics. During my career, a couple of my friends asked me for help in adding mathematical chapters to their already complete work.

But some people think that in order to compare the quality of sciences one has to put them on one scale, and that is wrong. The "soft" sciences have their own destiny and goal – they develop languages adequate for describing natural objects (mountains, languages, cells, perception etc). In particular, perception psychology tries to find out how different people similarly describe identical visual objects and scenes.

The authors' aim for mathematization sounds like they see themselves as missionaries in an underdeveloped country. They explain their plan of salvation:

1. "No translation invariance principle is proposed in Gestalt theory, in contrast with signal and image analysis where it takes a central role".

"Translational principle" doesn't exist in Gestalt psychology because there is serious apprehension about the very existence of translation invariance in visual perception. I am aware of some articles on that matter. One is *"Some results on translation invariance in the human visual system"* (Nazir 1990), which states that the "Visual system does not apply a global transposition transformation to the retinal image to compensate for translations." The title of a second paper announces the finding: *"Limited translation invariance of human visual pattern recognition"* (Dill, M. 1998).

2. "Gestaltists ignored the mathematical definition of digital image and never used resolution as a precise concept".

The Gestaltists don't ignore the problems of the digital nature of computer images – they have none of these problems, nor others mentioned in the book (like the role of noise and blur).

3. "Most of grouping laws and principles, though having an obvious mathematical meaning, remained imprecise".

And here is an example of proposed "precise mathematical meaning": "The vicinity law applies when distance between objects is small enough with respect to the rest" (this refers to an image containing two clusters of dots).

Such a description of two clusters of dots makes no sense, for many reasons: 1) what is meant by "objects"? There are two kinds of objects in that image – dots and clusters - and because I learn the *proximity principle* not from that book, I suppose that they are dots; 2) "distance between objects" makes no sense until a particular pair of dots is chosen; 3) "the rest" is not defined until a pair of dots is chosen. As a matter of fact, two clusters of dots have been described many times. Usually it sounds like "distance to closest neighbor for each dot in a cluster is sufficiently smaller than distance between clusters". It is worth mentioning that this definition is not a correct definition for clusters (Guberman & Woitkowski 2002). But in the quoted sentence there is a more general inconsistency: "vicinity law" works not only on images containing two clusters but works on one cluster as well.

It is a pity that the book is populated with such imprecise and sometimes even senseless "definitions" and "explanations". Here are two more.

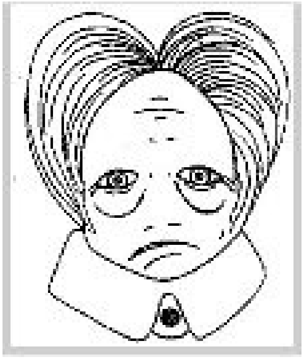
Quote: "All images are lacking any structure". As a matter of fact, all points of an image arrive in the most powerful of natural structures - Euclidian space. But in computer image processing it is a problem – each image is presented on a discreet rectangular grid of points (pixels) that don't keep the topology of Euclidian plane (i.e. each pixel doesn't know who are his neighbors; in every point it has to be calculated). But that is not a psychological problem; it is a problem of computer calculations.

Quote: "When we believe we see a straight line, the actual stimulus on our retina does not have much to do with the mathematical representation of a continuous, infinitely thin and straight stroke." As a matter of fact nobody pretends to see in an image of a short line the "mathematical representation of a continuous, infinitely thin and straight stroke". The majority of mankind knows nothing about mathematical abstractions. Geometry doesn't present a straight line as a rectangle which is infinitely thin. Euclidian geometry presents a line as an abstract object, which has only two features: 1) two lines intersect at one and only one point, and 2) a line contains at least two points. When one tries to illustrate the mathematical abstraction "straight line", one draws a straight line, which always has some thickness. To measure the thickness of a line, one has to draw a straight line crossing the line at a right angle. According to Euclidian postulates, two intersecting ideal lines have only one common point. So, that is the "thickness" of the line in popular description. But the psychology of perception doesn't deal with abstract lines.

The summary of the paragraphs above is that the idea of using mathematics as an unconditional benefit to Gestalt psychology is wrong. The problems that the authors intend to solve by use of mathematics are not problems of Gestalt psychology - they are problems of computer image processing. Despite the goal announced in the title – use of Gestalt psychology's findings to improve computer image processing – the book proposes a plan of dressing Gestalt psychology in fashionable image processing clothes with no visible benefits for Gestalt psychology.

