Quantitative description of the laws of perceptual grouping by means of the principle of representational minimum description length

A. S. Potapov

NPK S. I. Vavilov State Optical Institute, St. Petersburg

V. G. Petrochenko

St. Petersburg State University of Information Technologies, Mechanics, and Optics, St. Petersburg

(Submitted February 29, 2008)
Opticheskiĭ Zhurnal **75**, 42–47 (August 2008)

This paper discusses the problem of quantitatively describing the laws of perceptive grouping developed in Gestalt psychology and their use in computer vision. For a unified description of the Gestalt laws, the principle of representational minimum description length is proposed. Psychophysical experiments have been carried out in which it is established that the probability that a person will distinguish a stimulus depends predominantly on the information content of the stimulus, expressed in the reduction of the description length (in the framework of the given representation) of the elements of the visual scene when the stimulus is described separately from the field elements. A computer model is constructed and experimentally checked that groups the structural elements of the image, based on an informational criterion and a representation that embodies some of the Gestalt laws. © *2008 Optical Society of America*.

INTRODUCTION

One of the classical questions in the investigation of problems of both computer vision and human visual perception is the question of how to achieve complete perception of images from the local information contained in the individual pixels.[1] In the psychology of perception, this question was investigated in the framework of Gestalt theory, in which a series of "laws" of perceptive grouping of the elements of the field of the human visual system was detected. However, these laws have no quantitative or still less algorithmic character, and therefore little attention has been paid to them for a long time in the area of computer vision.[2]

Attempts have recently been made to connect the results obtained in Gestalt theory and in computer vision—in particular, in the framework of theoretical-informational,[1] neural-network,[2] and statistical[3] approaches. For Gestalt psychology, this can give the possibility of quantitative improvement of the laws of perceptive grouping and the possibility of additional experimental checking by means of computer modelling and, in the area of computer vision, can serve as the development of new ideas and biologically based algorithms for analyzing images.

A promising approach both to the problem of quantitatively expressing Gestalt laws[1] and to the problem of interpreting images in computer-vision systems[4] is the theoretical-informational approach based on the principle of minimum description length (MDL). According to this principle, the best model of the available data is the model for which the sum of the description length of the data by means of a model and the description length of the model itself is minimized.[5]

However, this principle does not rigorously allow the criteria of perceptive grouping to be deduced without resorting to additional heuristic considerations. This paper proposes an extension of this principle, taking into account the concept of image representation—the principle of representational minimum description length (RMDL). Psychological experiments have been set up that show that the efficiency of the grouping by the human visual system predominantly depends on the information content of the stimulus in the framework of the given representation (at least for the types of stimuli used in the experiment). The problem of algorithmization of the laws of perceptive grouping is thus connected not so much with the search for adequate quantitative grouping criteria as with the search for the optimum representations of images.

THE MAIN CONCEPTS OF GESTALT THEORY

Gestalt theory was developed in the first half of the Twentieth Century as an attempt to explain the features of human perception. Little could be known at that time concerning the organization of the brain, and Gestalt psychologists attempted to give a thermodynamic interpretation of the mechanisms of perception, assuming that the "perceptive field" is reorganized under the action of certain forces into the state with the minimum energy (i.e., possessing the greatest symmetry). Although such an interpretation was never confirmed, it made it possible to accumulate a large amount of empirical material, in particular, on the study of visual perception, and to form a number of laws of perceptive grouping.

The law of grouping by symmetry can serve as an example of a grouping law, according to which similar or like

1070-9762/2008/080509-05$15.00

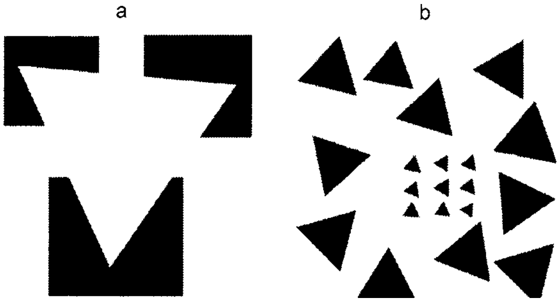

FIG. 1. Examples of images that illustrate the laws of grouping by "good continuation" (a) and similarity (b).

elements of the visual field have a tendency to combine into groups. Other grouping laws include grouping by closeness, "good extension" (elements that lie on a certain regular curve, for example, a line, are combined), closure of shape, etc. Figure 1 presents typical examples that show the action of Gestalt laws.

