A Decision Theoretic Approach to Motion Saliency in Computer Animations

Sami Arpa, Abdullah Bulbul, and Tolga Capin

Bilkent University, Computer Engineering Department,
Faculty of Engineering, 06800 Ankara, Turkey
{arpa,bulbul,tcapin}@cs.bilkent.edu.tr

Abstract. We describe a model to calculate saliency of objects due to their motions. In a decision-theoretic fashion, perceptually significant objects inside a scene are detected. The work is based on psychological studies and findings on motion perception. By considering motion cues and attributes, we define six motion states. For each object in a scene, an individual saliency value is calculated considering its current motion state and the inhibition of return principle. Furthermore, a global saliency value is considered for each object by covering their relationships with each other and equivalence of their saliency value. The position of the object with highest attention value is predicted as a possible gaze point for each frame in the animation. We conducted several eye-tracking experiments to practically observe the motion-attention related principles in psychology literature. We also performed some final user studies to evaluate our model and its effectiveness.

Keywords: saliency, motion, computer animation, perception.

1 Introduction

The impact of psychology and neuroscience disciplines to mature the techniques developed in computer graphics has increased through newly discovered biological facts. In psychological science, the sensitivity of human perception to motion is still a contemporary research issue. The fact that motion attracts attention is a former claim. New arguments focus on the attributes of motion. Recent developments in neurobiological science have helped to discover the reaction of brain to visual stimuli in very low detail. Recently, as opposed to the previous beliefs, it has been experimented that motion per se does not attract attention, but some of its attributes do [1].

We use neurobiological facts on motion and the hints provided by motion-related psychophysical experiments to develop a metric calculating the saliency of the objects due to their motion in computer graphics scenes. The model estimates the relative saliencies of multiple objects with different movements.

To evaluate our model, we performed a user study in which the observers' reactions to objects having high and low saliencies are analyzed. The results of this experiment verify that the proposed metric correctly identifies the objects with high motion saliency.

This paper presents a model to determine the perceptually significant objects in animated video and/or game scenes based on the motions of the objects. Instant decisions provided by the model could simplify interactive target decision and enable controlling the difficulty level of games in a perceptual manner.

The rest of the paper is organized as follows: In Section 2, a review of previous studies in computer graphics utilizing the principles related to motion perception and the psychological principles that affected our model are presented. Section 3 presents our model and the details of the user studies are given in Section 4 before concluding the paper in the last section.

2   Related Work

The related work for motion saliency could be divided into two parts as the usage of motion perception in Computer Graphics and the psychological studies related to the effect of motion on visual attention.

Computer Graphics. Saliency could be seen as the bottom-up stimulus driven part of the visual attention mechanism in which task dependent attention does not have a role. Recently, saliency computation has gained more interest in computer graphics. One of the earliest attempts to compute saliency of 2D images was described by Itti et al. [8]. A widely known saliency computation framework that works on 3D mesh models was proposed by Lee et al. [11] which computes saliency according to the surface curvature properties of the meshes. Both of these models compute saliency based on the center-surround mechanism of human visual attention. Compared to these more general saliency computation frameworks, there are less works on motion based saliency computation.

The saliency computation framework proposed by Bulbul et al. [4] calculates motion saliency in a center-surround fashion and combines it with geometric and appearance based saliencies to generate an overall saliency map. The motion saliency computation in this work is based on the idea that the regions having a distinct motion compared to their surroundings become more salient. Halit and Capin [6] proposed a metric to calculate the motion saliency for motion-capture sequences. In this work, the motion capture data is treated as a motion curve and the most salient parts of these curves are extracted as the keyframes of the animation.

Peters and Itti [13] observed the gaze points on interactive video games and concluded that motion and flicker are the best predictors of the attended location while playing video games. Their heuristic for predicting motion-based saliency (as for other channels like color-based and orientation-based) works on 2D images and it is also based on the center-surround machanism.

Visual sensitivity to moving objects is another aspect of motion perception that is utilized in computer graphics. Kelly [9] and Daly [5] studied to measure the spatio-temporal sensitivity and fit computational models according to their observations. Yee et al. [17] built on these studies and used the spatio-temporal sensitivity to generate error tolerance maps to accelerate rendering.

Psychological Literature. Andersen's research [2] elaborated low level processing of motion cues in visual system by showing specific roles of Where system in the brain, which is a part of visual cortex processing motion [12]. Spatial awareness based on motion cues and its hierarchical processing have been investigated in this study by analyzing different receptor cells having varied roles in Where system.

