Vision Research 126 (2016) 330–346

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Bent out of shape: The visual inference of non-rigid shape transformations applied to objects

Patrick Spröte *, Roland W. Fleming

Department of Psychology, Justus-Liebig-Universität Giessen, Germany

ARTICLE INFO

Article history:
Received 3 February 2015
Received in revised form 31 July 2015
Accepted 10 August 2015
Available online 26 September 2015

Keywords:
Perceptual organization
Scission
Causal history
Transformation
Generative model
Shape representation

ABSTRACT

In everyday life, we can often identify when an object has been subjected to some kind of transformation that alters its shape. For example, we can usually tell whether a can has been crushed, or a cookie has been bitten. Conversely, our ability to recognize objects is often robust across such shape transformations: we can still identify the can even though it has been dented. This ability to determine and discount the causal history of objects suggests the visual system may partially decompose the observed shape of an object into original (untransformed) elements plus the transformations that were applied to it. We sought to shed light on this possibility, using 'bending' as an example transformation. In one experiment subjects matched the degree of bending applied to random 3D shapes. We find that subjects could match the degree of bend, although there was a tendency to overestimate bends, especially for the least bent objects. In two other experiments, observers had to identify individual objects across different degrees of bending. Subjects performed significantly above chance although not as well as when the objects differed by rigid rotations without any bends (cf. traditional mental rotation experiments). Together our findings suggest that subjects can to some extent extract information about transformations applied to shapes, while ignoring other differences. At the same time subjects show a certain degree of invariance across shape transformations. This suggests scission of a shape's representation into its causes – a base shape and transformations applied to it.

1. Introduction

Both the cookie and the croissant in Fig. 1 exhibit a concavity. However, we can easily tell that those concavities originated from quite different processes or transformations. In the case of the cookie, the concavity was created by a process of forceful removal of cookie matter—a 'bite'—whereas the concavity in the croissant was created by shaping the dough around the concavity—a 'bend'. The fact that we can visually understand the differences in causal history between these two quite similar shapes suggests that the visual system readily seeks to identify the generative processes that create or modify objects, as part of its visual representation of shape (Feldman & Singh, 2006; Hoffman & Richards, 1984; Leyton, 1989; Pentland, 1986a, 1986b; Richards, 1988). In this study we wanted to shed some light on the representation of transformations within the representation of whole objects. In other words, how subjects infer transformations from shape.

Inferring transformations from observations of objects is computationally challenging. Under special conditions when both non-transformed and transformed versions of an object can be observed (e.g., before and after the transformation), inferring the transformation is not trivial, but the computations required can at least be defined in a relatively straightforward way. First, the visual system would have to establish correspondence between locations on the two versions of the object, and then identify a geometrical mapping from one point set to the other (i.e., solve the 'non-rigid registration' problem; for a review see Crum, Hartkens, and Hill (2004)). Then, to express the complex pattern of correspondences in terms of a simple global transformation, such as a 'twist' or 'bend' of a certain magnitude, some factorization or re-parameterization of the mapping might be required. Moreover, there may be difficulties when the transformation accretes or deletes portions of the shape—as in the bitten cookie—causing unmatchable features (i.e., points that occur on one version of the object that have no counterpart on the other version, precluding correspondence).

Nevertheless, the problem of inferring transformations becomes significantly more difficult in the more typical situation, as in

* Corresponding author at: Justus-Liebig-Universität Giessen, Department of General Psychology, Otto-Behaghel-Straße 10F, 35394 Giessen, Germany.
E-mail address: patrick.sproete@googlemail.com (P. Spröte).

http://dx.doi.org/10.1016/j.visres.2015.08.009
0042-6989/

Fig. 1. Two objects with similar concavities that originated from different processes. The cookie is perceived as 'bitten' whereas the croissant is perceived as 'bent'.