The indicated laws are the result of generalization of the observed features of human perception. As a rule, each law relates to one attribute: shape, size, direction, etc. However, the Gestalt laws are insufficiently concrete, so that there are no clear boundaries in which one law or the other acts, and the problem arises of resolving conflicts between the laws.[1] When several laws are applicable that give different grouping results, Gestalt theory cannot reliably predict how a person will perceive an image (see Fig. 2).

Moreover, it is not understood in the framework of Gestalt theory why it is these laws that are used in the process of perception, and whether they are optimal and compulsory when computer-vision systems are implemented. To answer these questions, we consider the principle of RMDL.

THE PRINCIPLE OF REPRESENTATIONAL MINIMUM DESCRIPTION LENGTH

According to the principle of MDL, the model m^* that best describes a certain set of data f is the model that makes it possible to minimize the sum of the description length $L(m)$ of the model and the description length $L(f|m)$ of the data in terms of the model:

$$m^* = \arg\min(L(m) + L(f|m)). \tag{1}$$

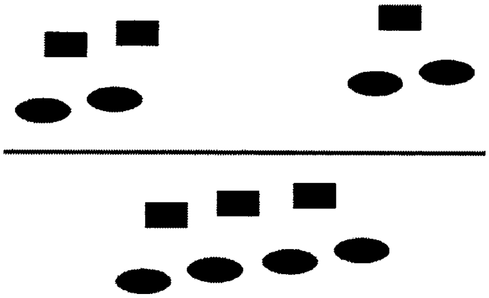

FIG. 2. Example of conflict of the laws of grouping by similarity and good continuation; Gestalt theory does not predict at what distance regrouping occurs between the elements.

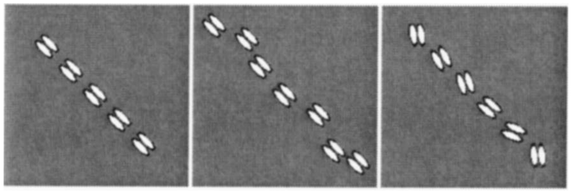

FIG. 3. Examples of stimuli with different values of the parameters (number of elements, errors of position, errors of orientation).

The algorithmic complexity of the lines of the symbols that correspond to data f and model m appears as the description lengths L in theoretical work. In this case, the set of programs for a Turing machine appears as the model space, while the data f contain all the available information.

In solving image-analysis problems, the images need to be independently interpreted, with certain *a priori* information being used to describe each image. Under these conditions, it is necessary to refine the MDL principle as the principle of representational MDL including two positions, considered below.[6]

1. Model m^* that best describes some set of data f in terms of representation S is the model that minimizes the sum

$$L_S(f,m) = L_S(m) + L_S(f|m). \tag{2}$$

Here L_S is the description length in terms of the given representation, which can formally be defined as $L_S(x) = L(Sx)$, where $L(Sx)$ is the algorithmic complexity of the concatenation of the lines of S and x.

The best representation for the given selection of images $F = \{f_1, \ldots, f_n\}$ is the representation S for which the sum of the length $L(S)$ of the representation and the sum of the lengths of the descriptions of images

$$\sum_{i=1}^{n} MDL_S(f_i),$$

is minimized, where $MDL_S(f_i) = \min_m L_S(f,m)$.

Representation S is information that is common to all images of selection F. Choosing the optimal image representation is thus an empirical problem, the solution of which can be directed by averaging the description length over some selection.

A number of researchers have pointed out[1] that Gestalt laws can be expressed in terms of the description length. However, according to the RMDL principle, it is necessary to construct an adequate representation of the images in order to construct the informational criterion that quantitatively expresses the Gestalt laws and correctly predicts the results of the perception of specific stimuli.