In spatial domain, visual system is tend to group stimuli by considering their similarity and proximity as introduced in Gestalt principles. It is shown that visual system searches similarities also in temporal domain and can group stimuli by considering their parallel motions [10]. A bunch of moving dots with the same direction and speed could be perceived as a moving surface with this organization.

Along with color, depth, and illumination; central-surround organization is also applied to motion processing in visual system. The neurons processing motion have a double-opponent organization for direction selectivity [2].

Visual motion may be referred as salient since 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. Phases of motion have different degrees of influence on attention. Hence, each phase of motion should be analyzed independently.

Abrams and Christ [1] experimented 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 with the effect of inhibition of return and attentional sensitivity to that stimulus is lost.

Singletons, having different motion than others within stimuli, capture attention in a bottom-up, stimulus driven control. If there is a target of search, only feature singletons attract attention. If it is not the target, observers' attention is not taken. However, abrupt visual onsets capture attention even if they are not the target [16].

Other than motion onset, the experiments in the work of Hillstrom and Yantis [7] showed that the appearance of new objects captures attention significantly compared to other motion cues.

3   Approach

By considering psychological literature given in the previous section, we describe a model to compare attention values of objects in terms of motion. Initially, we consider the motion attributes to discriminate different states of an object's motion. Following six states form the essence of a motion cycle (Fig. 1):

Static: No change of location.
Object Appearance: Appearance of an object which was not previously present on the screen.
Motion Onset: The start of motion (static to dynamic).
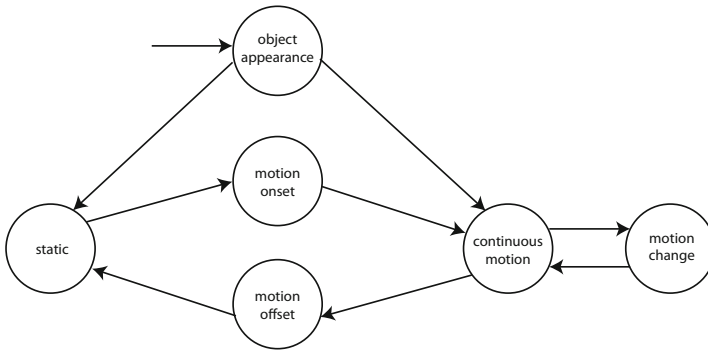**Motion Offset:** The end of motion (dynamic to static).

Fig. 1. Motion cycle of an object in an animation. Perceptually distinct phases are indicated as six states. Object appearance is the initial state for each object.

> **Continuous Motion:** The state of keeping the motion with the same velocity.
> **Motion Change:** Change in the direction or speed of a motion.

3.1 Pre-experiment

We performed an eye-tracker based pre-experiment to observe the relations between the defined states and the attentively attractive objects.

In the experiment, subjects were asked to watch the animations including movements of five balls within a 3D room. Eight different animations were used (Fig. 2). In each of them, different motion states and the observers' reactions to them are analyzed. We used an SMI Red Eye Tracker to observe and analyze gaze points of subjects during the animations. Eight voluntary graduate students who have normal or corrected to normal vision participated in the experiments. The subjects were not told about the purpose of the experiments and each of them watched eight different animations.

Experiment cases. The cases used in the pre-experiment are as follows:

