An Automated High Level Saliency Predictor for Smart Game Balancing

George Alex Koulieris, George Drettakis, Douglas W. Cunningham, Katerina Mania

An Automated High Level Saliency Predictor for Smart Game Balancing

GEORGE ALEX KOULIERIS, Technical University of Crete, Greece
GEORGE DRETTAKIS, Inria[1]
DOUGLAS CUNNINGHAM, Brandenburg Technical University Cottbus, Germany
KATERINA MANIA, Technical University of Crete, Greece

Successfully predicting visual attention can significantly improve many aspects of computer graphics: scene design, interactivity and rendering. Most previous attention models are mainly based on low-level image features, and fail to take into account high-level factors such as scene context, topology or task. Low-level saliency has previously been combined with task maps, but only for predetermined tasks. Thus the application of these methods to graphics – e.g. for selective rendering – has not achieved its full potential. In this paper we present the first automated high level saliency predictor incorporating two hypotheses from perception and cognitive science which can be adapted to different tasks. The first states that a scene is comprised of objects expected to be found in a specific context as well objects out of context which are salient (scene schemata) while the other claims that viewer's attention is captured by isolated objects (singletons). We propose a new model of attention by extending Eckstein's Differential Weighting Model. We conducted a formal eye-tracking experiment which confirmed that object saliency guides attention to specific objects in a game scene and determined appropriate parameters for a model. We present a GPU based system architecture that estimates the probabilities of objects to be attended in real-time. We embedded this tool in a game level editor to automatically adjust game level difficulty based on object saliency, offering a novel way to facilitate game design. We perform a study confirming that game level completion time depends on object topology as predicted by our system.

Categories and Subject Descriptors: H.1.2 [**Models and Principles**] User/Machine Systems, Human Factors; I.3.3 [**Computer Graphics**] Picture/Image Generation; I.3.7 [**Computer Graphics**] Three-Dimensional Graphics and Realism, Virtual Reality; I.6.0 [**Computer Graphics**] Simulation and Modeling, General

General Terms: Human Factors

Additional Key Words and Phrases: Computer Graphics, Game Balancing, Scene Schemata

1. INTRODUCTION

The prediction of attention can significantly improve many aspects of computer graphics and games. For example, image synthesis can be accelerated by reducing computation on non-attended scene regions [Cater et al. 2003]; attention can also be used to improve Level-of-Detail (LOD) [Lee et al. 2009]. Another interesting case is computer game design. Many game genres rely on a search or target detection task to solve riddles or find game objects. If attention can be automatically predicted, several tasks in game design would be simplified. For example adjusting the difficulty of a game level could be facilitated by relocating objects estimated to attract attention [Feil and Scattergood 2005].

Existing visual attention models such as Feature Integration Theory (FIT) are mostly driven by low level image features such as contrast, luminance and motion [Treisman and Gelade 1980]. FIT is a commonly used model of attention in computer graphics [Itti and Koch 2001; Longhurst et al. 2006]. However, it often fails to predict saccadic targets [Borji and Itti 2013] because high-level properties such as scene semantics and task strongly affect the planning and execution of fixations [Henderson and Hollingworth 1999; Einhäuser et al. 2008; Borji and Itti 2013]. There has been previous work

[1] Inria authors are at REVES/Inria Sophia-Antipolis Méditerranée.

accounting for task when modelling goal-oriented attention in computer graphics, but the task has always been predetermined [Cater et al. 2003; Sundstedt et al. 2004; Sundstedt et al. 2005].

Our goal is thus to develop an automated high-level saliency predictor which can be adapted to different tasks. In this paper we present the first such predictor incorporating two hypotheses from perception and cognitive science: the *scene schema hypothesis* and the *singleton hypothesis*. The scene schema hypothesis states that a scene is comprised of objects we expect to find in a specific context and salient objects that are not expected in a scene (see Figure 1) [Bartlett 1932; Henderson et al. 1999; Hwang et al. 2011]. The singleton hypothesis states that the viewer's attention is ordinarily captured by stimuli that are locally unique in a basic visual dimension such as color, orientation, etc. i.e. isolated [Theeuwes and Godijn 2002]. In our work, the singleton state is also a context dependent measure not purely image-driven: Figure 2 shows that the spatially isolated vase attracts attention, though not salient in terms of color.

We propose a new model by incorporating the schema [Bartlett 1932; Henderson et al. 1999; Hwang et al. 2011] and singleton [Theeuwes and Godijn 2002] hypotheses into the Differential-Weighting Model (DWM) [Eckstein 1998; Eckstein et al. 2002; Eckstein et al. 2006] that employs Bayesian priors. To find the parameters of this model, we perform several perceptual experiments, which also verify that high-level saliency guides attention. Using this new model, we estimate the posterior probability that a viewer will fixate an object based on high-level contextual features, independent of the viewer's task [Eckstein et al. 2006]. We use our automated high level saliency predictor to facilitate game level balancing, offering a novel way to ease game design.

We make three primary contributions:

—We propose a new model to account for high-level object saliency as predicted by the scene schema and singleton hypotheses by extending the Differential Weighting Model using Bayesian priors.

—In three perceptual experiments, we verify that high-level saliency guides attention and we obtain perceptual parameters to calibrate our model.

—We develop a tool based on the model to automatically predict high-level saliency in real-time. We then validate the tool's efficacy in helping to adjust game difficulty in a game-level editor.

We used a modern game engine for our experiments, our game level design and its validation. This choice underlines the relevance of our results for realistic use cases.

2. RELATED WORK

Visual Attention Visual perception can be thought of as the active extraction and manipulation of environmental information. The visual perception pipeline starts with low-level processes which extract simple image regularities such as edges or color [Marr 1982]. Subsequently, mid-level processes combine these properties to form higher-level features such as the shape of an object [Shipley and Kellman 2001]. Finally, high-level processes map these mid-level features to meaning and semantics [Palmer 1999]. To efficiently concentrate the limited brain resources of the mid- and high-level processes on those few low-level features that are likely to be important, the human brain is equipped with a selection mechanism known as focal attention. Some low-level features such as edges can automatically attract focal attention in an almost reflex-like fashion [Koch and Ullman 1987]. Likewise, mid- and high-level features as well as goal-oriented properties can direct focal attention [Henderson et al. 1999; Yarbus et al. 1967]. For example, the contextual validity or appropriateness of an object's location will affect visual search; when looking for a chimney, usually we direct our gaze first to the rooftops. The fundamental question of how the visual system combines the influence of low-, mid-, and high-level components is a challenging research issue and remains largely unanswered [Theeuwes 2010]. A recent review of the theories can be found in [Borji and Itti 2013].

The most common form of focal attention model is the two-stage model, such as the popular Feature Integration Theory (FIT) [Treisman and Gelade 1980]. In two-stage models, a privileged set of low-level features are initially extracted everywhere in an image in parallel. The focal attention mechanism then selects a few locations in the image based on these features for further processing. In the second stage, the low level features at the selected locations are integrated and subjected to further processing in a slow, serial (i.e., one region at a time) fashion. A widely used saliency model inspired by FIT [Itti and Koch 2001] employs low-level features such as contrast, luminance, and motion to determine which areas are likely to attract attention. Although FIT plausibly emulates many aspects of focal attention, it has been shown that: (i) complex stimuli such as surfaces are processed simultaneously and not in a serial fashion [Nakayama et al. 1986], (ii) visual attention is directed to objects in a scene rather than their low level visual attributes [O'Craven et al. 1999] and, (iii) observers may achieve multiple simultaneous foci of attention in the visual field, not supported by FIT [Awh and Pashler 2000]. In other words, attention models based on low-level features often fail to predict saccadic targets [Borji and Itti 2013], in part because they do not take into account high level factors such as scene context, task, or object topology [Einhäuser et al. 2008; Henderson and Hollingworth 1999; Rensink 2000].

