SMCia/07
2007 Three-Rivers Workshop on Soft Computing in Industrial Applications
University of Passau, Germany, August 1 - 3, 2007

Gestalt Theory in Image Processing: A discussion paper

Mario Köppen, Kaori Yoshida
Kyushu Institute of Technology,
680-4, Kawazu, Iizuka, Fukuoka 820-8502 Japan
mkoeppen@ieee.org,yoshida@ai.kyutech.ac.jp

Pablo A. Valle
Technical University Berlin
Strasse des 17. Juni 135, 10623 Berlin, Germany
valepnbb@mailbox.tu-berlin.de

1 Introduction

The apparent phenomenon of Gestalt laws in visual perception has stimulated a steady flow of research contributions in the image processing and pattern recognition research community. In this paper, we want to provide a critical review of a representative number of such contributions and point out some issues that have not been reflected well in all these past works.

A rough estimate gives a few hundert research contributions on the topic of Gestalt (theory, psychology, laws) in computer science[1]. The recent number of publications is about twice as large as 20-30 years ago, which is conform with the general increase of the number of research publications. Thus, there seems to be a small, but rather constant share of papers devoted to the general theme Gestalt. It is also notable that mostly these contributions are isolated ones, with few amount of referencing between them. So far, we could not witness any generally accepted means of introducing Gestalt in image processing.

The comparable low amount of research devoted to the direct linking between Gestalt theory and image processing does not entail a low practical impact. There are many potential industrial applications waiting for any more thorough approach to the Gestalt theme. The most obvious one is gaining a deeper insight into the functionality of human (or higher animal) vision, which has been the pre-cursor of any image processing related development so far. But also, it has a strong relation to any design study. There are iconic signs in use all over the world (issuing warnings, guidance, information), which have to be designed in a manner to be easily comprehensible in their meaning by a human observer (and often from a larger distance, within a short time, and while being only partially visible). These signs often directly employ Gestalt concepts, but this is usually the result of the considerations of an experienced graphic designer, that means, a human. At present, there is no test known for the suitability of a, for example, warning sign that does not involve a human observer. Same for specific, content-related designs, like a webpage: the guidance of a webpage visitor's attention is clearly following Gestalt laws, but after publishing a webpage, any measure for its content delivery will involve a human. In a similar manner, product design, advertisement banner, human-machine computer interfaces and the more are all referring to Gestalt concepts in their design today (and there are much more Gestalt laws appearing in an urbanized environment than in the free nature), but the effectiveness of fulfilling their purpose can only be valued by the human, and not a machine.

In the following, we will discuss these issues more deeply, especially focussing on the fact that there is barely a visual scene in the real world, where only one Gestalt law is active. We will present some examples in section 2, and then consider the relevance of Gestalt theory related publications to these examples in section 3. Then, we will conclude with a short comment about the suitability of soft computing methodologies for Gestalt-related processing, which has not been considered much in the past.

[1]For example, the SCOPUS lists about 200 relevant publications for the past 30 years.

2 Joint effect of multiple Gestalt laws

The situation that was stated in the introduction is somehow unsatisfying, as there is much evidence for the interplay of Gestalt laws in a manner that can hardly be based on the processing of primary sensorial input alone (for example, in a computer program). The most apparent and well-known phenomenon in Gestalt theory is the perceptual grouping, which is a cognitive aspect of a visual scene that is immediately perceived, but obviously not part of the raw sensorial input of the visual system. Figure 1 is intended to give a guidance to the complex interplay of different Gestalt laws in a simple visual scene.

Sub-figure (a) shows a circle that is perceived as being in front of the character string "ABCDEF." The point of interest here is that the circle, while obviously partially covering the characters "C" and "E" (and likewise a hidden "D", but this is a kind of inferencing from the scene that is not of interest here) is perceived as being in front of the character "A" as well. By itself, this is a remarkable fact, since there is no directly evaluable evidence in the image that directly relates the segment comprising the character "A" to the circle. Gestalt theory explains this by the perceptual grouping of the character string. Since the circle is covering a part of the perceptual group, it is also in front of the group as a whole and of any of its members. Both, the group relation and the foreground-background relation are perceived at the same time, and also at the same time linked together in the inference given above ("the circle is before the A"). By pure will only, we are not able to see something different.