- Case 1: Motion onset and motion offset are tested. Five initially static objects start to move in sequence with at least 3 seconds intervals. At the end, forty seconds later, all objects stop again in sequence with at least 3 seconds intervals.
- Case 2: Motion onset and motion offset are tested for shorter intervals. All objects start to move in sequence with 0.3 seconds intervals and stop the same way.
- Case 3: Object appearance is tested. Some objects starts to move and one object abruptly appears. There is an interval of five seconds between the latest motion onset and the object appearance.
- Case 4: Object appearance, motion onset, and motion change are tested. One object starts to move, one object appears, and another moving object changes its direction at the same time.
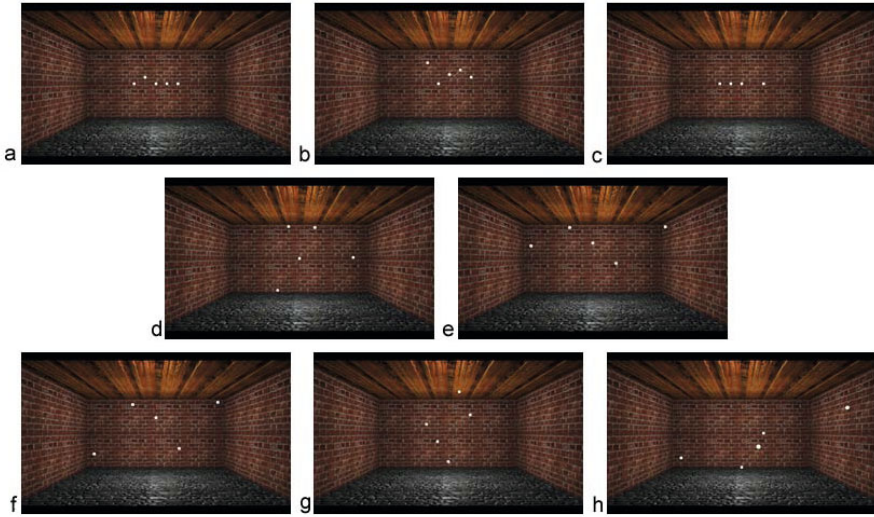
Fig. 2. Screenshots from the eight pre-experiment animations. **a:** Motion onset and offset are tested. **b:** Motion onset and offset are tested for shorter intervals. **c:** Object appearance is tested. **d:** Object appearance, motion onset, and motion change are tested. **e:** Object appearance, motion onset, and motion change are tested in larger spatial domain. **f:** Object appearance, motion onset, and motion change are tested in sequential order. **g:** Velocity difference is tested. **h:** States are tested altogether.

- Case 5: Object appearance, motion onset, and motion change are tested in larger spatial domain. Only difference from the forth case is that the objects are placed further points on the screen.
- Case 6: Object appearance, motion onset, and motion change are tested in sequential order. This time, three states do not occur at the same time, but in sequential order with intervals of 3 seconds.
- Case 7: Velocity difference is tested. Each object starts to move at the same time and the same direction with different velocities.
- Case 8: States are tested altogether.

Observations. The observations according to the pre-experiment results can be expressed as follows:

Motion onset, *motion change*, and *object appearance* states strongly attract attention. In the first case (Fig. 2-a), five object onsets were performed and among eight subjects, five subjects looked at all off the motion onsets. Each of the other three subjects missed only one motion onset and the missed objects were different for each of them. The gaze of all subjects was captured by the appearance of the object in the third animation (Fig. 2-c). In the cases four and five, none of the three objects with three different state sets dominate over the others (Fig. 2-d,e); indicating that *motion onset*, *motion change*, and *object appearance* states do not suppress each other. However, *motion offset*, *continuous motion*, and *static* states could be suppressed by any of the other three states.

Continuous motion and *motion offset* states capture attention merely over static state. The gazes of the subjects were rarely transferred to the objects with a motion offset event. However, when we asked the subjects, they have informed us that they were aware of all motion offsets but they did not move their gazes to see them. We can conclude that, in most cases, subjects perceive motion offsets in their peripheral view and their attention is not captured by these events.

None of the objects in *static state* captures attention while others are in different states in all of the experiments.

An important observation is that, if the time interval between two state transitions (not for the same object) is smaller than 0.3 seconds, the one which is later set was not captured by the subjects. In case two (Fig. 2-b), for instance, where all objects start to move in sequential order with 0.3 seconds intervals, none of motion onsets truly captured attention of the subjects. This is an expected behavior since following the first event, human attentional mechanism remains on this location in the first 0.3 seconds, causing new events to be discarded [15]. After this very short period, inhibition of return (IOR) mechanism slows the response to the current focus of attention enabling previously unattended objects to be attended. This decay time for the state is 0.9 seconds. After 0.9 seconds, the effect of state disappears [14] (Fig. 2-b,d,e,f).

Gaze is transferred to the closest object upon multiple events attracting attention. In the final case (Fig. 2-h), we observed that if more than one motion change, object appearance, motion onset, or motion change appears at the same moment, subjects gazes are commonly transferred to the closest object to the current gaze point.

Lastly, if more than one object have the same state with the same speed and the same direction of motion, they are recognized as a single object according to the Gestalt principles introduced in the previous section. In the case eight, subjects did not check each object separately in that moments, instead they looked at a point in the middle of this object group having the same motion direction and speed.