II. The theoretical basis of the work presented in the book consists of four principles for Gestalt psychology: isotropy principle, Shannon sampling principle, Wertheimer contrast invariance principle, and – the most important - Helmholtz's principle.

1. Quote: "Isotropy principle requires image analysis to be invariant with respect to translations and rotations". I mentioned above that perception of an image sometimes depends on its position on the retina. One can add that in scenes containing a perspective (as on images of city streets), interpretation of an object's size depends on location in the perspective. But the demand of the rotation invariant contradicts what we know about perception. Last year in *Gestalt Theory Journal* was published a big paper on different perceptions of simple figures (like triangles and quadrangles) depending on their orientation on the plane (Pinna 2011). It is even more true for complicated stimuli (see fig.).



2. Quote: "Shannon-Nyquist's principle: Any image or signal, including noisy signals, is a band-limited function sampled on a bounded, periodic grid".

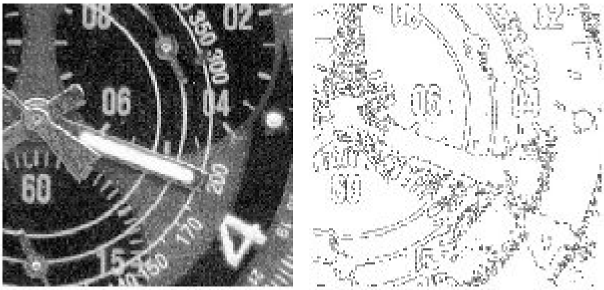
I can't imagine for which psychological problem knowledge of that fact could be useful, nor did I find in the book any evidence that the authors themselves know of such a problem.

The "accusation" of the authors against Gestalt psychology is that "Gestalt theory did not address two fundamental matters: image sampling and image information measurements". But image sampling is not a psychological problem. It is a problem of image processing and presentation. The illustration on that topic shows that changing the sampling can transform an image into another one with a different pattern, a different gestalt. (So, be careful when transforming images on a computer.) But that case doesn't create a problem in perception of these two patterns, nor in recognition of these two images on a computer as different ones. That means that it is not a problem of psychology of perception.

Information measurements, too, are far from perception problems. Shannon's concept of measuring information made remarkable progress in the theory of communications and coding. At the same time it was widely used in descriptive sciences like geology, biology, psychology, and medicine, without visible success. The first reason for the failure is that: to calculate the amount of information, one has to define a set of events and the distribution of probabilities on that set of events. Depending on the chosen set of events, the calculations give dramatically different numbers. At one time it was popular to calculate the information contained in our brain. First, the calculations were based on the hypothesis that the basic events are neurons with two stable conditions. Later, the set of all synapses were chosen as the basic set of events, and because each neuron has hundreds of synapses the amount of information increases by many orders of magnitude. The same problem exists for images. One can calculate the quantity of information in an image by taking as an event a point, but one can do it also by taking a stroke as a basic event, and receiving very different numbers.

The second reason why the majority of applications of Shannon's theory to "soft" sciences failed is that quantity of information is defined through entropy, which is defined on a set of *random* variables. It is not appropriate to treat a pixel in an image as a random variable.

3. Wertheimer's principle: "The level lines contain the whole shape information". Authors illustrate that statement with fig.



On the right image the isophots are marking only borders between black and white objects. As a result, white and black objects on the gray background are gone. The contrast between black and white objects becomes zero: it doesn't keep the contrast, and it has nothing to do with the *contrast invariance principle.* It illustrates a quite opposite principle – for correct interpretation sometimes we don't need to know the levels of brightness or their ratios. The unconditional proof of that is the existence of pen and pencil linear drawings.

"Because of the same principle, we shall only retain the gradient orientation and not the modulus of gradient as relevant information in images. "This is absolutely wrong. Every image processing developer knows the dilemma of choosing the threshold for gradients' modulus: if the threshold is high, many borders of important objects in the image would be lost. If it is low, many noisy lines would be picked up, and the relevant information will be lost in noise. So, not to care about the modulus of gradient (i.e. to put the threshold to zero) as it is proposed in the book is a dead-end solution.

4. Helmholtz principle: "Gestalts are sets of points whose (geometric regular) spatial arrangement could not occur in noise". That statement is wrong from both a mathematical and a psychological point of view.

1. When we are talking about a Gestalt of an image we never mean the particular set of points. Take, for example, 10 points arranged along a straight line. If we look for the same Gestalt in a noise, we are looking not for 10 points located on a line of the same direction and separated by the same space. We are looking for a Gestalt, which is a generalization of the particular stimuli – for a line of dots in any direction and with any spacing.

