Neurocomputing 147 (2015) 485–491

Contents lists available at ScienceDirect

Neurocomputing

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

# Motion sequence analysis in the presence of figural cues

Pawan Sinha [a], Lucia M. Vaina [a,b,c,d,e,*]

[a] Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA
[b] Brain and Vision Research Laboratory, Department of Biomedical Engineering, Boston University, Boston, MA 02215, USA
[c] Graduate Program in Neuroscience, Boston University, USA
[d] Department of Neurology, Boston University, USA
[e] Department of Neurology, Harvard Medical School, Boston, MA, USA

ARTICLE INFO

*Article history:*
Received 23 May 2014
Received in revised form
9 June 2014
Accepted 11 June 2014
Communicated by Witali Dunin-Barkowski
Available online 28 June 2014

*Keywords:*
Motion
Form-cues
Structure-from-motion
Stroke patients

ABSTRACT

The perception of 3-D structure in dynamic sequences is believed to be subserved primarily through the use of motion cues. However, real-world sequences contain many figural shape cues besides the dynamic ones. We hypothesize that if figural cues are perceptually significant during sequence analysis, then inconsistencies in these cues over time would lead to percepts of non-rigidity in sequences showing physically rigid objects in motion. We develop an experimental paradigm to test this hypothesis and present results with two patients with impairments in motion perception due to focal neurological damage, as well as two control subjects. Consistent with our hypothesis, the data suggest that figural cues strongly influence the perception of structure in motion sequences, even to the extent of inducing non-rigid percepts in sequences where motion information alone would yield rigid structures. Beyond helping to probe the issue of shape perception, our experimental paradigm might also serve as a possible perceptual assessment tool in a clinical setting.



## 1. Introduction

Motion of objects in the environment induces complex transformations in their images. The human visual system can recover the 3-D structure of the viewed objects and their motion in space by interpreting these image transformations [4,34]. As early as 1953, Wallach and O'Connell demonstrated humans' capacity to interpret structure from motion while studying what they termed the 'kinetic depth effect' [39]. In their experiments, an unfamiliar object was rotated behind a translucent screen with its shadow being observed from the other side of the screen. In most cases, the viewers were able to describe correctly the 3-D shape of the hidden object and its motion, even when each static shadow projection of the object was unrecognizable and contained no 3-D information.

Any vision system that attempts to compute 3-D structure from motion must contend with the problem that the recovery of structure is under-constrained; there are infinitely many 3-D structures consistent with a given pattern of motion in the changing 2-D image. Additional constraint is required to establish a unique interpretation. Computational studies have used the rigidity assumption to derive a unique 3-D structure and motion; they assume that if it is possible to interpret the changing 2-D image as the projection of a rigid 3-D object in motion, then such an interpretation should be chosen [2,6,10,13,17,21–25,33–35,43]. The rigidity assumption was suggested by perceptual studies that described a tendency for the human visual system to choose a rigid interpretation of moving elements [8,14,15,39].

The rigidity assumption has proven to be a powerful constraint, one that appears sufficient to explain how the human visual system solves the structure from motion problem in general settings. However, some interesting perceptual effects suggest that this notion of sufficiency might need to be revisited, at least insofar as modeling human performance is concerned. According to the rigidity assumption, a rigid object in motion should necessarily be perceived as rigid. But, a few studies have reported instances where displays of rigid objects in motion can give rise to the perception of distorting objects [3,5,9,40,44]. As detailed below, we suggest that these breakdowns of rigidity perception hint at a significant contribution from figural shape cues in the perceptual analysis of dynamic sequences. In this paper, we develop the hypothesis of a role for figural cues in sequence analysis, and present experiments designed to test this hypothesis.

The remainder of this paper is organized as follows: we first briefly review the current state of research related to the recovery of 3-D structure from motion and static image attributes; subsequently we present a hypothesis regarding interactions between shape information derived using motion and that derived from figural cues; and in Section 2 we describe the psychophysical

* Corresponding author at: Brain and Vision Research Laboratory, Department of Biomedical Engineering, Boston University, Boston, MA 02215, USA.
Tel.: +1 617 353 2455; fax: +1 617 353 6766.
*E-mail address:* vaina@bu.edu (L.M. Vaina).

http://dx.doi.org/10.1016/j.neucom.2014.06.037
0925-2312/

experiment and in Section 3 we present evidence from normal observers as well as stroke patients with perceptual impairments in support of the basic thesis of this paper.

### *1.1. Possible strategies for the analysis of dynamic sequences*

In a laboratory setting, it is possible to generate motion stimuli that cleanly dissociate between static and dynamic sources of 3-D information. For instance, the random clusters of dots often used in perceptual studies of recovering structure from motion are carefully controlled so that no single frame has any discernible static organization that may provide hints about the 3-D configuration of the dots. This has been the dominant paradigm of structure from motion research so far. Since there is no static 3-D information in such displays, the question of how statically and dynamically derived 3-D shape estimates interact with each other has been sidestepped.

In the real world, however, the static and dynamic shape cues are almost always confounded. The objects we see moving, e.g. cars, people and airplanes have non-random static configurations that may be used to derive good 3-D shape estimates. Since these estimates are available simultaneously with those from motion cues, we are faced with the question of whether they play a role in determining the eventual 3-D percepts.