3.2   Overview

The overview of the proposed model is shown in Figure 3 and it contains two main parts. In the first part, individual motion saliencies of the objects are calculated. In the second part, the relations among the object are examined and final focus of attention is decided.

3.3   Individual Motion Saliency

Based on the observations from the pre-experiments and findings from the literature in psychology, we build an individual motion saliency model. The proposed model calculates instant saliency values for each object in an animation. Once the motion state of an object is detected, its saliency value is calculated as a time dependent variable.
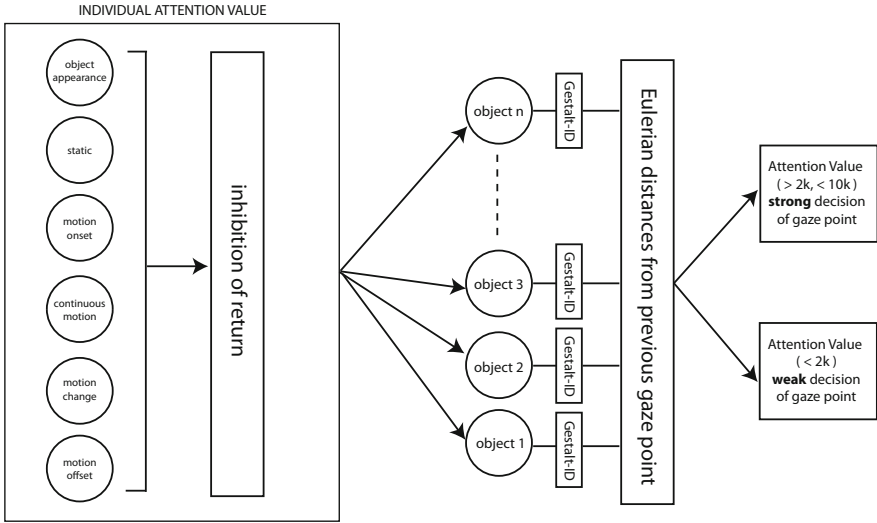
Fig. 3. Overview of the proposed model

We define an initial saliency value for each motion state according to the dominancy among the states (Fig. 4). These initial values are the peak saliencies and decays in time. Although these constants do not reflect an exact proportional dominancy result of the states among each other, they are used to make a correct decision with respect to their attentional dominancy. The initial saliency values for three most dominant states motion onset, motion change and object appearance are assigned as $10k$ while they are $2k$ for motion offset and continuous motion. For static state, it is $1k$. Usage of k as a coefficient enables converting the calculated saliencies for each object to probabilities defining the chance of getting the observers' gazes.

For each visible object, calculated individual saliency values change between $1k$ and $10k$ according to the formulas shown in Table 1 where t stands for the elapsed time after a state is initialized. Obviously, the saliency value for invisible objects is zero.

Saliency of an object in the static state is always $1k$, because we observed it never captures attention among other states. Also, saliency value is permanently $2k$ for continuous motion since a moving object may get attention anytime if nothing interesting, e.g. states of all other objects are static, happens on the screen and subject is not performing a target search [16]. Likewise, motion offset could get attention over static or continuous motion. However, it is under decay with the effect of IOR until the state becomes static.

Motion onset changes to continuous motion state; therefore, its value decays to $2k$ with IOR. It is exactly the same for motion change state. Hence, during inhibition of return (0.3sec$< t <$0.9sec) [5] [6], attention value decays linearly from
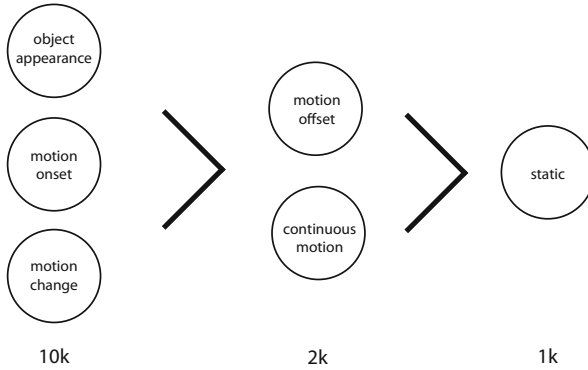
Fig. 4. Attentional dominancy of motion states over each other. Object appearance, motion onset and motion change are the most dominant states capturing attention. Motion offset and continuous motion is slightly more dominant over just static state.