Fig. 1, when *only* the transformed version of each object is observable. Under these conditions, how can the visual system make the requisite comparisons to work out how the object has changed? Indeed, the task of inferring transformations from observation of a single object can be thought of as a form of blind source separation problem. Any given feature on the shape could be affected to an unknown extent by the transformation. So, for any given feature, how does the visual system separate out the contribution of the original non-transformed shape from the contribution of the transformation to the observed geometry? Some features may be essentially unaffected by the transformation, while others might be almost entirely artifactual, introduced into the shape by the process of bending, twisting, stretching or scraping that has been applied to the original object. Thus, the inference of transformations becomes under-constrained when only the end-state is observable. To infer causal history, we presumably draw on previous experience with objects as well as general assumptions about untransformed objects and the geometrical signatures of specific transformations (e.g., the introduction of helical ridges into an object caused by 'twisting', or the curvature of an object caused by 'bending'). In addition to shape cues, there may also be other surface features and markings that act as signatures of processes. For example, in Fig. 1 the isotropic texture of the cookie remains unaffected by the bite transformation, whereas for the croissant, the bending transformation structures the anisotropic surface texture markings into patterns that indicate the direction and form of the transformation. This results in radial patterns and compressed texture markings within regions of higher bend. In other words, bend and bite transformations affect surface texture in different ways, indicating how the local coordinate frame has been affected.

It is therefore interesting to ask to what extent the visual system can separate shape features into those that pre-dated a specific transformation—that is, 'intrinsic' features of the object—and those that were introduced or modified by the transformation—that is, 'extrinsic' properties of the shape. Here, we seek to measure the extent to which the visual system can achieve this separation. Before describing how we approach this experimentally, we set the study in context by briefly reviewing related work on the perceptual organization of shape and the inference of generative processes from shape.

1.1. Background: perceptual organization of shape

One of the major problems for the visual system in 3D shape perception is the lack of three-dimensionality in the retinal image. The visual system cannot measure the third dimension directly, so it has to reconstruct it from the two dimensional retinal image. A huge amount of research has dealt with questions of how the brain uses stereopsis, motion, shading, texture, or other cues to reconstruct local 3D shape properties such as position in depth (Bülthoff & Mallot, 1988; Stevens & Brookes, 1987), surface orientation (Johnston & Passmore, 1994; Koenderink, van Doorn, & Kappers, 1992) or local curvature (Johnston & Passmore, 1994; Rogers & Cagenello, 1989). However, it should be clear that estimating local surface structure is not all there is to 3D shape perception. Even a complete and error-free map of local surface properties would leave out important information about object shape, because many important object properties consist in how the different portions of the shape *relate* to one another (e.g., whether the object is top-heavy; whether it is symmetrical; whether it is composed of multiple distinct parts). Other researchers have investigated some of these more 'global' aspects of shape representation. For example, when looking at objects we can estimate their center of mass (Baud Bovy & Soechting, 2001; Cholewiak, Fleming, & Singh, 2013), report certain kinds of symmetrical relations (Sawada & Pizlo, 2008; for a review see Treder, 2010) or identify different functional and meaningful parts (de Winter & Wagemans, 2006; Hoffman & Richards, 1984). For example, the handle of the pot in Fig. 2 is usually perceived as being a different part and serving a different purpose than the bowl. There is no way we could do that without perceptually organizing local information into more global units and assign meaning to them.

Another important line of research is devoted to the format of the shape representation used by the visual system. A widely recognized approach coming from computer vision (Blum, 1973; see also Feldman and Singh (2006), Twarog, Tappen, and Adelson (2012)) is the medial axis transform. The medial axis of a shape or object (see Fig. 2) can be imagined as its underlying 'skele ton'—similar to the bones of a human body—which captures the local symmetry axis of the constituent parts. All skeletal representations are organized hierarchically, representing object features at different levels of resolution. Bigger branches (or parent branches) thereby resemble more global object features and smaller branches code the fine structure of objects (local features). Kovacs and Julesz (1994) found that Gabor targets were easier to detect when located on the medial axes of objects. More recently, by asking hundreds of subjects to tap once anywhere within a shape, Firestone and Scholl (2014) found that the pattern of tapping points reflected the shape's medial axis. Such findings suggest that at very least, medial axes are locations that satisfy important geometrical conditions that are important for visual processing, and possibly reflect explicit representation of medial axes by the visual system.

Despite their potential importance, there is very little research on the interpretation of the meaning of parts or part structure (but see Kim and Feldman (2009), Spröte and Fleming (2013)). In other words, what causal origins do we assign to certain parts of an object and how do these inferred causes influence our perception of shape?