Normalized Graph Cuts: Pages 16 - 34

Chapter 16

SEGMENTATION USING CLUSTERING METHODS

An attractive broad view of vision is that it is an inference problem: we have some measurements, and we wish to determine what caused them, using a mode. There are crucial features that distinguish vision from many other inference problems: firstly, there is an awful lot of data, and secondly, we don't know which of these data items come from objects — and so help with solving the inference problem — and which do not. For example, it is very difficult to tell whether a pixel lies on the dalmation in figure 16.1 simply by looking at the pixel. This problem can be addressed by working with a compact representation of the "interesting" image data that emphasizes the properties that make it "interesting". Obtaining this representation is known as **segmentation**.

It's hard to see that there could be a comprehensive theory of segmentation, not least because what is interesting and what is not depends on the application. There is certainly no comprehensive theory of segmentation at time of writing, and the term is used in different ways in different quarters. In this chapter we describe segmentation processes that have no probabilistic interpretation. In the following chapter, we deal with more complex probabilistic algorithms.

Segmentation is a broad term, covering a wide variety of problems and of techniques. We have collected a representative set of ideas in this chapter and in chapter **??**. These methods deal with different kinds of data set: some are intended for images, some are intended for video sequences and some are intended to be applied to **tokens** — placeholders that indicate the presence of an interesting pattern, say a spot or a dot or an edge point (figure 16.1). While superficially these methods may seem quite different, there is a strong similarity amongst them[1]. Each method attempts to obtain a compact representation of its data set using some form of model of similarity (in some cases, one has to look quite hard to spot the model).

One natural view of segmentation is that we are attempting to determine which components of a data set naturally "belong together". This is a problem known as **clustering**; there is a wide literature. Generally, we can cluster in two ways:

[1] Which is why they appear together!

Figure 16.1. As the image of a dalmation on a shadowed background indicates, an important component of vision involves organising image information into meaningful assemblies. The human vision system seems to be able to do so surprisingly well. The blobs that form the dalmation appear to be assembled "because they form a dalmation," hardly a satisfactory explanation, and one that begs difficult computational questions. This process of organisation can be applied to many different kinds of input. *figure from Marr, Vision, page101, in the fervent hope that permission will be granted*

- **Partitioning:** here we have a large data set, and carve it up according to some notion of the association between items inside the set. We would like to decompose it into pieces that are "good" according to our model. For example, we might:
  - decompose an image into regions which have coherent colour and texture inside them;
  - take a video sequence and decompose it into **shots** — segments of video showing about the same stuff from about the same viewpoint;
  - decompose a video sequence into motion blobs, consisting of regions that have coherent colour, texture and motion.
- **Grouping:** here we have a set of distinct data items, and wish to collect sets of data items that "make sense" together according to our model. Effects like

occlusion mean that image components that belong to the same object are often separated. Examples of grouping include:

- collecting together tokens that, taken together, forming an interesting object (as in collecting the spots in figure 16.1);
- collecting together tokens that seem to be moving together .

16.1 Human vision: Grouping and Gestalt

Early psychophysics studied the extent to which a stimulus needed to be changed to obtain a change in response. For example, Webers' law attempts to capture the relationship between the intensity of a stimulus and its perceived brightness for very simple stimuli. The Gestalt school of psychologists rejected this approach, and emphasized grouping as an important part of understanding human vision. A common experience of segmentation is the way that an image can resolve itself into a **figure** — typically, the significant, important object — and a **ground** — the background on which the figure lies. However, as figure 16.2 illustrates, what is figure and what is ground can be profoundly ambiguous, meaning that a richer theory is required.

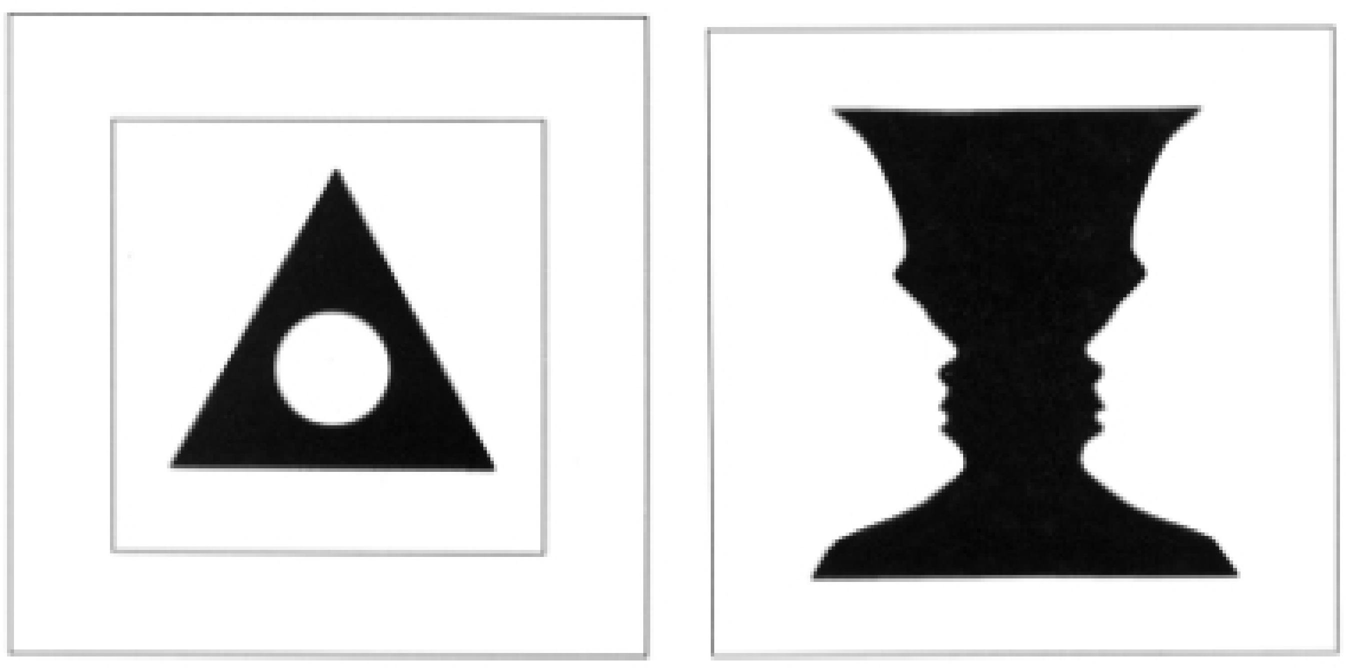

Figure 16.2. One view of segmentation is that it determines which component of the image forms the figure, and which the ground. The figure on the left illustrates one form of ambiguity that results from this view; the white circle can be seen as figure on the black triangular ground, or as ground where the figure is a black triangle with a circular whole in it — the ground is then a white square. On the right, another ambiguity: if the figure is black, then the image shows a vase, but if it is white, the image shows a pair of faces. *figure from Gordon, Theories of Visual Perception, page 65,66 in the fervent hope that permission will be granted*

The Gestalt school used the notion of a **gestalt** — a whole or a group — and of its **gestaltqualität** — the set of internal relationships that makes it a whole