Two phenomena within the perception literature point to specific roles that high-level information can play in focal attention. The first – the *scene schema* effect – is based on the observation that a high proportion of objects in a scene can usually be expected to be found there. They are "consistent" with the scene. Sometimes, however, objects are in a scene or a location that is very atypical. Such "inconsistent" objects are potentially salient (see, e.g., Figure 1) [Bartlett 1932]. Research has shown that previously-acquired knowledge of stereotypical object placement in a scene combined with the on-going visual experience of a scene can attract focal attention [Brewer and Treyens 1981; Bar et al. 1996; Henderson et al. 1999]. The ratio and location of consistent and inconsistent objects in a specific context can also influence whether the scene is perceived to be congruent overall [Einhäuser et al. 2008; Rayner 2009; Hwang et al. 2011]. The second effect – the *singleton effect* – refers to the finding that stimuli that are locally unique in terms of color or orientation capture attention (Figure 2) [Theeuwes and Godijn 2002]. Object perception is based on context-dependent processing of low-level variables i.e. pixels, therefore the singleton state is a high level semantic property of spatially isolated objects.

More recently, a number of single stage models have been proposed, which are very effective at describing visual attention. For example, Eckstein has proposed a single-stage model of attention called the Differential-Weighting Model (DWM), [Eckstein 1998; Eckstein et al. 2002; Eckstein et al. 2006] which incorporates both low-level features as well as prior knowledge about scene context. The DWM models attentional processing using physiological noise in brain neurons and Gaussian combination rules. Contextual information in the DWM is embodied in the Bayesian priors provided to the model beforehand. For example, when searching for a chimney in a picture that contains a house, the visual elements depicting the roof of the house are given a higher prior probability than other scene elements. DWM has never been used to predict high-level saliency or gaze patterns in interactive Virtual Environments (VEs) incorporating scene schemas and singletons.

Attention in Computer Graphics In an effort to predict attention in pre-determined task areas, it has been shown that task importance maps may be used to accelerate rendering by reducing quality in regions that are unrelated to a given task [Cater et al. 2003]. Selective rendering guided by a FIT-based saliency model renders perceptually important parts of a scene in high quality while the remaining areas of the image are rendered at lower quality, thus saving in computational cost [Longhurst et al. 2006]. As mentioned, FIT only uses low-level image characteristics. Other research has combined task maps with a low-level saliency map and validated the results using eye-tracking [Sundstedt et al. 2004; Sundstedt et al. 2005]. Predicting gaze behavior in games may be used to optimize the distri-

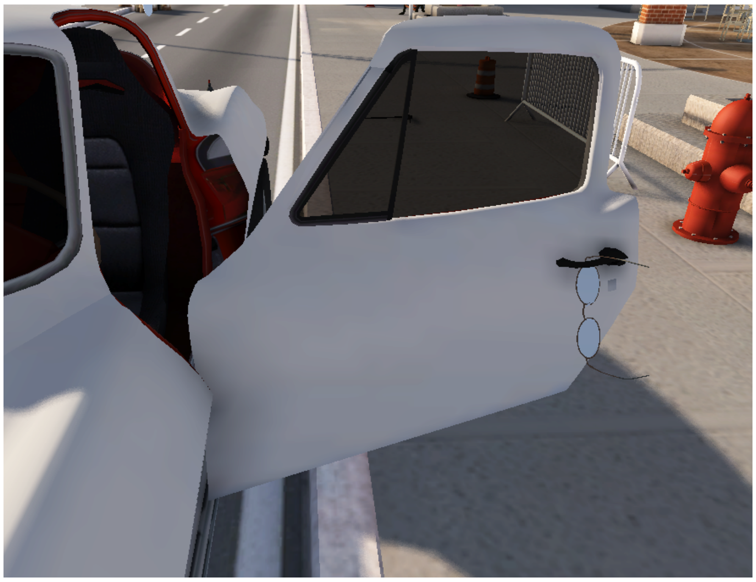

Fig. 1. The spectacles attract attention as they are *inconsistent* with the car door context.

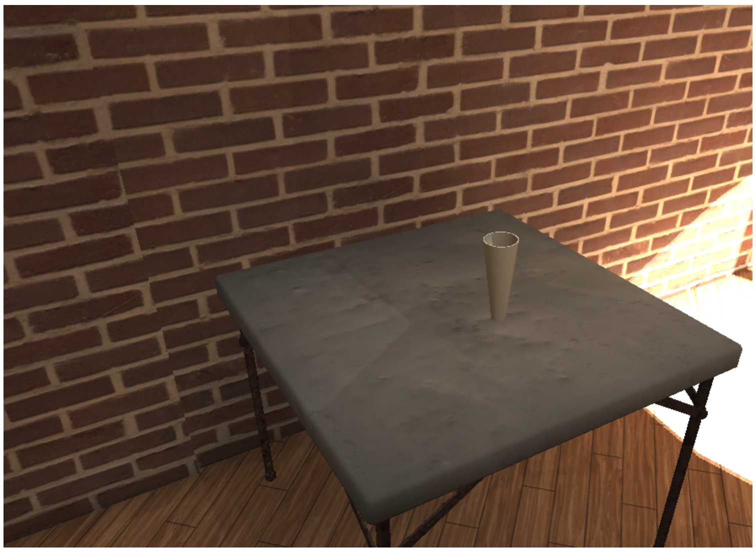

Fig. 2. The spatially isolated vase attracts attention as it is a *singleton* object.

bution of computing resources [Sundstedt et al. 2008]. Saliency models and task related data have been linearly combined to track visually attended objects in a VE in task-specific areas [Lee et al. 2009]. Task relevant gaze behavior associated to first-person navigation in a virtual environment has been estimated by combining bottom-up and top-down components to compute user gaze point position on screen [Hillaire et al. 2010]. Attention in games may also get manipulated. A guiding principle and method based on the Guided Search theory [Wolfe 1994] has been proposed to direct attention to target items that should be noticed by an observer in a video game e.g. an advertisement. When a frequently searched game object is modified to share perceptual features such as color or orientation with a target item, the item will attract attention [Bernhard et al. 2011]. Saliency models have been employed to animate the gaze behavior of virtual characters [Oyekoya et al. 2009] and crowds [Grillon and Thalmann 2009].

Although task-based saliency estimations competently predict salient regions in pre-determined task-specific areas [Cater et al. 2003], the challenge is to estimate salient regions in all areas of a scene for different tasks via an integrated model. Research in interactive VEs has confirmed that attention is influenced by the semantic context of objects in the form of scene schemas [Mania et al. 2005; Mourkoussis et al. 2010; Zotos et al. 2009]. In one step towards implicitly modelling high-level effects, machine learning techniques have been applied to eye tracking data in order to train a model to detect salient regions in a pre-defined set of static photographs [Judd et al. 2009]. A pipeline to derive gaze prediction heuristics from eye-tracking data for 3D Action Games has been proposed [Bernhard et al. 2010]. To date, a model that explicitly links in a physiologically plausible manner experimental outcomes on attention with object saliency is missing.

We propose an innovative visual attention model based on DWM which takes into account high level information about the context of the scene. Our new model can be directly used in computer graphics. We validate our theoretical hypotheses in a formal perceptual study involving game level balancing.

Deployment of Attention in Computer Games Gameplay greatly depends on attention deployment [Sundstedt et al. 2013]. Eye tracking data has revealed that players playing First Person Shooter games tend to concentrate on the center of the screen searching for enemies while in an Action-Adventure game players mostly explore the entire screen for game props to advance the gameplay [El-Nasr and Yan 2006].