2. A set of points is neither a Gestalt nor an image. A set of points on the plane can become an image after the points are colored differently than the background.

3. Any given set of points could occur in noise.

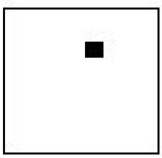
4. Face profile is not geometrically regular but we perceive it as a gestalt.

5. The probability that any given set of points will appear in a randomly distributed set of points (i.e. in noise) depends very much on the number of points in the given set and the number of points in the set presenting the noise. The bigger the given set and the smaller the noise set, the smaller is the probability that the given set will occur in the noise. If the number of points in the noise is infinite, the probability that the noise contains the given configuration is "1.0". By the way, it is worth mentioning that probabilities of the occurrence in a noise of a geometrically regular set of ***n*** points and a random set of ***n*** points are equal.

6. The crucial element of that statement is the verb "occur". Use of "occur" means that it pretends to be a mathematical statement: God can check if a given configuration occurs in the noise but we are talking about psychological principles, principles of perception. The psychological question has to be: could a human being recognize that configuration in noise. An appropriate experiment was performed thousands of times: to a given image was added noise (randomly distributed points) in increasing amounts. At some moment the human eye fails to recognize the initial image despite the fact that the image is still there. So, did the given arrangement of points occur in the noise? Definitely "Yes". Did the human eye perceive the appropriate Gestalt? Definitely "No".

III. Despite my disappointment with the introductory part of the book, I know that sometimes when authors turn to solutions of practical problems, they forget ideological constraints and deliver good solutions (it was particularly true in the Soviet Union, where any publication had to be started with a panegyric to Marxist ideology). I decided to go through the figures and images, which have to present the essence of the work on human perception. My comments on some of them follow.

1. A black square on a white background.



a) The enthusiast is sure that he sees a square and that its existence can be proven by probabilistic arguments. The skeptic responds: "You think you see a square; but, all I see is a set of white or black pixels. They just fell together by chance and built this square just by chance."

This perplexing statement raises a number of questions. The "enthusiast" sees a square.

Why does he need to prove that he sees a square? We are discussing problems of psychology of perception, and we are interested in the subject's report only (look at the papers of Wertheimer, Koehler, Metzger). The "skeptic" sees a square as well ("pixels built this square"), so they dispute not the Gestalt (both see the same) but the origin of the stimulus. The "enthusiast" might be sure that the square has been created intentionally by somebody, the "skeptic" might state that it happens by chance. But that is not a problem of perception.

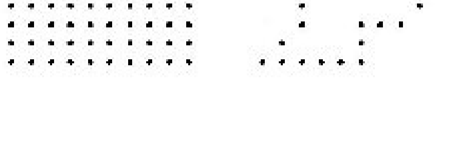
b) A square alone is a global gestalt, but it is the synthesis of a long list of concurring local groupings, leading to parts of the square endowed with some gestalt quality. Such parts we shall call partial gestalts. The sides and corners of the square are therefore partial gestalts.

Visual stimulus (like "square alone" or "corner" or "side") is not a Gestalt. To the contrary, Gestalt is a product of our brain, and it is always a generalization of a single image.

c) Description of a square: "parallelism between opposite sides, orthogonality between adjacent sides, and constant **width** (? SG) of both pairs of **opposite** sides".

As a matter of fact, it is a description of a rectangle: in a square not only opposite sides have the same length, but all sides have equal length (of course, "length" – not "width" as in the quotation above).

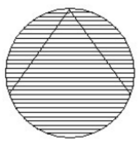
2. Quote: "According to Kanizsa's paradox, the figure on the right is potentially present in the figure on the left, and would indeed appear if we colored the corresponding dots. This illustrates the fact that the figure on the left contains a huge number of possible different shapes".



I agree that the figure will appear "if we colored the corresponding dots". To me it means that it will appear if someone would like to send a message presented by that shape. And because the left figure doesn't contain any message it doesn't contain any shape. If the messenger has a pen and colored ink, he can draw any figure: why does he need the matrix of black dots? He can do it on a white sheet. Are we willing to seriously discuss the proposition that a sheet of paper potentially contains all possible drawings? The entire problem is translated from the computer domain. Images are displayed on screens covered with a rectangular grid of dots. One can say that the screen potentially contains all possible images if(!) we colored the corresponding dots in corresponding colors (what our monitors do 30 times a second). That is the truth, but is it worth mentioning in a psychological context? Once more the authors substitute psychological problems with computer ones.

