Generating spatial attention cues via illusory motion

Jensen, Janus Nørtoft; Hannemose, Morten; Wilm, Jakob; Dahl, Anders Bjorholm; Frisvad, Jeppe Revall; Belongie, Serge

Published in:
CVPR Workshop on Computer Vision for Augmented and Virtual Reality

Publication date:
2019

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Generating Spatial Attention Cues via Illusory Motion

Janus Nørtoft Jensen\textsuperscript{1}\textsuperscript{*}  Morten Hannemose\textsuperscript{1}\textsuperscript{*}  Jakob Wilm\textsuperscript{2}  Anders Bjorholm Dahl\textsuperscript{1}  Jeppe Revall Frisvad\textsuperscript{1}  Serge Belongie\textsuperscript{3}

\textsuperscript{1}Technical University of Denmark  \textsuperscript{2}University of Southern Denmark  \textsuperscript{3}Cornell Tech

\textsuperscript{*}These authors contributed equally to this work.

Abstract

For many applications in augmented reality (AR), the user has a much more enjoyable experience if the AR system is able to properly guide the user’s attention. In this extended abstract, we explain how to create patterns of light that when projected onto an object are perceived as if the object itself is moving. This can be used as a spatial attention cue. We accomplish this with a calibrated projector-camera setup to synthesize an image from the projector’s point of view. This image is filtered to create local phase changes that are then projected back onto the object and perceived as motion. Our method will be shown as a live demonstration at the CV4AR/VR workshop at CVPR 2019.

1. Introduction

Visual attention is important as it affects our performance in many visual tasks [2]. To successfully accomplish tasks such as visual search or tele-assistance we need effective spatial attention cues to attract the attention of a user. What constitutes an effective cue is in part determined by the task at hand. A very obvious cue, such as a big bouncing red arrow, might be the best, if the purpose is to alert the user of possible danger. In many other situations, however, more subtle cues might be preferred. This could be in the setting of an escape room game where we would guide players towards the next clue if they are stuck. A very obvious cue could possibly ruin the fun of the game, however a subtle one, which would take longer to notice, could be useful.

As described by Carrasco and Barbot [2], our attention is involuntarily captured by sudden changes in the environment. We seek to exploit this effect by creating apparent motion through light projection. While the human ability to detect motion is not better in the peripheral vision [10], the speed of visual processing does increase in the peripheral vision [3]. When something moves differently compared to the movement of the observer, it becomes a powerful cue to attract attention, especially in the periphery of the visual field [12]. This is referred to as a relative-motion cue. Our idea is to guide attention by creating apparent relative-motion cues in a scene by means of a light projector that modifies the appearance of a physical object to make it look as if it were moving, thereby creating illusory motion.

2. Projecting Illusory Motion

Prior work in guiding visual attention has mainly been focused on head-mounted displays and images displayed on a screen. The spatial attention cues used include flickering [16, 17], blurring [8], and color manipulation [9]. Spatial projection has been shown to be more effective than head-mounted displays in providing spatial instructions in assembly tasks [1]. Research in attention cues for spatial projection is however limited. Some use laser projection [14, 7] as a simple cue. Similar to our approach Taki-
moto et al. [15] uses a calibrated projector-camera setup. They modulate color information recorded with the camera and project the result back onto the object. For humans, the sensitivity to color variations declines faster compared with the sensitivity to luminance [6]. It thus seems natural to investigate modifying the luminance instead.

Our approach to create the illusion of motion is based on adapting the work by Freeman et al. [4] to a projector-camera setup. In short, they apply local filters with continuously varying phase over time to the image. This is based on the observation that local phase changes are interpreted as global motion. We synthesize an image from the viewpoint of the projector (Figure 1) and use it as input for their method. The filter response is then projected back onto the object. In Figure 2, examples of the object with the filter response projected onto it are shown along with the images projected.

2.1. Projector-camera calibration

We use two cameras and a projector mounted in a fixed setup, and we model the projector and cameras as pinhole cameras with radial distortion. To calibrate the system we encode the projector's pixel coordinates using structured light [13], detect corners in images of a checkerboard and convert these to the projector’s pixel space by local homographies [11]. The projector and cameras are then calibrated with Zhang’s method [18].

2.2. Synthesizing image from projector’s point of view

To synthesize an image from the projector’s point of view we use structured light to create two 3D-scans of the object - one based on each camera. The resulting point clouds and the pixel intensities associated with the points are then projected back to the projector to form the desired image (Figure 1). Creating two separate 3D-scans enables us to scan all points visible in the projector and at least one camera, thereby getting better coverage of the object. After the points from each scan have been projected to the projector’s pixel space they are rounded to the nearest integer pixel. Because the projector has a lower resolution than the cameras, each pixel in the projector contains multiple measurements. We compute medians of these to obtain a single value per pixel.

2.3. Creating Illusory Motion

With the image from the projector’s point of view, we can directly apply the method of Freeman et al. [4] to this
image. As the filter responses contain both positive and negative values, and the projective setting has the constraint of only adding light, we add a constant value to the filter response to make all values positive. The filtered image is multiplied with a mask to restrict light to the object, and the resulting image is projected onto the object. Examples of the projected image and the resulting effect are in Figure 2.

3. Future work

We can perhaps utilize the 3D information obtained through our process to make the projected patterns less view dependent. Because we know the 3D position of each pixel in addition to its intensity, we can compute a normal at each point. If we, based on these normals, choose a consistent orthonormal basis at each point, such that the basis changes smoothly over the object [5], and assuming that the object is locally planar, we can project the filter onto this plane, and thereby approximate convolution along the surface of the object instead of in image space.

Our current work has been focused on creating the illusion of motion. Further work is needed to determine under which circumstances it is perceived as motion and to determine its effectiveness as a spatial attention cue. Furthermore, examining how it affects the user’s experience of interacting with the AR system, and how our cue compares with using alternative spatial attention cues is important to examine. Reasonable variables to look into would be how much light is necessary to make the effect noticeable and how this varies across the visual field.

References