Computers & Graphics 43 (2014) 11–20

Contents lists available at ScienceDirect

Computers & Graphics

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

Technical Section

A clustering-based method to estimate saliency in 3D animated meshes ☆

Abdullah Bulbul [a,b,*], Sami Arpa [b,c], Tolga Capin [d]

[a] Vision Science Graduate Group, University of California at Berkeley, Martin Banks' lab, 360 Minor Hall, Berkeley, CA 94720, USA
[b] Department of Computer Engineering, Bilkent University, Turkey
[c] School of Computer and Communication Sciences, EPFL, Switzerland
[d] Department of Computer Engineering, TED University, Turkey

ARTICLE INFO

Article history:
Received 2 October 2013
Received in revised form
18 April 2014
Accepted 29 April 2014
Available online 16 June 2014

Keywords:
Computer graphics
Visual perception
Saliency
Animated meshes

ABSTRACT

We present a model to determine the perceptually significant elements in animated 3D scenes using a motion-saliency method. Our model clusters vertices with similar motion-related behaviors. To find these similarities, for each frame of an animated mesh sequence, vertices' motion properties are analyzed and clustered using a Gestalt approach. Each cluster is analyzed as a single unit and representative vertices of each cluster are used to extract the motion-saliency values of each group. We evaluate our method by performing an eye-tracker-based user study in which we analyze observers' reactions to vertices with high and low saliencies. The experiment results verify that our proposed model correctly detects the regions of interest in each frame of an animated mesh.

1. Introduction

The human visual system (HVS) frequently shifts focal attention to the subsets of a scene, that is, to salient feature points. The visual acuity and the details transferred from the real world to the HVS change with these shifts. Fixation movements are the most important movements of the eye; the fixation mechanism allows us to direct the eyes towards objects of interest. While this process is automatically employed by the HVS, saliency detection mechanisms are not yet fully understood. For 3D graphics, automatic detection of salient features may provide significant advances in various problems, including selective rendering, view-point selection, retargeting, symmetry detection, segmentation, and 3D model compression.

Many view-independent saliency detection models have been proposed for 3D scenes in the graphics literature. In these methods, geometric features such as mean curvature differences at different scales and average variations between two polygons are considered, but temporal variations of the geometry are not well integrated. Regarding this drawback, we use HVS mechanisms supported by motion-related psychophysical experiments to develop a metric calculating the saliency of 3D objects based on their motion. Current research shows that while motion by itself does not attract attention, its attributes, such as initiation, may make it more salient. This paper presents a saliency model based on the effect of motion states on the attractiveness level of a visual stimulus.

The main contribution of this paper is a new approach to determine perceptually significant elements in animated 3D scenes using a motion-saliency model. Our proposed approach is based on clustering vertices with similar behaviors. To cluster vertices in each frame of a deforming mesh sequence are analyzed according to their motion properties. Vertices with similar motion behaviors are perceptually grouped with a Gestalt approach, thus each cluster is analyzed as a single unit. Representative vertices for each cluster are therefore used to extract the motion-saliency values of their clusters. To evaluate our model, we performed a user study to analyze observers' reactions to objects with high and low saliency values. The results of the experiment verify that the proposed metric correctly identifies the mesh regions with high motion saliency.

The paper is organized as follows: In Section 2, we present a review of previous studies in computer graphics utilizing the motion perception principles and the psychological principles that inform our method. Section 3 presents our 3D cluster-based motion-saliency estimation method. Section 4 presents the user study and its results. Section 5 presents a discussion and Section 6 concludes.

2. Related work

2.1. Concepts in visual attention and saliency

The visual attention mechanism can be divided into two components: bottom-up and top-down attention.

☆This article was recommended for publication by M. Spagnuolo.

* Corresponding author at: Vision Science Graduate Group, University of California at Berkeley, Martin Banks' lab, 360 Minor Hall, Berkeley, CA 94720, USA.
Tel.: +1 510 642 7679, fax: +1 510 643 5109.

http://dx.doi.org/10.1016/j.cag.2014.04.003
0097-8493/

Bottom-up attention: The bottom-up component of visual attention is driven merely by the properties of the visual scene, regardless of the viewer's intention. Viewer-independent factors (irrespective of personal tasks, experiences, etc.) direct the visual attention and are part of the bottom-up component.

Saliency, a bottom-up property, is mainly related to the differences of an object's various visual properties and its surroundings. The neurons employed in the visual system respond to image differences between a small central region and a larger surrounding region [1], in a process called the center-surround mechanism. Through this mechanism, the differences between a property and its surroundings stimulate the visual system. If an object is notably different from its surroundings, it becomes salient. This difference can be in terms of one or more properties of the object, such as hue, luminance, orientation or motion. A highly salient object pops out from the image and immediately attracts attention. This process is unconscious and operates faster than top-down, or task-oriented attention. The speed of bottom-up attention is usually between 25 to 50 ms per item, while task-oriented top-down attention takes more than 200 ms [1].