Table 1. Individual Attention Values

States	$t \leq 300ms$	$300ms < t < 900ms$	$900ms \leq t$
STATIC	1k	1k	1k
ONSET	10k	$10k(1.25 - 7t/6)$	2k
OFFSET	2k	$10k(1.25 - 5t/6)$	1k
CONTINUOUS	2k	2k	2k
CHANGE	10k	$10k(1.25 - 7t/6)$	2k
APPEARANCE	10k	$10k(1.3 - 4t/3)$	1k

$10k$ to $2k$ as shown on Table 1. Differently, object appearance decays to $1k$ because its state may become static. Similarly, motion offset decays linearly from its peak value $2k$ to $1k$, since its following state is static.

Inhibition of return is not applied to any of the states if the elapsed time on a state is smaller than 300ms. As we mentioned earlier, during this time other state changes are not captured by subjects in the experiments.

This model provides calculating individual motion saliency values of each object in real time. However, it is still not sufficient to decide the focus of attention without examining the relationships of the objects among each other.

3.4   Global Attention Value

So far we have shown a model to calculate individual attention values of the objects. On the other hand, our visual system does not interpret objects in the scene individually. It tries to represent any stimulus in the simplest way. As Gestalt psychologists concluded that all similar objects in our vision are grouped

and perceived as a simple object [3], we included a Gestalt organization into
our model. Each object has a *Gestalt ID*. The objects having identical motion
direction and speed are labeled with the same Gestalt ID. The pseudocode for
this procedure is as follows.

```
SetGestaltIDs(objects[]){
  curGestalt = 1;
  for i = 1 to NUMBEROFOBJECTS
    gestaltSet = false;
    for j = 1 to i-1
      if objects[j].velocity equals to objects[i].velocity
        objects[i].gestaltID = objects[j].gestaltID;
        gestaltSet = true;
        break;
    if gestaltSet is false
      objects[i].gestaltID = curGestalt;
      curGestalt++;
}
```

All objects with the same Gestalt ID obtain the highest individual saliency
among this group. In the pre-experiments, we clearly observed that subjects'
gaze circulates around not one object if the objects have the same speed and
direction with some other objects.

Another problem that is not solved by the individual saliency model is the case
of equivalence. If more than one object with different Gestald IDs have the same
highest individual attention value, which object will be chosen as the possible
gaze point is not addressed. An observation we made during pre-experiments
suggest a solution to this problem. If more than one states were set at the same
time, subjects commonly looked at the closest object to the previous gaze point.
Therefore, we included an attribute to our model to consider Eulerian distance of
each object from previous decision of gaze point. It is calculated as the pixelwise
Eulerian distance from the previously decided focus of attention in the screen.

Finally, if there are multiple Gestalt groups with the highest calculated
saliency, the closest one to the previously decided focus of attention is selected
as the current focus of attention.

4   Experiment

Experiment Design. We have performed a formal user study to validate our
model. In the experiment, subjects have looked at a 22" LCD display where
twenty spheres animate for two minutes in a 3D room (Fig. 5). For each half-
second of animation, a random object changes its motion state to another ran-
dom state. The subjects' task is pressing a button when the color of a sphere
becomes the target color which is shown to the subjects during the experiment.
Interval time for two color change is at least three seconds. In order to avoid
color or shape related bias, all spheres have the same size and randomly selected
colors having close luminances. The reason for selecting different colors for each

sphere is to decrease the pop-out effect for the spheres that change color. During an animation, we have three cases of color changes. The first case is the color change of highly salient spheres, which are chosen with strong decisions of our model. In second case, among spheres having the lowest saliencies, the most distant spheres to the previously calculated gaze points are chosen. In third case, we choose the sphere in a fully random fashion. All cases are shown to the users in a mixed manner multiple times.

Fig. 5. Sample screenshot from the experiment

16 voluntary graduate or undergraduate level students (4 females, 12 males) whose average age is 23.75 attended to our experiment. All subjects have normal or corrected to normal vision and they are not informed about the purpose of the experiment. The experiment is introduced to the subjects as a game. In the game, to get a higher score, they should press the button as soon as possible when an object changes its color to the target color. Before starting the experiment, each subject performed a trial case to learn how to play the game.

Experiment Results. The results of the experiment can be seen in Fig. 6. We expect the response times to color changes of salient spheres to be shorter since they are expected to occur on the focus of attention. As expected, observers responded to color changes of salient spheres in a shorter time compared to those appeared on lowly salient spheres and randomly selected spheres. The differences for both cases are statistically significant ($p < 0.05$) according to the applied paired t-test.