The other sub-figures show variations of this theme, and thus also some dependencies of the phenomenen from the constituting parts of the figure. In sub-figure (b), there is no perception of a circle at all, but by some bordering hints, the perception of a virtual contour of a circle can be triggered (sub-figure (d)), and then, the circle is in front of the "A" again. Sub-figure (c) shows the opposite situation: the group is there, but the circle is not covering a part of it (and we also do not perceive any hidden character behind the circle). Thus, circle and "A" appear to be on the same depth level. Even if there are two perceptual groups, a kind of meta-grouping makes "A" appearing behind the circle in sub-figure (e) again. Below, sub-figure (g), provides an ambiguous situation. Here, there can be the perception of a black circle in front of the outlined "A", but the perception of a "punch hole" is also possible. Finally, sub-figures (f) and (h) demonstrate this phenomenon to be independent of the size of the circle.

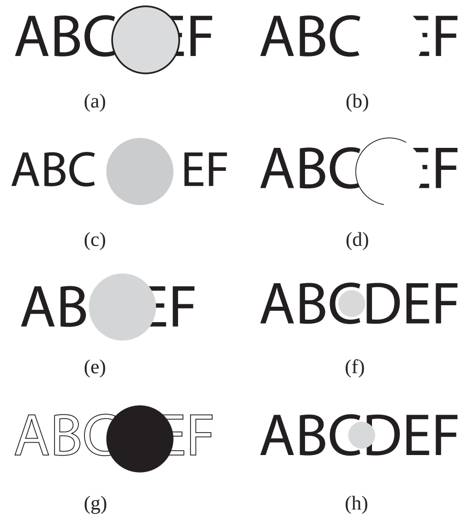

Figure 1: Foreground-background separation of perceptual groups.

So, the question that comes up is if there is any substantial contribution to the understanding of such phenomena, making it appearing as particular features and aspects of a more general model of visual perception. The answer will be negative, as we mostly find all such studies in relation to isolated aspects of Gestalt laws. The simple interplay of Gestalt laws, as apparent in fig. 1 is not reflected at all.

3 A survey on using Gestalt in image processing

To demonstrate this, we are considering a relevant cross-section of about twenty contributions, to solicit general ideas of researchers in relation to the concept of Gestalt. In summary, some of the relevant features of these contributions are as follows:

Methodology Due to the interest of providing a computer-based evaluation of images, it is no wonder that nearly all papers are either presenting algorithms [Cao03, GLM97, HB05, WKSW05, CL96, TOS93, DMM03] or (often hierarchical) frameworks [ZTB04, WG99, Sar03, CN89, AA99] to incorporate Gestalt laws in the computational processing flow. A smaller number also considers the relevance of computational models for the biological modelling of Gestalt perception [Ros89, eGZW03, CN89, APE99, BSZ05]. Few papers focus on the real-world application of such algorithms, models or frameworks [Sai99, GLM97]. In [AA99, WG99], the interest is in the provision of features for image retrieval. In general, all these works neglect the interplay between Gestalt laws in the same scene. A single algorithm may capture a virtual boundary like the circular contour in fig. 1(d), but it will not come to the point where this circle appears to be in the foreground of a totally disconnected part of the same scene. It should also be noted that such intermediate processing steps of an algorithm or framework can only be of functional nature. As human, we always *see* the scene and are not aware of any intermediate or hidden processing.

Referenced Gestalt laws Among the various stated Gestalt laws, nearly all reference is going to perceptual grouping and (more behind in number) good continuation. Only in a few cases, other Gestalt laws are also considered (like symmetry in [APE99] and closure in [CL96]). We never find the interplay of more than one Gestalt law, as a simple example like the one given in fig. 1 would need for its explanation.

Studied image structure The extraction and completion of boundaries and edges in images seems to gain highest interest for applying Gestalt laws [Ros89, Cao03, Sai99, GLM97, WKSW05, CL96, TOS93]. A smaller number of works is focussing on segmentation approaches [Ros89, RW90, WG99], and from the references some overlap can also be seen. Here, Gestalt laws are mostly considered a source of inspiration for interesting and new algorithms (which is not a wrong attitude *per se*).

Visual concepts Only in a few cases, the presentation is accompanied by a concept of visual processing. Notable works here are focussing on visual primitives [Sar98] and salient features or salient boundaries [Sar98, WKSW05]. Both, the meaning of primitive and the meaning of salient remains more or less self-evident and unspecified. Incorporating (or one can also say forcing) the Gestalt laws into the standard feature classification approach of pattern recognition has been done in [AA99, Zhu99, Sar98]. Regarding the latter one, and glancing on figure 1, it seems very unlikely that the only effect of Gestalt perception is the solicitation of some set of features against other features. These works do much more remind on a "cargo cult" (Feynman) and are purely pragmatic - in the sense of just taking an inspiration from evident phenomena instead of seeking an explanation.