Top-down attention: What are we looking for greatly affects our visual perception. When we look for a specific type of object or for a specific property we may perceive many details that we may not perceive in a casual glance. On the other hand, biasing perception towards a specific target can make other objects less perceivable [2].

After objects have been selected from the scene in a bottom-up fashion, goal-oriented top-down attention determines *what* is perceived. This phase of attention includes constraining the recognized scene based on scene understanding and object recognition [1]. When a scene is constrained by the visual system, the region that gets the most attention is promoted, which is known as the winner-take-all principle [1].

With a search task to browse a scene, the HVS is optimally tuned according to the search goal, such that the features of the target become easily recognizable [3]. Interestingly, our visual system is not adjusted to the exact features of a search target, but adjusted to differentiate these features in the optimal way. For example, if our goal is to find a slightly right slanting object among objects oriented in an upwards direction, our sensitivity is tuned to that exaggerated feature in the target object to simplify differentiation. Similarly, when our attention is tuned to a search goal, we may not notice objects unrelated to our task even if they are easily visible; this phenomenon is called inattentional blindness [4].

Inhibition of return: Another principle of visual attention is called inhibition of return, first described in 1984 by Posner and Cohen [5], and which allows our visual system to perceive an entire scene rather than focusing only on the visually most attractive region. According to this principle, when a region is attended to once, our perception of that region is inhibited after the first 0.3 s and object recognition in that location decreases over approximately 0.9 s. As a result, our attention moves to a new region, enabling a search of different and novel regions on the visual periphery.

Motion perception: A difference of position in our visual field results in a sense of motion; this process requires a temporal analysis of the contents in our visual field. When two different images fall into our retina sequentially, our visual system must identify whether those images represent the same object in different positions or whether they are different objects. If the HVS determines that it is the former case, it has established that the object is moving. The HVS can easily perceive objects as smoothly moving, but the mechanism that detects motion is not that simple. Working out spatial relations is easier than solving temporal relations [6].

A proposed model to explain motion detection in the HVS is Reichardt's motion detector [7]. This device is based on small units responsible for detecting motions in specified directions. These units compare two retinal image points, and if the same signal appears in these two points with a small delay, the units detect motion in their specific direction [6]. Along with color, depth, and illumination, center-surround organization is also applied to motion processing in the HVS. The neurons processing motion have a double-opponent organization for direction selectivity [8], meaning that motion-detecting modules can inhibit their surroundings; motion must be differentiable compared to its surroundings to be detected.

In the spatial domain, the HVS tends to group stimuli by considering their similarities and proximity as introduced in the Gestalt principles. It is shown that the HVS also searches for similarities in the temporal domain and can group stimuli by considering their parallel motions [9]. According to this process, a group of moving dots with the same direction and speed could be perceived as a moving surface.

Visual motion may be referred to as salient because it has temporal frequency. On the other hand, recent studies in cognitive science and neuroscience have shown that motion by itself does not attract attention. However, phases of motion (e.g., motion onset, motion offset, continuous motion) have different degrees of influence on attention. Hence, each phase of motion should be analyzed independently. Abrams and Christ [10] experimented with different states of motion to observe the most salient one. They indicated that the onset of motion captures attention significantly compared to other states. Immediately after motion onset, the response to stimulus slows from the effect of the inhibition of return, and the attentional sensitivity to that stimulus is lost. Singletons, having a different motion than others within stimuli, capture attention in a bottom-up, stimulus-driven way. If there is a search target, only feature singletons attract attention. However, abrupt visual onsets capture attention even if they are not a target [11]. Hillstrom and Yantis [12] also showed that the appearance of new objects captures attention significantly compared to other motion cues and that motion offset and continuous motion do not capture the same level of attention.

2.2. Computational models of visual attention and saliency

Itti et al. [13,1] describe one of the earliest methods to compute the saliency of 2D images. To calculate the saliency of a region, they compute the Gaussian-weighted means of intensity, orientation, and color opponency properties in narrow and wide scales; the differences between these scales provide information on how a region is compared to its surroundings.

Lee et al. [14] introduced the concept of mesh saliency in 3D graphical models. In their work, the saliencies of mesh vertices are computed based on the mesh geometry. Their proposed mesh saliency metric is based on the center-surround operator on Gaussian-weighted mean curvatures. They use the computed saliency values to drive the simplification of 3D meshes, implementing Garland and Heckbert's Qslim method [15] for simplifying objects based on quadric error metrics.

The mesh saliency metric was improved by Liu et al. [16], who discuss two main disadvantages of Lee et al.'s work [14]. One disadvantage is that the Gaussian-weighted difference of fine and coarse scales can result in the same saliency values for two opposite and symmetric vertices because of the absolute difference in the equation. The other one is that combining saliency maps at different scales makes it difficult to control the number of critical points. Therefore, instead of the Gaussian filter, Liu et al. use a bilateral filter and define the saliency of a vertex as the Gaussian-weighted average of the scalar function difference between the neighboring vertices and the vertex itself. Kim et al. [17] presented a user study that compares the performance of the