Foreword

The representation of objects by their parts has a long tradition in computer-aided design, simulation, and in cognitive psychology. Indeed, in these areas it is the dominant strategy for representing complex 3-D objects. It is absolutely clear, therefore, that part representations are excellent for many computational and cognitive tasks. What has not been so clear is how they might be useful in computer vision and robotics.

The first parts representation was suggested by Binford (Binford, 1971); this is the idea of generalized cylinders. Unfortunately, the recovery of this type of representation seems to require elaborate line grouping and reasoning, which is a difficult and largely unsolved problem. Moreover, because such descriptions are often not unique, it is unclear how they aid in object recognition.

The idea of generalized cylinders has subsequently been elaborated in two very different ways. One variation is due to Biederman (Biederman, 1985), who suggested using the Cartesian product of qualitative properties such as tapering, cross-section, etc., in order to create a qualitative taxonomy of generalized cylinders. Correct choice of these qualitative properties can make the recovery process much simpler. By using deformable superquadrics to model these qualitative properties, Dickinson, Pentland, and Rosenfeld (Dickinson et al., 1992a; Dickinson et al., 1992b) were the first to successfully use this approach in real imagery.

A second alternative to generalized cylinders was suggested by Pentland (Pentland, 1986), in which he proposed a parametric version of generalized cylinders based on deformable superquadrics. He argued that the use of a parameterized implicit function, such as the superquadric, converts the problem of recovering a description into a relatively simple numerical optimization that can be made overconstrained and therefore robust.
Although this initial formulation for superquadric recovery proved unstable, by 1987 Pentland had developed a stable method for both segmentation and fitting deformable superquadrics using the Minimum Description Length (MDL) principle in a segment-and-fit paradigm (Pentland, 1987). However, this solution was extremely slow, and it was finally Solina and Bajcsy (Bajcsy and Solina, 1987) that developed the first practical method for recovering superquadric parameters. This method immediately became (and remains to this day) the standard method for fitting superquadrics.

Since then, the idea of fitting parameterized deformable models has been extended in several directions. Perhaps the most popular extension has been to employ physics-based techniques for fitting and tracking. This approach provides a robust framework for fitting and offers the possibility for natural extension to moving, dynamic scenes.

Pentland (Pentland, 1990) was the first to take this approach, using an eigenvector analysis of the resting shape to produce a new parameterization of the superquadric deformations that is linear and orthogonal (and consequently unique). This method has been successful at solving several difficult recognition and tracking problems, such as the recognition and tracking of people (Pentland and Sclaroff, 1991).

Metaxas and Terzopoulos (Metaxas and Terzopoulos, 1991) have further extended the physics-based approach by developing a class of deformable models in which both global and local deformations are physics-based. The global deformations capture the salient structure of object parts, while the local deformations capture the object’s details.

Biederman’s qualitative representation and the parametric superquadric representation have complementary properties. The qualitative representation of part structure has proven useful for grouping regions and edges, while the parametric representations have proven useful for recovering precise descriptions of shape. It is therefore natural to try to combine the strengths of the two approaches, using one for grouping, and the other for fitting and description. This idea has been developed by Metaxas and Dickinson (Metaxas and Dickinson, 1993) and by Raja and Jain (Raja and Jain, 1994), and shows great promise.

Despite all this progress, more work remains to be done, particularly in the difficult area of segmentation. There are in principle two types of segmentation methods: “segment-then-fit,” in which segmentation and fitting are only loosely connected (Gupta et al., 1989b; Pentland, 1990; Darrell et al., 1990; Ferrie et al., 1993), and “segment-and-fit,” in which segmentation and fitting are accomplished simultaneously (Pentland, 1987; Gupta and Bajcsy, 1993; Leonardis et al., 1997; Horikoshi and Suzuki, 1993), usually using an MDL criterion. We think that this
book shows that great strides towards reliable recovery of superquadric parameters for complex and articulated objects are being made.

Finally, we would like to thank Franc Solina, Aleš Leonardis, and Aleš Jaklič not only for writing this book, but also for their vital contributions to the problem of recovering parametric descriptions from complex scenes. They have been a critical force in advancing this area of vision science.

Ruzena Bajcsy
University of Pennsylvania
Philadelphia

Alex Pentland
Massachusetts Institute of Technology
Cambridge