Player enjoyment is crucial for the success of a computer game. An enjoyable/optimal experience, also termed *flow*, is shown to be so satisfying that players take pleasure in the game with little concern for what they will get out of it [Czikszentmihalyi 1990]. Enjoyable experiences in games arise primarily from challenge [Sweetser and Wyeth 2005]. Challenge refers to the ability of a game to be sufficiently intriguing and match the player's skill level [Pagulayan et al. 2003; Desurvire et al. 2004]. Improper challenge levels provoke anxiety in a discouragingly hard game or apathy in a boringly easy game [Johnson and Wiles 2003].

Looking for an object is a common task in Adventure or Action-Adventure video games, often guiding level advances. The time spent searching for an object in a game should be in proportion to the advantage it conveys in game play. Designers mostly rely on their experience and instinct while calculating cost/benefit ratios by manually placing objects and obstacles in their levels [Pagulayan et al. 2003]. Multiple rounds of Play-Testing and observation can stabilize choices in a level [Sweetser and Wyeth 2005]. However, because players' abilities vary and play-testers are not abundant to every game designer, a sophisticated approach such as the model we propose, that guides automatic object manipulation and game balancing based on high-level visual attention is crucial.

3. HIGH LEVEL SALIENCY MODELING

In this section, we present our new model of high-level attention. Before presenting our new model, we first describe the DWM. We then explain how we extended DWM by encoding the interaction of schemas and singletons based on the Bayesian priors of the original model.

3.1 The Differential-Weighting Model

The Differential-Weighting Model (DWM) [Eckstein 1998; Eckstein et al. 2002; Eckstein et al. 2006] estimates the interaction between visual evidence concerning a target in a scene and Bayesian prior probabilities indicating expectation and context of a scene. By combining sensory data with existing knowledge it calculates the posterior probability that a location will be fixated in a visual search task and thus predicts saccadic targeting.

DWM assumes that when searching for a target, each location in a scene elicits neuronal activity in relevant sensory units of each visual feature. This response is subject to Gaussian independent neutral noise, i.e. the outcome of the perceptual processing of this response is probabilistic. When a sensory unit is tuned to observe a specific feature, it responds at a higher rate when the observed feature is present. Neurons are subject to internal noise and have a response following a Gaussian distribution [Tolhurst et al. 1983]. After many trials, Figure 3 depicts the internal response probability density functions for noise-alone (left curve) and for signal-plus-noise trials (right curve). The model calculates the ratio of the joint likelihood of observing the feature's neural responses in each image region given that the target is present and the joint likelihood of observing the feature's responses given that the target is absent according to a selected probability. This noisy response is then weighted by context effects encoded in Bayesian priors relevant to specific stimuli. The Bayesian priors embody the probability of these stimuli to co-occur with other highly visible visual features of the image.

For each image frame f and each visual field location (x, y), each sensory unit responds in a noisy manner for each feature λ_j. DWM calculates the likelihood $l_{j,x,y,f}$ of observing the response λ_j given the presence of the target's j^{th} feature at that location and the likelihood of the response given the absence of the feature. The response has a Gaussian distribution [Tolhurst et al. 1983] with a mean of d'_j and a standard deviation σ. The likelihood $l_{j,x,y,f}$ that the j^{th} sensory unit takes a value $\lambda_{j,x,y,f}$ given the presence of the target's j^{th} feature at (x,y) on frame f is then

$$l_{j,x,y,f}(\lambda_{j,x,y,f}|s) = \frac{1}{\sqrt{2\pi\sigma^2}} exp - \left(\frac{(\lambda_{j,x,y,f} - d'_j)^2}{2\sigma^2} \right) \quad (1)$$

s stands for signal and denotes the presence of the target.
The likelihood that the j^{th} sensory unit takes a value λ_j given the absence of the target's j^{th} feature is

$$l_{j,x,y,f}(\lambda_{j,x,y,f}|n) = \frac{1}{\sqrt{2\pi\sigma^2}} exp - \left(\frac{(\lambda_{j,x,y,f})^2}{2\sigma^2} \right) \quad (2)$$

n stands for noise and denotes the absence of the target.
A likelihood ratio LR [Green et al. 1966] can be calculated as

$$LR_{j,x,y,f} = \frac{l_{j,x,y,f}(\lambda_{j,x,y,f}|s)}{l_{j,x,y,f}(\lambda_{j,x,y,f}|n)} = exp\left(\frac{\lambda_{j,x,y,f} d'{}_j - 0.5 d'{}_j^2}{\sigma^2} \right) \quad (3)$$

3.2 A New High-level Attention Model

We propose a new model by integrating high-level information implied from semantic (schema inconsistency, Figure 1) or physical (singletoness, Figure 2) context represented by Bayesian priors in the DWM. We assume that (i) the internal response associated with high level saliency components is also subject to noise, (ii) dedicated, high-level sensory units are analogous to low-level sensory units. The high-level units fire at their highest rate when fed with the correct high-level feature, much as a low-level edge-detection unit reacts highest when an edge with the proper orientation is presented [Eckstein 1998; Eckstein et al. 2002; Eckstein et al. 2006]. Whether the neural mechanism underlying a high-level sensory unit is a single neuron or a cluster of neurons does not matter. What matters is that there is an internal (neural) state reflecting whether this high-level feature is present or not. For

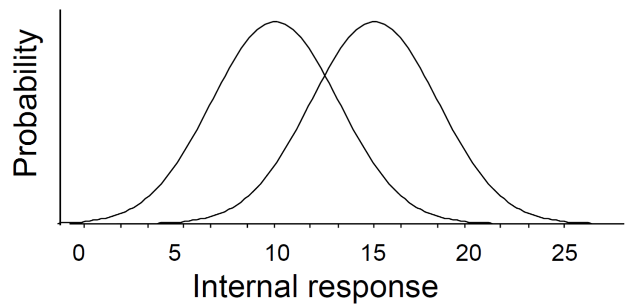

Fig. 3. Internal response probability density functions for noise-alone (left curve) and for signal-plus-noise trials (right curve).

example, we propose a sensory unit that monitors the degree to which an object is isolated. Such a unit would fire when a singleton object is in the field of view [Steinmetz et al. 2000].

We extended the original DWM equations to describe two high-level sensory units tuned to schema inconsistencies and the singleton state of objects. Equations $(1)-(3)$ assume that the internal response generated by the presence of each visual feature is known a priori. Since neuronal response strength is unknown concerning scene schemata and singletons, we alter the DWM and instead calculate the posterior probability that the target is present at each pixel as a sum of K different feature strengths d'_k associated with scene schemata and singletoness.

$$P_{semantic,x,y,f}(s|\lambda) = \sum_{k=1}^{K} LR_{semantic,x,y,f,k} \tag{4}$$

$$P_{physical,x,y,f}(s|\lambda) = \sum_{k=1}^{K} LR_{physical,x,y,f,k} \tag{5}$$

We then average the components (4), (5) using weights $w_{semantic}$ and $w_{physical}$ that we obtain from perceptual studies (see Section 4) to calculate the posterior probability that a location will be attended. A linear combination of components is a common practice in saliency detection algorithms [Frintrop et al. 2010].

$$P_{x,y,f} = w_{semantic}P_{semantic,x,y,f} + w_{physical}P_{physical,x,y,f} \tag{6}$$

As an example consider a bar counter. A coffee mug which is consistent with the context and a medical kit which is inconsistent with the context are shown in Figure 4 (top). Consider a sensory unit that tracks schema inconsistencies. The $\lambda_{semantic}$ of the image regions corresponding to the medical kit is higher than the λs associated to the mug and counter. The $\lambda_{semantic}$ communicates a subjective rating of consistency, e.g. the higher the number, the more inconsistent the object is in relation to the context (Figure 4 bottom left). Let us assume that $K = 1$, $d'_{semantic} = 0.6$ and $\sigma = 0.2$. Because the medical kit is inconsistent, we assume that $\lambda_{semantic} = 1.0$, similarly for the mug $\lambda_{semantic} = 0.16$, for the bar counter $\lambda_{semantic} = 0.22$ because they are both highly consistent. The likelihoods ratios of observing the medical kit, mug and counter are $LR_{medikit} = 36315.5$, $LR_{mug} = 0.1$, $LR_{counter} = 0.3$ respectively as derived from equation (3). The schema inconsistency unit would then estimate the medical kit as the most salient (Figure 4 bottom right). Similarly, $\lambda_{physical}$ is used to calculate the likelihood ratios of observation based on whether an object is placed as singleton.