3. Quote: "Masking by embedding in a texture. The basis of the triangle becomes invisible as it is embedded in a group of parallel lines."

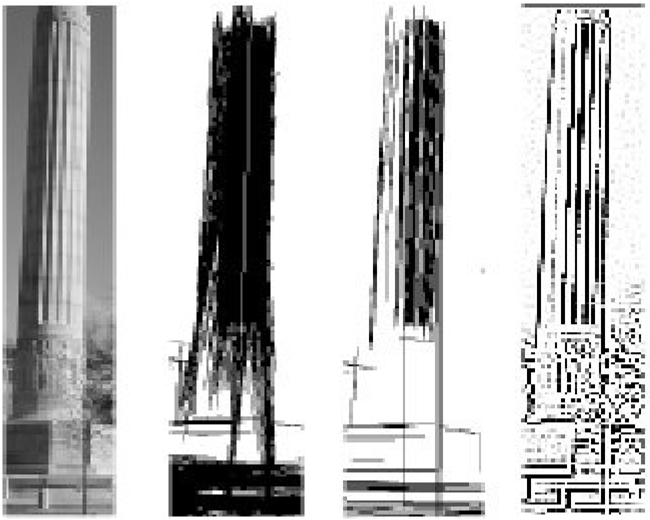
The correct explanation comes from the very essence of Gestalt. According to Bongard's imitation principle, our perception imitates the way the stimulus was created. When subjects reproduce



this drawing in experiments they draw a circle, then a set of parallel lines, and then an angle. The triangle in question has never appeared in redrawing experiments. So, the perception of that figure will always be "circle, parallel lines, and angle".

Many illustrations in the book refer to the alignment problem – how to find straight lines in a digital image? That problem and its solution were presented with no connections to Gestalt psychology. During the last 50 years, hundreds of algorithms were developed for that task, but that described in the book is the most loaded with mathematics: 105 formulas and equations, 4 theorems, 4 lemmas, 6 definitions, 5 corollary, 2 conjunctions, 13 propositions, 7 algorithms, and 1 principle in 40 pages of text and figures. Some of the equations are integro-differential ones, one equation occupies 3 lines. It was interesting to look at the results. Here is one of them. The first figure is the initial photo, the second one is an intermediate result, the next figure shows the final result after cleaning the previous one, and the last shows the **gradient** calculated by a 20-years-old standard image processor PaintShopPro. Conclusion: the mountain gives birth to a mouse. By the way, the first formula of 105 dedicated to the alignment problem was the formula for calculating the **gradient**.

But what is a shame concerning this finding of straight lines is the complete ignorance of Gestalt psychology principles. One of the basic principles of Gestalt psychology is the holistic approach to perception: interpretation of a part of an image depends on its surroundings. The algorithms proposed in the book for finding straight lines are local – they don't take into account what is in the image around that line. In reality, when one recognizes an urban scene (by recognizing cars, road, pedestrians) one will identify vertical lines even in big noise. When a straight line in any direction is recognized, other lines parallel to it will be picked up even when poorly presented (as on the photo below).



IV. The authors found Gestalt psychology guilty of not using mathematics before the case was presented to the scientific community. After analyzing the case, I find that it has to be dismissed.

Shelia Guberman

References

Bongard, Mikhail M. (1970): *Pattern Recognition.* New York: Spartan Books.

Dill, M. & Fahle, M. (1998): Limited translation invariance of human visual pattern recognition. *Percept Psychophys. 60(1),* 65-81.

Guberman, S. & Woitkowski, W. (2002): Clustering Analysis as a Gestalt Problem. *Gestalt Theory, 24 (2),* 143 - 158.

Guberman S. (2007): Algorithmic Analysis of "Good Continuation" Principle. *Gestalt Theory, 29(2),* 148- 168.

Nazir T. & O'Regan J. (1990): Some results on translation invariance in the human visual system. *Spat Vis. 5(2),* 81-100.
Pinna, B. (2011): What is the Meaning of Shape? *Gestalt Theory 33* (3/4), 383-422.

Shelia Guberman, PhD in Nuclear Physics, PhD in Computer Science. Former Chief Scientist at Institute of Applied Mathematics (Moscow). He is author of first application of Pattern Recognition to technical problems (oil exploration, medical diagnosis, earthquake prediction), of D-waves theory (seismicity), theory of similarity in Nuclear Physics, handwriting recognition technology. He published a number of papers on algorithmical analysis of Gestalt theory and its application to Artificial Intelligence.
Address: PO Box 2411, Cupertino, CA, 95015, USA.