A PSYCHOLOGICAL EXPERIMENT

To determine the characteristics of perceptive grouping in the visual system, we carried out the following experiment: A stimulus was presented to a subject (see Fig. 3), and then two images were presented (see Fig. 4), one of which

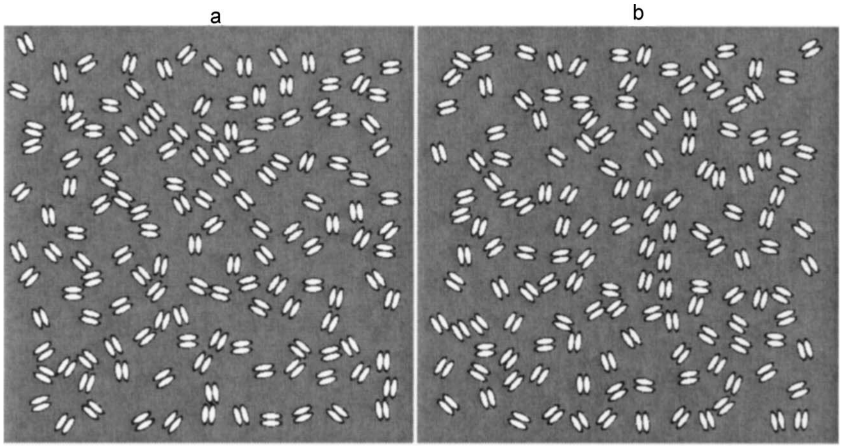

FIG. 4. Examples of images presented in the course of an experiment, one of which (a) contains the desired stimulus, while the other (b) does not.

contained the stimulus, while the other did not (i.e., it fulfilled the role of a distracter). Each of the images was presented for 1sec.

The stimuli varied in the number of elements, in the rms deviation of the elements from being positioned on a straight line, and in the variance of the orientation. For each set of stimulus-formation parameters, a series of experiments was carried out with the subjects, as a result of which the mean percent of the errors was established that was acceptable by people when they choose the image that contains a stimulus.

Let us consider the question of the representation of information concerning the elements on images of this type. Each element is described by a set of parameters x_i, y_i, φ_i, where x_i and y_i are the coordinates of the ith element, and φ_i is its orientation angle. If the elements do not break up into groups, the parameters of each of them are described independently. The coordinates of a certain element can be described by the relative position of the element closest to it, for which it is necessary to indicate its number (this requires $\log_2 N$ bits, where N is the total number of elements) and the relative coordinates (on the average, $\log_2 \sigma_r + C_1$ bits are required to indicate them, where σ_r is the mean distance between elements, and the constant C_1 is determined by the accuracy with which the coordinates are described. It is also required to describe the azimuth of the direction to the closest element; however, because the distribution of the azimuths is isotropic, this component of the description is neglected). Let the orientation angles of the field elements be uniformly distributed in the range $[-\varphi, \varphi]$. Then $N(\log_2 \varphi + C_2)$ bits are required to describe them (C_2 is a constant, determined by the accuracy of the description of the orientation). To describe the field elements that are not broken up into groups, we require

$$L_0 = N(\log_2 N + \log_2 \sigma_r + \log_2 \varphi + C_1 + C_2) \text{ bits}. \tag{3}$$

We separate out a group of n elements from the field elements. $L = (N-n)(\log_2 N + \log_2 \sigma_r + \log_2 \varphi + C_1 + C_1)$ bits are required to describe the remaining elements. Let it be known *a priori* that these elements are located on some curve with a definite step (as was the case in the experiment, see Fig. 3). The following representation of the information can then be proposed to describe the given group. It is indicated in the framework of the given representation precisely what elements are discriminated into the group, for which $n\log_2 N$ bits are required. The coordinates of the elements of the group are described through their deviation from the curve, and, for the average deviation σ_c, this requires $n(\log_2 \sigma_c + C_1)$ bits. Let the orientation angles of the elements of the group be uniformly distributed in the range $[-\psi, \psi]$. Their description then requires $n(\log_2 \psi + C_2)$ bits. The description of all the elements for the discriminated group requires