Fig. 4. A bar counter context (top), $\lambda_{semantic}$ visualization (bottom left) and highest $LR_{semantic}$ highlighted (bottom right).

4. REAL-TIME EVALUATION OF HIGH LEVEL SALIENCY COMPONENTS

We examined the real-time effect of singleton and scene schemata on gameplay for two reasons:

—The role of scene schemata and singletons in interactive, synthetic environments is unknown, even though their effects are well-documented for target detection displays or static real photographs [Henderson et al. 1999; Rensink 2000; Einhäuser et al. 2008; Theeuwes and Godijn 2002].

—Our extension of the DWM requires an empirical classification of objects in relation to scene schemata and the determination of the weighting factors w_j that signify the interaction between semantic (scene schemata) and physical context (singletons).

Inspired by Adventure games [Ju and Wagner 1997], a suitable game genre to apply our method, we designed an environment that allows us to investigate the impact of high-level saliency on visual attention & gameplay and recorded the time it took to search for plot-critical objects. The storyboard was based on the popular video game L.A. Noire^{TM}, a 2011 Action-Adventure neo-noir crime video game developed by Team Bondi^{TM} and published by Rockstar Games^{TM}. A scene depicting a Coffee Shop inspired by the "Driver's Seat" case of the game was heavily modified to include multiple areas representing a car schema and a cafeteria schema inclusive of sub-schemata representing a coffee shop counter and a lounge loft. We systematically controlled the semantic and physical states of plot-critical objects. Each object could be in a schema-consistent or a schema-inconsistent location, and could be in either a singleton state (positioned by itself) or a compound state (positioned in cluttered surroundings) (please see accompanying video).

4.1 Experiment 1: Defining object consistency

Here, we empirically classify scene objects as either consistent or inconsistent in relation to the context of each part of the scene. Specifically, a list of 50 objects was assembled and given to 21 graduate students (14 male, 7 female). Each participant used a 7-point Likert scale to rate how likely each item was to appear in a given scene. A rating of 7 meant that the object was very much expected to be in that location and 1 meaning the object was very much not expected. Half of the objects were tested in the Coffee Shop counter context and the other half were tested in the car context. We then selected a set of consistent objects from the high end of the scale and a set of inconsistent ones from the low end (based on the approach used in [Brewer and Treyens 1981]). The classification of objects in relation to scene schemata is independent from a specific game scenario, i.e. a teapot is consistent with a kitchen context irrespectively of a background story. A taxonomy of common objects in relation to scene schemata that can be used in any game will be provided as part of the production level version of our system.

4.2 Experiments 2 & 3: Determining the Roles of Semantic and Physical Context

In Experiments 2 and 3 we examine the effect of physical (singletoness) and semantic (consistency effects) manipulations on game task completion time for two common tasks appearing in (Action-)Adventure games. In both tasks, the same general scenario was used: "*Adrian Black, a married man and a barista at the Coffee Shop decides to start a new life with his customer Nicole staging his own murder to cover a getaway with her.*" Participants were instructed in both tasks to find three decisive objects as quickly and as accurately as possible in order to solve the mystery (Figure 5). Experiment 2 used a *Search* task (participants knew exactly what they were searching for). Experiment 3 used a *Non-Search* task (participants did not know what they were searching for, and as such were exploring the environment with less of the specific purpose).
Our two main predictions are:

—Singleton objects will require less time to be recovered compared to objects in compound state because they capture attention no matter what the task is [Theeuwes and Godijn 2002].

—When searching for an object, consistent locations will attract attention and therefore will require less time to be recovered than inconsistent locations. When not searching, on the other hand, objects at inconsistent locations should attract attention and therefore will require less time to be recovered compared to consistent locations [Eckstein et al. 2006].

Fig. 5. One of the desicive objects, the spectacles, as positioned in different conditions: Consistent/Compound (left) vs Inconsistent/Singleton (right).

4.2.1 *Method.* Each of the two main factors (Semantic Context and Physical Context) had two levels, which were factorially combined to produce four experimental conditions: Consistent/Compound, Inconsistent/Compound, Consistent/Singleton and Inconsistent/Singleton object placement. The objects were positioned so as to maintain constant navigation time while reaching them across conditions and on similar visual angles within the VE. The four conditions above were the same for both experiments. A between-participants design was used, meaning that each person participated in only one experiment and in only one experimental condition.
Participants A total of 80 participants (56 male, 24 female; ages between 21 - 33) were recruited from the undergraduate and research population of our institution and were rewarded with pastry for their participation. All participants were familiar with first person perspective navigation and had normal to corrected vision. Upon arrival, each participant was randomly assigned to one of eight groups so that each group had 10 participants. Each group participated in only one of the experimental conditions.
Procedure and Apparatus Upon arrival, the participants signed a consent form and were then allowed to practice navigating in a training scene. The participants were then informed of the experimental scenario and positioned about 60cm from a 20" flat screen monitor (screen width of 44cm) at a resolution of 1680x1050. The game environment was rendered in real-time at a 60Hz constant refresh rate. First person viewing mode was used for navigation. The virtual camera was positioned at the level of the eyes of the subject's avatar which was 1.80m in height. The avatar had three degrees of displacement freedom. Yaw and pitch angles of the camera were controlled with the mouse, while walking was controlled with the arrow keys of the keyboard. Task completion time as well as inspection start/end timings indicated by a mouse over a possible clue, collect attempts, collected (decisive or not) objects were stored in a database along participants' age and gender.

4.2.2 *Results.* We subjected the completion times to a Multiple Linear Regression (MLR) analysis which, like the ANOVA, is a subclass of *general linear modelling*. Unlike the ANOVA, a linear regression also provides an explicit, quantitative model of how the different experimental factors affect performance along with the relative importance of the different factors [Cunningham and Wallraven 2011]. This information is critical for deriving the DWM weights.

In MLR, the line $y = m_1x_1 + \cdots + m_nx_n + b$ is fit to the data, with y being the participants' performance (e.g., task completion time) and each x_i being an experimental factor (e.g., physical or semantic context) and b being the intercept. Since our two factors are categorical, they must be *dummy coded*. We gave Compound a value of 0 and Singleton a value of 1. Likewise, Inconsistent and Consistent were set to 0 and 1, respectively. Each regression coefficient m_i indicates how many seconds faster a unit change (i.e., from 0 to 1) in the factor x_i will cause the completion time to be. Critically, the ratio of the mean squared prediction error of a model to the variance in completion time is directly related to the Pearson correlation coefficient [Cunningham and Wallraven 2011] and indicates how much of the variance in completion time can be "explained" or predicted by the change in the independent variables. We will use this relative predictive values to derive the DWM weights.
Experiment 2: Search Task On average, participants needed 64.81, 72.10, 135.03 and 164.5 seconds to complete the Singleton/Consistent, Singleton/Inconsistent, Compound/Consistent, and Compound/Inconsistent conditions, respectively (see Figure 6). Regressing physical context onto completion time yields a model that explains 80.7% of the variation in completion time. This is a significant amount, $F_{1,38} = 159.1, p < .001$, showing the significant effect of physical context. There was also a significant effect of semantic context: a two predictor model regressing both physical and semantic context onto completion time explains 84.8% of the variance. This increase in predictive power of 4.1% is statistically significant, $F_{1,37} = 10.068, p < 0.0031$. Finally, the interaction between physical and semantic context was marginally significant: adding a term to capture the variance jointly explained by

semantic and physical context – while controlling for multicolinearity – explains an additional 1.5%, $F_{1,37} = 3.9621, p < 0.055$. The intercept, regression coefficients and statistical significance of each predictor in the two and three predictor models can be seen in Table I. As can be seen in the table, the two predictor model predicts that performance in the Compound/Inconsistent condition should be 158.962 (the intercept) which is close to the actual value of 164.5. Changing from compound to singleton should speed up performance by 81.309 seconds (the regression coefficient for physical context), and changing from inconsistent to consistent should speed up performance by 18.381 seconds. Thus, performance in the Singleton/Consistent condition is predicted to be 59.272, which matches the actual value of 64.81 well.