Probability A remarkable large number of contributions makes reference to probabilistic concepts. This is done explicitely in [Cao03, eGZW03, DMM03], and with reference to probability distributions in [RW90, Zhu99]. Relating Gestalt perception with (im)probability of occurrences of certain visual phenomena [DMM03] does not look like a promising approach. A simple argument against this can be found in [Sar03]: in the perceptual grouping, the actual number of elements of a group does not matter. In figure 1 (a) there is a perceptual group of five characters, in sub-figure (e) a similar one of four elements. However, for computing probabilities actually it does matter whether having four or five, or twenty elements.

Semantics We find few explicit reference to the question of semantics (in the sense of a "visual grammar"), the bridging of the semantic gap between primary data and content, and if Gestalt laws can be of help here [ZTB04]. Obviously, there is a kind of semantic in the effects of Gestalt law. As indicated by fig. 1, the inferencing of "A" being behind the circle is a semantical one, as we can use the relation between abstract concepts for its formulation.

Special functions We found only one paper relating Gestalt theory and Gabor functions [Zhu99]. As one of the major instruments for the modelling of perception and attention mechanisms in neuroscience, this is a little bit surprising. One reason could be the stronger focus on boundary processing, where the field aspects of a Gabor function are not so important.

4 Soft Computing and Gestalt theory

Only one of the selected contributions was considering a soft computing technique for the handling of Gestalt theory based concepts [Zhu99]. But also there, it was just the use of a Genetic Algorithm for tracking an accompanying optimization problem. We may conclude on a rather minor importance of soft computing for establishing computer models of Gestalt theory.

One of the main characteristic of soft computing methods is their ability to deal with imprecision. Basically, we may find this suitable for the given problem instance. For example, perceptual grouping may happen also if the group elements are not perfectly equal. But also other characteristics of soft computing methods are fitting with the circumstances of Gestalt perception and "evauation." Soft computing offers a vast variability of algorithms and their hybrid combinations, which may act together in collaboration and not competition. It offers the potential for hybrid and intelligent system design as well, and can thus overcome the "single algorithm approach" that we felt unsuitable for the handling of the given Gestalt related problems. Moreover, the visual perception of Gestalt is clearly based on the functionality of the brain (despite of the lack of knowledge of any Gestalt-related neuron processing so far), but the brain also gave raise to one of the main soft computing methodologies, the neural networks. Also several other soft computing methods are based on group processing, like the individuals of a population of an evolutionary algorithm. So, there should be not much doubt on the claim that soft computing can provide something to Gestalt-based image processing as well. What can not be seen easily is, whether the soft computing methods and their underlying paradigm of computational intelligence, and the "real" intelligence will ever converge.

It is not the goal of this communication to add another algorithm or framework to the already exisiting ones, as we feel supposedly still far away from a sufficient insight into the nature of all the Gestalt related phenomena and their potential use in the processing of technically acquired images. Only one vague and up to now unverified point should be mentioned at last, related to some recent works in neurophysiology. Especially one model, not being free of critique of other researchers, could be of some interest. We remind on the theory of Edelman and Tonini, mostly known as "neural darwinism" [ET01]. One of the main claims of this theory is that neural processing, in general, is selectively and not representationally organized. The so-called reentry mechanism gives an additional explanation for the obviously immediate and stable nature of neural processing. Reentry refers to rapid and recurrent linkings between remote groups of activated neurons in the thalamo-cortical system, strongly connecting parts of the cortex being as different as neural groups for planing, or proprioception with parts responsible for object recognition, language or sensorial processing. So far, the very first models of reentry and neural group selection studied by this school (including the works on the robot DARWIN) did not find much interest in the computer science community, but it is not so hard to see their potential value for establishing new neural architecture far beyond the multi-layer perceptron and including selectional techniques as they are very common to evolutionary compuation.

5 Summary

A number of contributions on Gestalt theory in image processing is missing the point that Gestalt laws do actually interact, as can be seen from the interplay of foreground-background separation of perceptual groups. It seems to need a selective and recurrent mechanism instead of a flow of a number of processing steps or levels to approach the computational modelling of this interplay.

Acknowledgment

This research was partially supported by the Japan Society for the Promotion of Science, Grant-in-Aid for JSPS Fellows, 18.06094. One author has also been supported by the Conacyt Mexico.