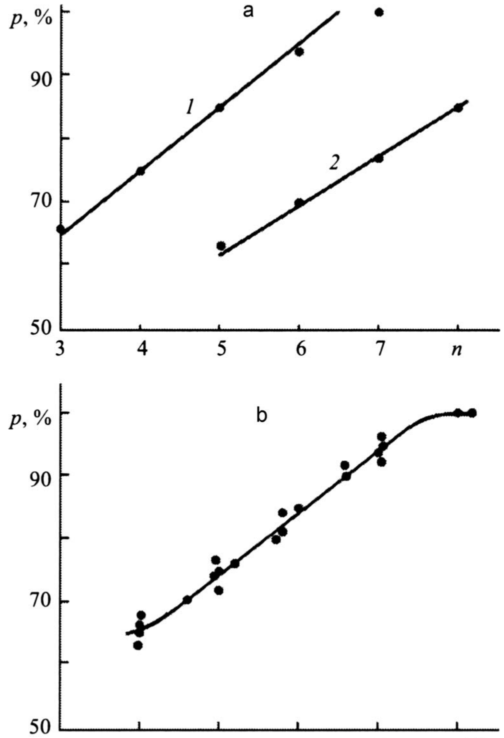

FIG. 5. Percent error p vs the number n of elements in the stimulus (a) and vs the information content ΔL of the stimulus (b) for various values of the other parameters of the stimulus. 1—$\psi = \pi/4$, 2—$\psi = \pi/2$.

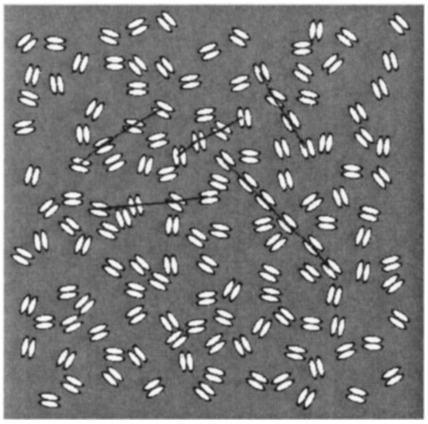

FIG. 6. Examples of automatic discrimination of a stimulus (as well as of several groups of regularly placed elements) from the field elements.

$$L_1 = (N - n)(\log_2 N + \log_2 \sigma_r + \log_2 \varphi + C_1 + C_2) + n(\log_2 N + \log_2 \sigma_c + \log_2 \psi + C_1 + C_2) \text{ bits}. \quad (4)$$

In the framework of the given representation, the gain in description length when the group is discriminated is

$$\Delta L = L_0 - L_1 = n\left(\log_2 \frac{\sigma_r}{\sigma_c} + \log_2 \frac{\varphi}{\psi}\right) \text{ bits}. \quad (5)$$

According to the RMDL principle, the gain in the description length is the main criterion for carrying out the grouping. Based on the experimental data, it was determined how the percentage error allowed by a person when detecting a stimulus depends on the information content of the stimulus—i.e., the gain ΔL in description length achieved when grouping was carried out.

The following values were used when the stimulus and the field elements were generated (Figs. 3 and 4): $\varphi = \pi$ (the field elements had arbitrary orientation), $\sigma_r = 32$; the number n of elements in the stimulus varied in the range from 3 to 8, and ψ and σ_c took values in the ranges $[\pi/8, \pi/2]$ and [4, 16], respectively.

Figure 5a shows examples of how the percentage error depends on the number of elements in the stimulus for various values of parameter ψ. The percentage error simultaneously depends on all the parameters of the stimulus: n, ψ, and σ_c.

Figure 5b shows how the percentage error depends on the information content ΔL of the stimulus. Each point corresponds to the percentage error for fixed parameters n, ψ, and σ_c, averaged over the series of experiments. Since all the points lie on the same curve, the most significant part of the variance in the percentage error can be explained only by the difference in the value of ΔL for various stimuli.

A COMPUTER MODEL

The proposed grouping criterion, Eq. (5), was used in constructing an algorithm for searching for sets of elements whose combination into a group reduces the description length. Experiments show that, on the images that were presented to the subject, this algorithm can also be used to find the desired stimulus (see Fig. 6). However, at this stage of the studies, it is hard to compare the percentage errors allowed by the program that implements the algorithm and by a person, since the time to perceive the stimulus was substantially limited for a person, and this had a great effect on the probability of detecting the stimulus.