Experiment 3: Non-Search Task On average, participants needed 89.74, 94, 173.05 and 144.90 seconds to complete the Singleton/Consistent, Singleton/Inconsistent, Compound/Consistent, and Compound/Inconsistent conditions, respectively (see Figure 7). The effect of physical context was again significant; a single predictor model explains 77.7% of the variance, a statistically significant amount, $F_{1,38} = 132.1, p < .001$. Semantic context was also significant; the two predictor model explained 80.2% of the variance, a statistically significant increase of 2.5%, $F_{1,37} = 4.578, p < 0.04$. The interaction was also significant; the three predictor model explains 84.7% of the variance, an increase of 4.5%, $F_{1,37} = 3.258, p < 0.003$. The intercepts and significance of the three predictors can be seen in Table II.

Table I. The regression coefficients and their significance on the overall model, for the case of a Search task

Coefficients	Estimate Time	p-value
Intercept	158.962	< 0.0001
+Singleton placement	-81.309	< 0.0001
+Consistent placement	-18.381	0.003
+Joint Term	22.190	0.055

4.3 Discussion

Both semantic and physical context play a statistically significant role in attention deployment, with physical context playing the dominant role. Moreover, an object is often inconsistent with its surroundings (and thus will probably grab attention) but neither in a singleton state nor salient in terms of low level features. In such cases, the scene schemata theory can predict its prominence. In agreement with our first prediction, placing an object in a singleton state decreased task completion time. The two predictor model indicates that performance in the singleton conditions is about 49% of that in the compound conditions for Search tasks, and about 59% for Non-Search tasks. In agreement with the first part of our second prediction, consistency decreases task completion time for a Search task. The significant interaction for Non-Search tasks, however, means that the effects of semantic was dependent upon physical consistency: inconsistent locations were only faster for compound objects. Contrary to prediction, inconsistency increased search time in a Non-Search task for singleton objects.

Table II. The regression coefficients and their significance on the overall model, for the case of a Non-Search task

Coefficients	Estimate Time	p-value
Intercept	153.008	< 0.0001
+Singleton placement	-67.111	< 0.0001
+Consistent placement	11.944	0.039
+Joint Term	-32.407	0.025

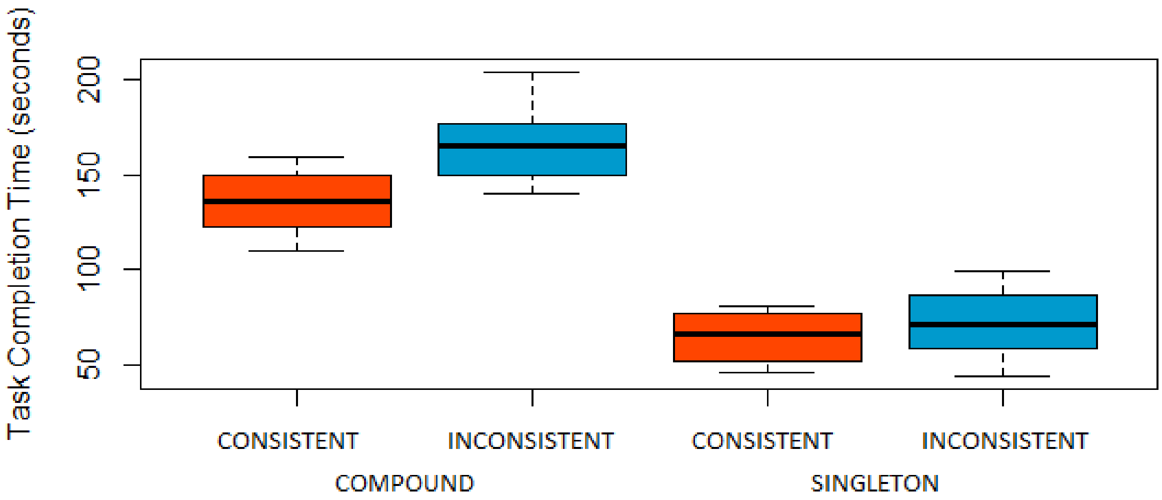

Fig. 6. Task completion time distribution in a Search task. The thick, horizontal line in each box represents the median for that condition. The colored box around the median represents the middle quartiles and the outer bars represent the extremes.

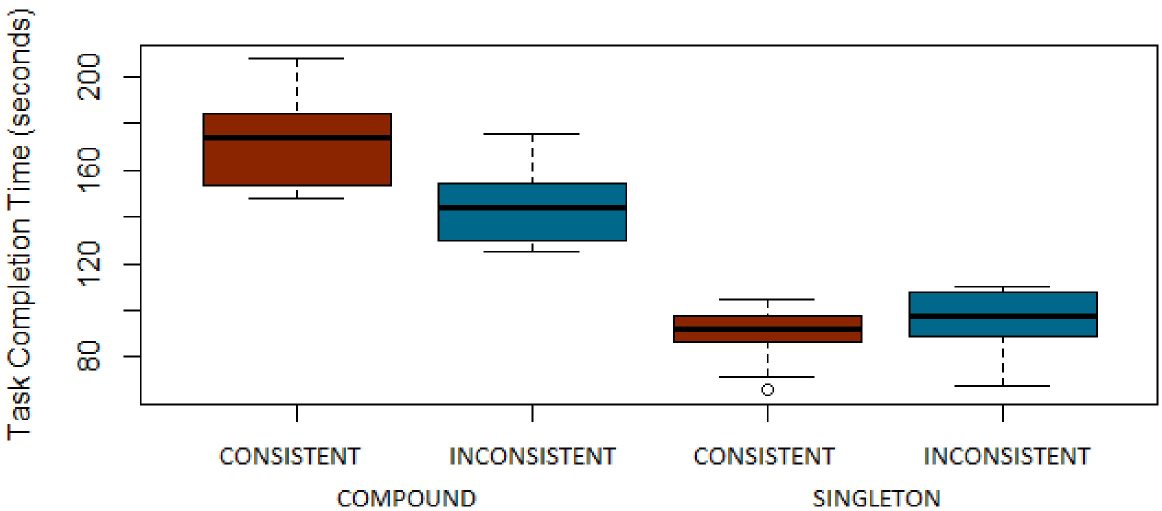

Fig. 7. Task completion time distribution in a Non-Search task.

4.4 Model Initialization

We used the results of the previous two experiments to derive weighting factors w_j for each dimension. In a Search task, a two predictor model explained 84.8% of the variance, with object singletoness explaining 80.7% and schema consistency 4.1%. Thus, $w_{physicalSEARCH} = 0.95$ (80.7% out of 84.8%) and $w_{semanticSEARCH} = 0.05$. In a Non-Search task, object singletoness explained 77.7% of the total 80.2%, giving us $w_{physicalSEARCH} = 0.97$ and $w_{semanticSEARCH} = 0.03$.