References

[AA99] S. Alwis and J. Austin. An integrated framework for trademark image retrieval using gestalt features and cmm neural network. In *Image Processing And Its Applications, 1999. Seventh International Conference on*, volume 1, pages 290–295, 1999.

[APE99] A. Assadi, S. Palmer, and H. Eghbalnia. Learning gestalt of surfaces in natural scenes. In *Neural Networks for Signal Processing IX,* *1999. Proceedings of the 1999 IEEE Signal Processing Society Workshop*, pages 380–389, 1999.

[BSZ05] Ohad Ben-Shahar and Steven W. Zucker. Continuation of general 2d visual features: Dual harmonic models and computational inference. In *Computer Vision, Tenth IEEE International Conference on*, volume 2, pages 1643–1650, 2005.

[Cao03] F. Cao. Good continuations in digital image level lines. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, volume 1, pages 440–448, 2003.

[CL96] Yung-Sheng Chen and Tsay-Der Lin. Simulation of closure process for line patterns. In *Pattern Recognition, 1996., Proceedings of the 13th International Conference on*, volume 2, pages 215–219, 1996.

[CN89] R.W. Conners and C.T. Ng. Developing a quantitative model of human preattentive vision. *IEEE Transactions on Systems, Man and Cybernetics*, 19(6):1384–1407, 1989.

[DMM03] A. Desolneux, L. Moisan, and J.-M. More. A grouping principle and four applications. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(4):508–513, 2003.

[eGZW03] Cheng en Guo, Song-Chun Zhu, and Ying Nian Wu. Towards a mathematical theory of primal sketch and sketchability. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pages 1228–1235, 2003.

[ET01] Gerald M. Edelman and Giulio Tonini. *A Universe of Consciousness: How Matter Becomes Imagination*. Basic Books, 2001.

[GLM97] P. Gamba, M. Lilla, and A. Mecocci. Extraction of discontinuous chains of symbols by means of perceptual grouping. In *Image Processing, 1997. Proceedings., International Conference on*, volume 2, pages 422–425, 1997.

[HB05] J. Huart and P. Bertolino. Similarity-based and perception-based image segmentation. In *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, volume 3, pages 1148–1151, 2005.

[Ros89] A. Rosenfeld. Computer vision: a source of models for biological visual processes? *Biomedical Engineering, IEEE Transactions on*, 36(1):93–96, 1989.

[RW90] T.R. Reed and H. Wechsler. Segmentation of textured images and gestalt organization using spatial/spatial-frequency representations. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 12(1):1–12, 1990.

[Sai99] F. Saitoh. Boundary extraction of brightness unevenness on lcd display using genetic algorithm based on perceptive grouping factors. In *Image Processing, 1999. ICIP 99. Proceedings. 1999 International Conference on*, volume 2, pages 308–312, 1999.

[Sar98] S. Sarkar. Learning to form large groups of salient image features. In *Computer Vision and Pattern Recognition, 1998. Proceedings. 1998 IEEE Computer Society Conference on*, pages 780–786, 1998.

[Sar03] Sudeep Sarkar. An introduction to perceptual organization. In *Integration of Knowledge Intensive Multi-Agent Systems, 2003. International Conference on*, pages 330–335, 2003.

[TOS93] M. Teranishi, N. Ohnishi, and N. Sugie. Subjective contours are useful for extracting contours with very weak contrasts. In *Neural Networks, 1993. IJCNN '93-Nagoya. Proceedings of 1993 International Joint Conference on*, volume 1, pages 139–142, 1993.

[WG99] A. Wardhani and R. Gonzalez. Perceptual grouping of natural images for cbir. In *Signal Processing and Its Applications, 1999. ISSPA99. Proceedings of the Fifth International Symposium on*, volume 2, pages 923–926, 1999.

[WKSW05] Song Wang, Toshiro Kubota, Jeffrey Mark Siskind, and Jun Wang. Salient closed boundary extraction with ratio contour. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(4):546–561, 2005.

[Zhu99] Song-Chun Zhu. Embedding gestalt laws in markov random fields. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 21(11):1170–1187, 1999.

[ZTB04] N. Zlatoff, B. Tellez, and A. Baskurt. Image understanding and scene models: a generic framework integrating domain knowledge and gestalt theory. In *Image Processing, 2004. ICIP '04. 2004 International Conference on*, volume 4, pages 2355–2358, 2004.