At the same time, the developed algorithm can be used for problems of automatically analyzing actual images (see Fig. 7), and this emphasizes the importance of studying and transferring the mechanisms of human visual perception into the region of computer vision.

CONCLUSION

The psychophysical experiments that have been carried out and the testing of the algorithm for grouping structural elements showed that it is possible to apply the principle of representational minimum description length to the problem of perceptive grouping, and consequently that it is possible to quantitatively describe the Gestalt laws from the viewpoint of reducing the description length of an image in the framework of a definite representation.

However, this paper used a simple type of stimulus and consequently a limited representation of the images. To ex-

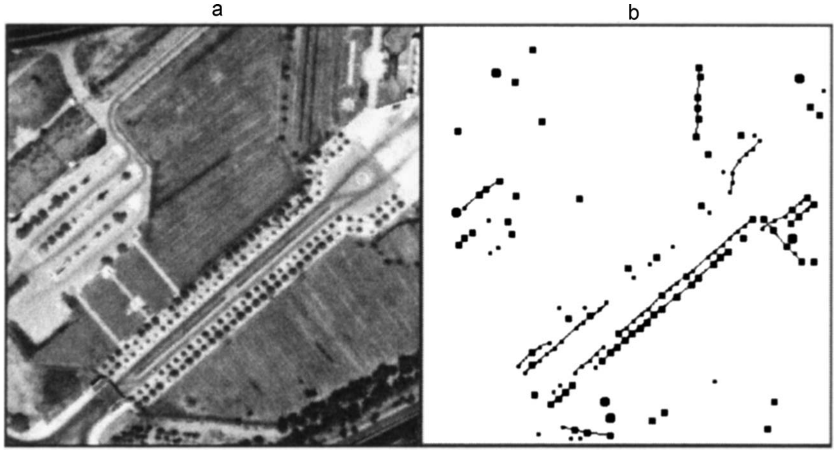

FIG. 7. Aerospace image (a) and the result of grouping the spots discriminated on it (b).

tend the resulting conclusions, it is necessary to set up an experiment with a wider set of stimuli. Based on the principle of representational minimum description length, it can be assumed that the percentage error depends on the chosen representation that can be distinguished by various people.

The choice of the representation that a person uses can also affect the quantity of the *a priori* information being communicated. In this experiment, all the parameters of the stimulus were communicated ahead of time to the person. If, however, the number of elements, for example, in the stimulus is not known ahead of time, it is necessary to distinguish about $\log_2 n$ bits in addition in order to describe it. If the orientation or other parameters of the stimulus are not known *a priori*, they must also be described in the process of grouping, and this reduces the gain in the description length. Additional experiments are needed to answer the question of how this will affect the probability of detecting the stimulus. A further expansion of the representation in terms of which the image is described is also required in order to expand the sphere in which the grouping algorithms are used in computer vision.

[1] A. Desolneux, L. Moisan, and J.-M. Morel, "Gestalt theory and computer vision," in *Seeing, Thinking and Knowing: Meaning and Self-Organisation in Visual Cognition and Thought*, ed. A. Carsetti (2004).

[2] A. Robert, "From contour completion to image schemas: a modern perspective on Gestalt psychology," Technical Report 9702, Department of Cognitive Science, University of California. 1997.

[3] S.-C. Zhu, "Embedding Gestalt Laws in Markov random fields—a theory for shape modeling and perceptual organization," IEEE Trans. Pattern Anal. Mach. Intell. **21**, 1170 (1999).

[4] A. S. Potapov, *Pattern Recognition and Machine Perception: A General Approach Based on the Principle of Minimum Description Length* (Politekhnika, St. Petersburg, 2007).

[5] P. M. B. Vitanyi and M. Li, "Minimum description length induction, Bayesianism, and Kolmogorov complexity," IEEE Trans. Inf. Theory **46**, 446 (2000).

[6] A. S. Potapov, "Study of image representations based on the principle of representation description length," Izv. Vyssh. Uchebn. Zaved. Prib. **51**, No. 7, 3 (2008).