In order to calculate the likelihood values associated to the scene schema hypothesis we compare the associated scene schema of each examined object determined in Experiment 1 against the scene schemata associated with the objects that surround it. We define an object neighborhood of radius N as a multiple of the examined object's radius. We define c the count of objects residing in this neighborhood and m the count of items tagged with the same schema inside the neighborhood.
We then define $\lambda_{semantic}$ as:

$$\lambda_{semantic} = \frac{c-m}{c} \tag{7}$$

Inconsistent objects signified by their varied schema relatively to their surroundings have greater $\lambda_{semantic}$ values than consistent objects.

In order to calculate the likelihood values associated to the singleton hypothesis, we both examine the number of neighbours for each examined object and employ the available image depth information. In particular, we can use the spatial derivatives to estimate the magnitude of the depth gradient. This operator indicates how distinct an object is from its environment and is a strong indication of whether it is a singleton.
We thus define $\lambda_{physical}$ as:

$$\lambda_{physical} = \frac{1}{1-c} \times \sqrt{(\frac{\partial f}{\partial x})^2 + (\frac{\partial f}{\partial y})^2} \tag{8}$$

5. IMPLEMENTATION AND GAME BALANCING

In this section we describe a GPU implementation of our model and its integration in a game engine to assess game level difficulty. The efficiency of our model in predicting attention deployment is evaluated in Experiments 4 & 5.

5.1 GPU based implementation

We developed a plug-in for Unity 3D™ game engine which we call High Level Saliency Modeler (HLSM). HLSM highlights objects expected to attract attention by estimating in real time the posterior probability term (Equation 6) of our new high level attention model in a shader (Figure 8) (please see accompanying video). Equations (1)-(5) are supplied with semantic consistency and object singletoness information in terms of the $\lambda_{physical}$ and $\lambda_{semantic}$ variables as determined by the experiments. The $\lambda_{semantic}$ (Equation 7) and $\lambda_{physical}$ (Equation 8) are calculated at runtime by both quering the scene graph and utilizing an edge detection kernel run over the depth buffer. The obtained likelihood ratio sums are then combined according to the w_j factors obtained from the regression analysis applied to the experimental task completion timings (see Section 4.4). The d', σ and K values are user controlled via the system's user interface. Manipulating these parameters either increases or decreases the system's sensitivity to saliency resulting in more or fewer objects to be highlighted as salient respectively (Figure 9).

The shader approach offers view-dependent estimations i.e. an object may or may not appear as singleton depending on the viewpoint. Additionally, the linearity of the likelihoods calculated allows for linear quantitative measurements. For instance, "an object x is more inconsistent than object z by a factor of q". This offers rich information about the semantic context of objects as opposed to the previously defined binary definition of an object being characterized as either consistent or inconsistent [Zotos et al. 2009].

5.2 Game Level Editing

Game balancing is a meaningful application of high level saliency modeling. Plot-critical objects are placed in their respective locations by game designers to achieve a purpose: Ease or make it difficult for the player when searching for them depending on the plot. Placing objects far from expected locations is standard in game balancing [Feil and Scattergood 2005]. Integration of a high level saliency model in a game level editor can assist the level artists by highlighting salient objects. Designers using the proposed editor are able to reposition or tint props to make them less/more visible in real-time. This way, designers modulate the search-cost/benefit curve for easier or harder object recovery in Adventure or Action-Adventure games. When working with our tool, the game level designer proceeds as normal to place game objects as desired. The designer observes saliency visualization and examines the attention prediction for the current view. The current view or object placement may then be modified and high level saliency can be re-assessed in real-time. The overhead of investigating the attention predictions is minimal since the game level designer may save on time by not needing to elaborate on suitable locations for prop placement depending on the current game difficulty level that is developed. We used the GPU implementation of our model to guide object placement in a game level editor in order to adjust game difficulty. Our plug-in works in parallel with the editor, allowing the game designer to play-test the level while designing it.

5.3 Experiments 4 & 5: Evaluation of the implementation

We designed an experiment to evaluate the efficiency of our model in predicting attention deployment by examining its effect on task completion time and by acquiring eye-tracking data. Since our tool is intended to be used by game level designers when creating game levels, the evaluation also indicates the model's potential as a means to adjust game level difficulty.

5.3.1 *Design.* We created four game levels corresponding to two experimental conditions (*Easy*/*Hard*) of a Search and a Non-Search Task. The placement of three critical objects was manipulated to systematically alter game difficulty. Our model implementation (HLSM) assisted object placement by highlighting objects that were expected to pop out in a Search task for the first two conditions (*Experiment 4*) and in a Non-Search task for the last two conditions (*Experiment 5*). Figure 8 shows a vase at a consistent/singleton layout expected to attract attention in a Search Task and thus marked as red by HLSM. When the vase is placed on the chair therefore being at an inconsistent/compound location it is not expected to pop-up in a Search Task. When objects pop out we expect a shorter task completion time, thus the easy level for the Search task was created by placing consistent objects at a singleton state in the scene. A hard game level expected to be completed slower was created by placing inconsistent objects at a compound state (Table I). In relation to the Non-Search task the easy level was created by placing consistent objects at a singleton state and the hard level by placing consistent objects at a compound state expected to have the fastest/slowest recovery times respectively (Table II). In all cases we use our saliency modeler, which indicates the appropriate configurations. We used the Saliency Toolbox [Walther and Koch 2006] to ensure that the requested objects exhibited a minimum low level saliency (Figure 10). Constant navigation time to the individual objects was maintained regardless of location. Similar visual angles within the VE were maintained for all objects.

5.3.2 *Participants and Apparatus.*
Forty participants (34 male, 6 female; mean age 23) were split in four groups; 10 played the easy Search task level, 10 played the hard Search task level, 10 played the easy Non-Search task level and the rest played the hard Non-Search task level. For the Search task participants were instructed to find three specific objects. For the Non-Search task participants observed the VE to identify three unknown ob-

jects that were indirectly described: ”identify objects necessary for a car trip” (Figure 11). For both tasks participants were instructed to find the objects as quickly and as accurately as they could. Each subject participated in only one of the experimental conditions. The VEs were presented in stereo at SXGA resolution on an NVIS nVisor SX111 Head Mounted Display with a Field-of-View of 102 degrees horizontal. An InterSense InertiaCube3, three degrees of freedom head tracker was utilized for rotation and a gamepad for translation. Attached to the HMD was an eye-tracker by Arrington Research recontructing the subjects eye position through the Pupil-Center and Corneal-Reflection method at a rate of 30Hz. The eye tracking was performed to the dominant eye of each subject.

5.3.3 *Completion Time Analysis.*

Experiment 4: Search Task An independent-samples t-test was conducted, revealing a significant difference between easy (M=42.83, SD=11.83) and hard (M=82.2, SD=21.88) level completion times, $t(9) = -4.54$, $p < 0.0001$. The easy task completion time was reduced to 52.1% of the hard task; 42.83 vs 82.2 seconds, that is consistent with the results of the regression analyses of Experiment 2: A consistent/singleton object placement is predicted to be reduced to 37% of an inconsistent/compound object placement completion time derived from 59.272 (intercept+singleton+consistency terms) vs 158.962 seconds (Table I).

Experiment 5: Non-Search Task An independent-samples t-test was conducted, revealing a significant difference between easy (M=61.86, SD=17.57) and hard (M=138.35, SD=16.1) level completion times, $t(9) = -14.48$, $p < 0.0001$. The easy task completion time was reduced to 44.7% of the hard task; 61.86 vs 138.35 seconds, that is consistent with the results of the regression analyses of Experiment 3: A consistent/singleton object placement is predicted to be reduced to 39.67% of a consistent/compound object placement completion time derived from 65.434 (intercept+singleton+consistent+joint terms) vs 164.952 seconds (Table II).