5 Conclusion and Future Work

For motion saliency in computer animations we analyzed psychological findings on the subject by conducting an eye-tracking experiment and developed a decision theoretic approach to momentarily determine perceptually significant objects. Our observations from the experiment show that each phase of the motion has a different impact on perception. To that end, we defined six motion states and assigned attention values to them by considering their dominance. Individual attention values are determined for each object by considering their
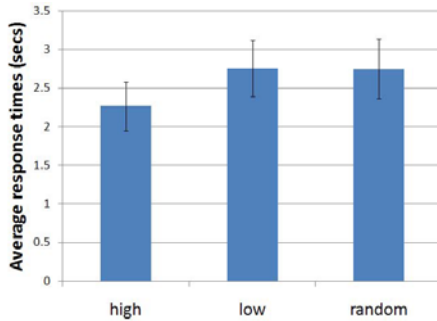
Fig. 6. The results of the experiment. Error bars show the 95% confidence intervals.

current state and elapsed time after a state initialization to include the impact of IOR. We elaborated our model by including the relationships of the objects on the scene with each other using Gestalt principles. Overall model makes decisions for the most salient object and its position is predicted as the gaze point. We carried-out a final experiment to evaluate the effectiveness of decisions our model make.

One of the limitations of our approach is the consideration of only bottom-up and stimulus-driven attention like most of other saliency works. Our final experiment had promising results although it included a simple task for users to search. On the other hand, more complex tasks could remove the effect of motion saliency. The model should be evaluated with further experiments and improved by the impact of new task-dependent cases. Furthermore, motion saliency and other object based saliency methods considering object material and shape should be compared. In our model we only consider saliency caused by motion. Dominancy of other saliency parameters on motion saliency is not evaluated in this paper.

References

1. Abrams, R.A., Christ, S.E.: Motion onset captures attention. Psychological Science 14, 427–432 (2003)
2. Andersen, R.A.: Neural mechanisms of visual motion perception in primates. Neuron 26, 776–788 (1997)
3. Arnheim, R.: Art and Visual Perception, pp. 50–60. University of California Press (1954)
4. Bulbul, A., Koca, C., Capin, T., Güdükbay, U.: Saliency for animated meshes with material properties. In: Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization (APGV 2010), pp. 81–88. ACM, New York (2010)

5. Daly, S.: Engineering observations from spatiovelocity and spatiotemporal visual Models. In: Proc. SPIE, vol. 3299, pp. 180–191 (1998)
6. Halit, C., Capin, T.: Multiscale motion saliency for keyframe extraction from motion capture sequences. Computer Animation and Virtual Worlds 22(1), 3–14 (2011)
7. Hillstrom, A.P., Yantis, S.: Visual motion and attentional capture. Perception - Psychophysics 55(4), 399–411 (1994)
8. Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 20(11), 1254–1259 (1998)
9. Kelly, D.H.: Motion and vision II. Stabilized spatio-temporal threshold surface. J. Opti. Soci. Ameri. 69(10), 1340–1349 (1979)
10. Koffka, K.: Principles of Gestalt psychology, pp. 106–177. Harcourt Brace, New York (1935)
11. Lee, C.H., Varshney, A., Jacobs, D.W.: Mesh saliency. In: ACM SIGGRAPH 2005, pp. 659–666. ACM, New York (2005)
12. Livingstone, M.: Vision and Art: The Biology of Seeing. ABRAMS, 46–67 (2002)
13. Peters, R.J., Itti, L.: Computational mechanisms for gaze direction in interactive visual environments. In: Proceedings of the 2006 Symposium on Eye Tracking Research & Applications (ETRA 2006), pp. 27–32. ACM, New York (2006)
14. Posner, M., Cohen, Y.: Components of visual orienting. In: Attention and Performance X, pp. 531–556 (1984)
15. Torriente, I., Valdes-Sosa, M., Ramirez, D., Bobes, M.: Vision evoked potentials related to motion onset are modulated by attention. Visual Research 39, 4122–4139 (1999)
16. Yantis, S., Hillstrom, A.P.: Stimulus driven attentional capture: Evidence from equiluminant visual objects. Journal of Experimental Psychology: Human Perception and Performance 20, 95–107 (1994)
17. Yee, H., Pattanaik, S., Greenberg, D.P.: Spatiotemporal sensitivity and visual attention for efficient rendering of dynamic environments. ACM Transactions on Graphics 20(1), 39–65 (2001)