Visual Groupings on the basis of Similarity and Proximity: A Model and Experimental Results

ABSTRACT
How objects are grouped in the visual field is a fundamental question that has yet to receive a satisfactory answer. Traditional Gestalt principles suggest a number of *ceteris paribus* rules that determine groupings such as proximity, similarity, closure, and common fate. But these rules are misleading, because they implicitly assume arbitrarily large groups can be formed in a single glance, and that different groups can be maintained in parallel. In fact, only one group can typically be focused on at any given time, and within this group, only a subset of the objects present will be full members of the attentional field. A neurophysiologically-realistic model of this effect is presented, and shown to be consistent with the experimental evidence.

SUMMARY
There has been a recent revival of interest in a very old problem, namely, how the brain forms groupings of visual objects [5]. A number of Gestalt principles were originally proposed to solve this problem, including those of proximity, similarity, closure, and common fate. However, a neurophysiologically-consistent model of these processes is still outstanding. Moreover, these principles do not take into account the dynamics of the grouping process, and in particular, the fact that within a single glance, only the dominant group is prominent in the attentional field, and even within this group, not all objects may be fully attended to. For example, assuming that there is a 5 by 5 matrix of identical objects, with greater spacing between columns than rows, the proximity rule dictates that grouping will be by columns. But it is also true that at any given time, only one column, typically centered on the fixation point will be fully attended to, and in addition, some of the objects far from the point of fixation in this dominant group will also be attenuated (see Figure 1).

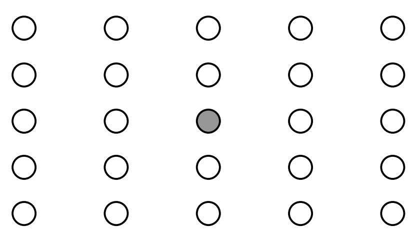

Figure 1. Proximity induces grouping by columns in this scene. However, notice that if the point of fixation is the center shaded circle, one typically cannot pay attention to other columns at the same time, and even within the column, there may be less attention to objects far from the point of fixation.

This talk first presents evidence to this effect. The method of Palmer [5] which allows one to convert the perception of groupings into reaction times has been adapted to this task. He asked subjects to find a pair of identical objects, such as two squares, that were present in a scene. If the objects were part of the same grouping, as indicated by Gestalt principles, the pair was identified roughly twice as quickly than if they were not. Here, the protocol consists of flashing an arrow on the screen to form a fixation point, and then flashing sets of visual groups. The hypothesis is that reaction time to locate another highlighted object on the screen will be a linear function of the distance from the fixation point, but that the slope of this function will be greater for objects outside the dominant group than for objects within the dominant group; these experiments are currently being carried out.

Once these results have been obtained, they will be correlated with a model of the formation of dominant groups. This model makes predictions for both the principles of proximity and similarity. The former is based on Li's [3,4] model of contour integration in V1. Li suggests that this can be modeled by assuming a Mexican hat weight function within this layer (his model is somewhat more complex, in that it also favors line orientations that are collinear as well as proximal, but this aspect is ignored here). The net effect of this is that proximal groups tend to be jointly activated, and as I additionally show, there is a fade-off in activity as a function of the distance from the point of fixation. Grouping by similarity can be accomplished by a resonance loop from tokens of a given type to a unit representing this type and back down to these tokens. The neural justification for this loop lies in the ubiquity of feedback connections in the visual cortex, and the fact that cells responding to objects in higher levels of visual processing are translation invariant (thus, more than one object can project to these units) [6]. The cognitive justification lies in the fact that these connections are necessary to explain phenomena such as the object-superiority effect. I have also shown that such a loop is useful in explaining object recognition in the case of polygons [1], and in face-recognition [2].

The model makes of number of predictions regarding grouping by proximity, similarity, and the interaction between the two. First, as expected, it shows that activation generally does not spread beyond the dominant group to other groups. For example, if there are a group of proximal or similar objects, forcing the fixation point to be centered on a group means that only the objects in that group will be active, and then only if they are not too distant from the fixation point. The model also predicts, less intuitively, that the effect of similarity largely disappears as objects become sufficiently close. That is, groupings are no more robust if identical objects are used rather than dissimilar ones if they are closely spaced. With this arrangement, they will be perceived as a line no matter what the makeup of the objects. The reason for this is that effect of similarity dominates over small distances because a powerful within-layer resonance loop develops between objects units as the distance from the center of the Mexican hat decreases. These predictions will be tested against the human results now being gathered. This talk will also touch on the predictions of the model for more interesting visual patterns, such as tilings. It will be shown that with the suitable extensions, the model can predict the basis grouping effects that are seen in these cases.

REFERENCES

[1] B.F. Katz, What makes a polygon pleasing? Empirical Studies of the Arts, in press.

[2] B.F. Katz, Evolution, the invisible artist. Empirical Studies of the Arts 17 (2000).

[3] Z. Li, Visual segmentation by contextual influences via intra-cortical interactions in the primary visual cortex. Network: Comput. Neural Sys., 10 (1999) 187-212.

[4] Z. Li, Pre-attentive segmentation in the primary visual cortex: Spatial Vision, 13 (2000) 25-50, 2000.

[5] S. Palmer, Vision science: Photons to phenomenology. (MIT, Cambridge) 1999.

[6] D.C. Vam Essen. D.C, and J.H. Maunsell, Hierarchical organization and functional streams in the visual cortex. Trends in Neuroscience, 6 (1983) 370-375.

[7] N. Weisstein, and C.S. Harris, Visual detection of objects: An object superiority effect. Science 186 (1974) 752-755.