Let us consider two extreme scenarios. In scenario 1, the visual system uses only motion trajectories of feature points to estimate structure ('features' are defined as points of high curvature or other punctuate discontinuities). This has been the typical approach for sequence analysis. The reason for the popularity of this approach is its ability to yield unique structure interpretations based on motion information from only a few frames and features. For instance, Ullman [34] has shown that under orthographic projection, three views of four non-coplanar points are sufficient to guarantee a unique 3-D solution. Longuet-Higgins and Prazdny [17] proved that the instantaneous velocity field and its first and second spatial derivatives at a point admit at most three different 3-D interpretations. Tsai and Huang [33] showed that two perspective views of seven points are also usually sufficient to guarantee uniqueness.

While scenario 1 is mathematically elegant and powerful, we should consider whether the human visual system does in fact adopt such a strategy, or perhaps it might incorporate other, non motion-based, cues as well in its analysis of dynamic sequences. This leads us to scenario 2. Here, we treat each frame of a motion sequence as an entity to be analyzed on its own, in terms of the figural cues it contains. These cues provide 3-D estimates on a frame-by-frame basis, rather than requiring the use of feature motion trajectories, as prescribed by scenario 1. Several figural cues, such as shading, or texture gradients can provide 3-D structure information [6,11,31,45]. Even in the absence of such gradients, global contour based cues provide powerful constraints for 3-D shape recovery [7,12,16,18,41,42] as demonstrated by computational schemes for the recovery of 3-D structures from simple 2-D line drawings [19,27,28].

Is human analysis of dynamic sequences more akin to scenario 1 (predominant use of motion information), or scenario 2 (predominant use of figural information, when such information is available)? Addressing this question presents a challenge in that for most dynamic sequences, both scenarios tend to produce identical results. For instance, a moving wire-frame cube would be seen as a cube irrespective of whether one uses structure from motion algorithms on the vertex trajectories, or applies shape from contour algorithms on individual frames. In order to overcome this constraint, we need dynamic sequences where the motion based and figural content-based strategies yield different results.

Here we use dynamic sequences showing rigid wire-frames in motion, where the wire-frame objects are specially constructed so that their different views suggest different 3-D shapes. Conventional structure from motion algorithms would easily recover the true rigid 3-D structure of these objects. However, the use of figural cues on a frame by frame basis would suggest that the underlying 3-D shape was changing over time. Thus scenario 1 would predict the percept of a rigid object, while under scenario 2, a non-rigid percept would result. Fig. 1 illustrates our proposal. The object depicted across all of the frames is a rigid wire-frame but its different views appear to derive from different underlying 3-D geometries. Thus, if scenario 2 is correct, then even though there is a unique rigid interpretation for the whole sequence, observers would opt for a strange non-rigid one. If, however, motion information takes precedence over figural cues, then the veridical rigid structure will be perceived. In the next section, we develop an experimental paradigm to test this hypothesis.

## 2. Methods

### *2.1. Experimental tests of the contribution of motion and figural cues in the analysis of dynamic sequences*

We describe our experimental paradigm by means of a specific example. Consider a 2-D image showing a cube. Up to a depth reversal, there is a unique set of depth values for all of the vertices that would be precisely consistent with a 3-D cube shape. Let us label this set of depth values as defining the 'base' structure. Now, we can add increasing amounts of random depth noise to the vertices. Since the noise is constrained to be only along the depth axis, the 2-D projections of the resulting objects are always the same (from the original viewpoint). However, the 3-D structures are now no longer consistent with an observer's expectations

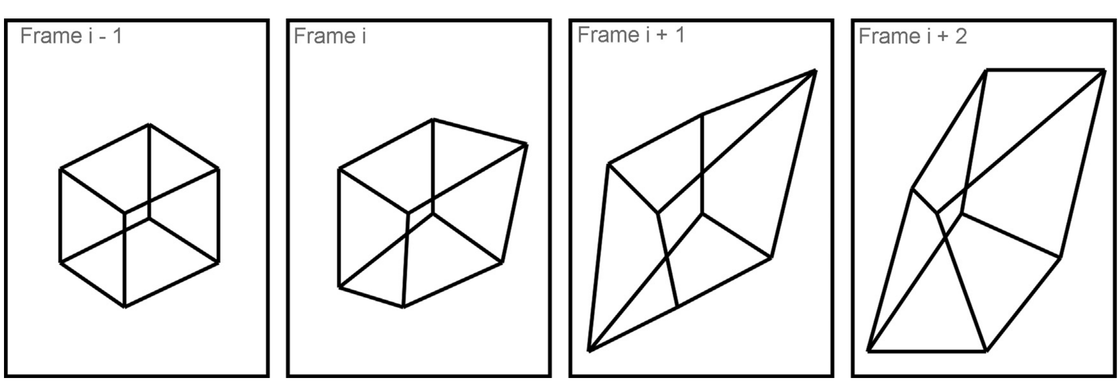


**Fig. 1.** A few frames from a motion sequence that show a specially constructed rigid 3-D object whose different views suggest different 3-D structures. Analysis of the motion trajectories of the vertices would yield the veridical rigid 3-D structure. However, an analysis based on figural cues within each frame would suggest non-rigidity.