The reduction of task completion time in the easy conditions when compared to the hard conditions for both the Search and Non-Search tasks validate our hypothesis that game level completion time depends on object topology as predicted by our system.

5.3.4 *Eye-tracking Data Analysis.*

For every object in quest a Region-Of-Interest (ROI) was defined. Each ROI held metadata indicating a consistent/inconsistent placement and a singleton/compound placement of the object in relation to its surroundings. In total 9837 fixations to the ROIs were recorded. As a fixation we considered every spatially stable gaze lasting for at least 300 milliseconds [Salvucci and Goldberg 2000].

For **Experiment 4** an independent-samples t-test was conducted on total object fixations per condition, revealing a significant difference between consistent/singleton (M=265.3, SD=15.41) and inconsistent/compound (M=182.6, SD=25.16) object placement, $t(9) = 7.45$, $p < 0.0001$.
For **Experiment 5** an independent-samples t-test was conducted on total object fixations per condition, revealing a significant difference between consistent/singleton (M=364.5, SD=44.92) and consistent/ compound (M=171.3, SD=19.04) object placement, $t(9) = 15.6$, $p < 0.0001$.

The results indicate a clear influence of context consistency in attention deployment for the Search Task. Singleton objects attracted attention in both conditions since the total number of fixations for ROIs defined for objects in a singleton state was higher for both the Search and Non-Search tasks. We aggregated fixations collected over raw eye data from all participants and visual angles in multiple heatmaps (Figure 11). Observing the heatmaps indicated that in a Search task eye gaze is directed significantly more often to consistent locations in relation to the requested object (Figure 11). In a

Non-Search task the eye scan pattern spans over the entire scene, which is consistent with previous literature stating that in an Action-Adventure game players mostly explore the entire screen for game props to advance the gameplay [El-Nasr and Yan 2006] (Figure 11).

Our model implementation succesfully predicts the saliency of objects (Figure 8) that were identified as non-salient in terms of low level features (Figure 10) further validated by the eye-tracking study (Figure 11). Adjusting game level difficulty by manipulating object topology is thus feasible in Adventure or Action-Adventure games.

Fig. 8. In a Search task, our tool highlights the vase at a consistent/singleton location signifying an easier recovery than at an inconsistent/compound location (on chair). The green hue indicates non-salient areas.

Fig. 9. The system's sensitivity to saliency can be adjusted, resulting in more (left) or fewer (right) objects to be highlighted as salient.

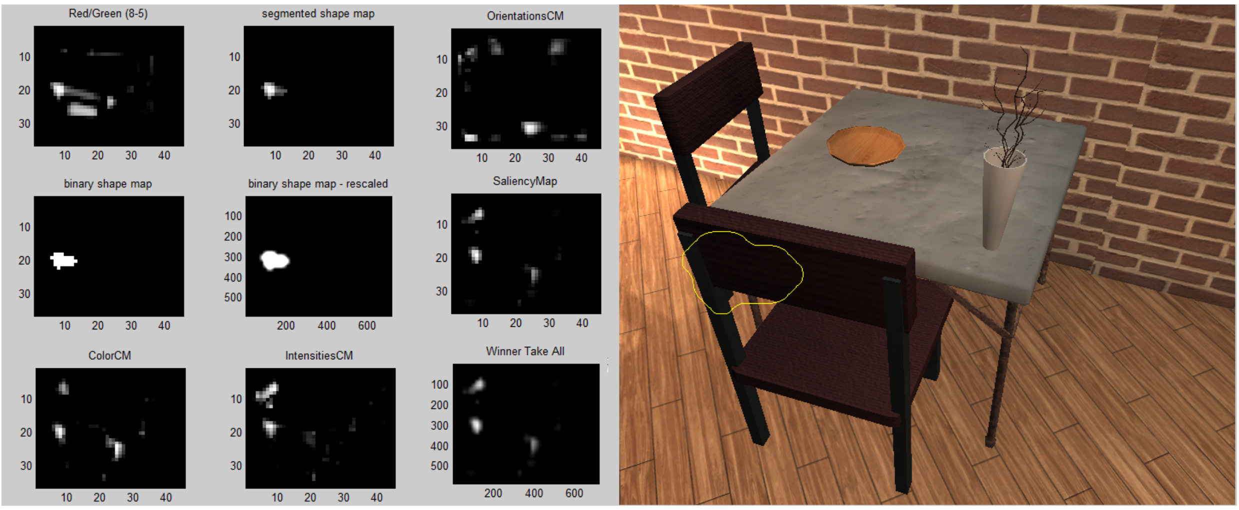

Fig. 10. The Saliency Toolbox [Walther and Koch 2006] indicates that the most salient area of the image is the dark area behind the chair.

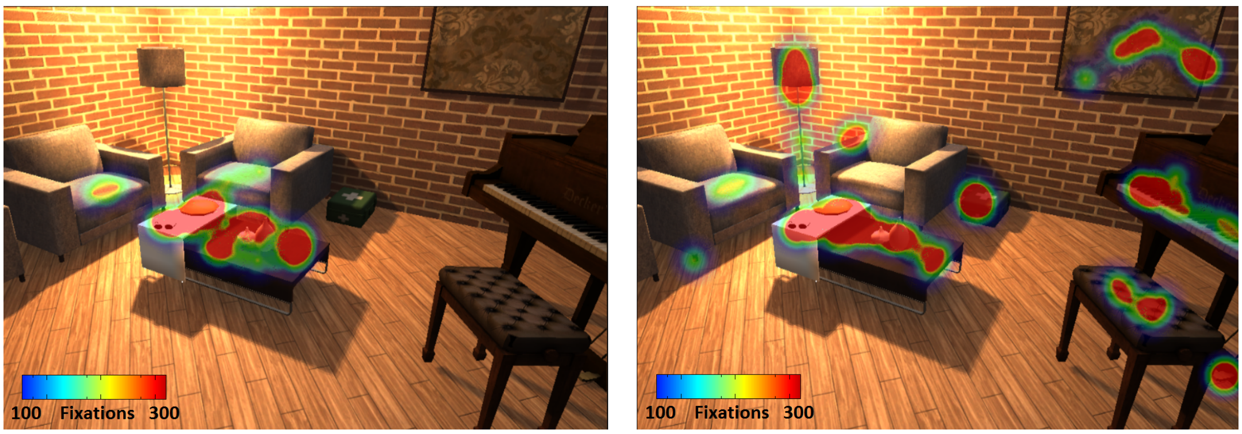

Fig. 11. The left image indicates fixations for a Search task where subjects were requested to find a pair of spectacles. The right image indicates fixations for a Non-Search task where subjects were requested to identify objects necessary for a car trip. Areas receiving less than 100 fixations are excluded to eliminate noise.

6. CONCLUSION & FUTURE WORK

This work presents a first attempt to devise a high level saliency predictor based on the topological relationships of objects with their surroundings and object-scene schema conformance for common tasks in (Action-)Adventure games. The framework automatically estimates attention deployment by identifying salient regions in the viewpoint. We conducted three experiments to verify that high level saliency of objects affects the time needed to find them in a VE and also obtained all the necessary weighting factors for our model. Finally, we developed a GPU based computational model that implements our new model incorporating high level saliency components. The system estimates the probabilities of individual objects to be foveated in real time and can be used in an innovative game level editor automatically suggesting game objects' positioning in order to adjust the difficulty of the game. The system can be adapted to additional tasks, different than the ones presented here by acquiring the necessary parameters using the methodology we presented.

We plan to include additional High Level Saliency Components in our model such as Perception of Sets. i.e. attention regarding large aggregations of similar objects as a single unit [Ariely 2001]; Violations of Canonical Form that are detected peripherally, are semantically salient and can affect the likelihood of fixating on an item [Becker et al. 2007] and novelty detection i.e. popping out objects becoming less salient over time [Markou and Singh 2003]. The production-level working system will provide a taxonomy of objects in relation to scene schemata as a library limiting the need for further experiments or manual input. We will evaluate the tool by presenting it to skilled experts in game level design as well as integrate and validate our model against documented low-level predictors. We plan to combine our saliency model with a low level saliency model to obtain even more accurate visual attention predictions. Simulating natural effects such as Depth-Of-Field, Camera motion and Dynamic Lighting could benefit from a list of potentially attended objects based on high level saliency. It has been shown that when these effects are dynamically adapted depending on gaze, users reproduce distances better in a VE [Moehring et al. 2009].

We are currently working on a Level-of-Detail framework for mobile devices that takes into account the dependence of attention deployment on scene context and object topology. The framework accelerates rendering by automatically removing perceptually non-important details in regions that are not expected to be attended. Taking into account the dependence of attention deployment on scene context and object topology the innovative renderer saves computational time by automatically and seamlessly removing perceptually non-important details. We will show that integration of a high level saliency model in a Level-of-Detail manager enables the usage of complex effects in low-power devices by applying them sparingly only in regions that are expected to be attended [2]. Finally, we intend to integrate our model in a game engine for on-the-fly level difficulty adjustments and a smarter game AI. Objects could be repositioned dynamically resulting in an adjustable level of difficulty depending on user performance so far. Object placement could automatically shift after every respawn when a player comes back to life after being killed. A smarter AI could use high level saliency data to spawn opponents that pop-out or appear inconspicuously.

ACKNOWLEDGMENTS

This research has been co-financed by the European Union (European Social Fund - ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program: Heracleitus II: Investing in knowledge society through the European Social Fund. We thank Adobe and Autodesk for generous donations; the work was partly supported by EU FP7 project ICT-611089-CR-PLAY www.cr-play.eu.

European Union
European Social Fund
OPERATIONAL PROGRAMME
EDUCATION AND LIFELONG LEARNING
investing in knowledge society
MINISTRY OF EDUCATION & RELIGIOUS AFFAIRS, CULTURE & SPORTS
MANAGING AUTHORITY
Co- financed by Greece and the European Union
NSRF 2007-2013
programme for development
EUROPEAN SOCIAL FUND

[2]During the review cycle for this paper we advanced on this work, which is now published as [Koulieris et al. 2014]

REFERENCES

ARIELY, D. 2001. Seeing sets: Representation by statistical properties. *Psychological Science 12,* 2, 157–162.

AWH, E. AND PASHLER, H. 2000. Evidence for split attentional foci. *Journal of Experimental Psychology: Human Perception and Performance 26,* 2, 834.

BAR, M., ULLMAN, S., ET AL. 1996. Spatial context in recognition. *PERCEPTION-LONDON- 25*, 343–352.

BARTLETT, F. C. 1932. Remembering: An experimental and social study. *Cambridge: Cambridge University*.

BECKER, M. W., PASHLER, H., AND LUBIN, J. 2007. Object-intrinsic oddities draw early saccades. *Journal of Experimental Psychology: Human Perception and Performance 33,* 1, 20.

BERNHARD, M., STAVRAKIS, E., AND WIMMER, M. 2010. An empirical pipeline to derive gaze prediction heuristics for 3d action games. *ACM Transactions on Applied Perception (TAP) 8,* 1, 4.

BERNHARD, M., ZHANG, L., AND WIMMER, M. 2011. Manipulating attention in computer games. In *IVMSP Workshop, 2011 IEEE 10th*. IEEE, 153–158.

BORJI, A. AND ITTI, L. 2013. State-of-the-art in visual attention modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence 35,* 1.

BREWER, W. F. AND TREYENS, J. C. 1981. Role of schemata in memory for places. *Cognitive Psychology 13,* 2, 207–230.

CATER, K., CHALMERS, A., AND WARD, G. 2003. Detail to attention: exploiting visual tasks for selective rendering. In *Proceedings of the 14th Eurographics workshop on Rendering*. Eurographics Association, 270–280.

CUNNINGHAM, D. W. AND WALLRAVEN, C. 2011. *Experimental Design: From user studies to psychophysics*. A.K. Peters.

CZIKSZENTMIHALYI, M. 1990. Flow: The psychology of optimal experience. *Praha: Lidové Noviny*.

DESURVIRE, H., CAPLAN, M., AND TOTH, J. A. 2004. Using heuristics to evaluate the playability of games. In *CHI'04 extended abstracts on Human factors in computing systems*. ACM, 1509–1512.

ECKSTEIN, M. P. 1998. The lower visual search efficiency for conjunctions is due to noise and not serial attentional processing. *Psychological Science 9,* 2, 111–118.

ECKSTEIN, M. P., DRESCHER, B. A., AND SHIMOZAKI, S. S. 2006. Attentional cues in real scenes, saccadic targeting, and bayesian priors. *Psychological Science 17,* 11, 973–980.

ECKSTEIN, M. P., SHIMOZAKI, S. S., AND ABBEY, C. K. 2002. The footprints of visual attention in the posner cueing paradigm revealed by classification images. *Journal of Vision 2,* 1.

EINHÄUSER, W., SPAIN, M., AND PERONA, P. 2008. Objects predict fixations better than early saliency. *Journal of Vision 8,* 14.

EL-NASR, M. S. AND YAN, S. 2006. Visual attention in 3d video games. In *Proceedings of the 2006 ACM SIGCHI international conference on Advances in computer entertainment technology*. ACM, 22.

FEIL, J. AND SCATTERGOOD, M. 2005. *Beginning game level design*. Course Technology PTR.

FRINTROP, S., ROME, E., AND CHRISTENSEN, H. I. 2010. Computational visual attention systems and their cognitive foundations: A survey. *ACM Transactions on Applied Perception (TAP) 7,* 1, 6.

GREEN, D. M., SWETS, J. A., ET AL. 1966. *Signal detection theory and psychophysics*. Vol. 1. Wiley New York.

GRILLON, H. AND THALMANN, D. 2009. Simulating gaze attention behaviors for crowds. *Computer Animation and Virtual Worlds 20,* 2-3, 111–119.

HENDERSON, J. M. AND HOLLINGWORTH, A. 1999. High-level scene perception. *Annual review of psychology 50,* 1, 243–271.

HENDERSON, J. M., WEEKS JR, P. A., AND HOLLINGWORTH, A. 1999. The effects of semantic consistency on eye movements during complex scene viewing. *Journal of experimental psychology: Human perception and performance 25,* 1, 210.

HILLAIRE, S., LÉCUYER, A., REGIA-CORTE, T., COZOT, R., ROYAN, J., AND BRETON, G. 2010. A real-time visual attention model for predicting gaze point during first-person exploration of virtual environments. In *Proceedings of the 17th ACM Symposium on Virtual Reality Software and Technology*. ACM, 191–198.

HWANG, A. D., WANG, H.-C., AND POMPLUN, M. 2011. Semantic guidance of eye movements in real-world scenes. *Vision research 51,* 10, 1192–1205.

ITTI, L. AND KOCH, C. 2001. Computational modelling of visual attention. *Nature reviews neuroscience 2,* 3, 194–203.

JOHNSON, D. AND WILES, J. 2003. Effective affective user interface design in games. *Ergonomics 46,* 13-14, 1332–1345.

JU, E. AND WAGNER, C. 1997. Personal computer adventure games: their structure, principles, and applicability for training. *ACM SIGMIS Database 28,* 2, 78–92.

JUDD, T., EHINGER, K., DURAND, F., AND TORRALBA, A. 2009. Learning to predict where humans look. In *Computer Vision, 2009 IEEE 12th international conference on*. IEEE, 2106–2113.

KOCH, C. AND ULLMAN, S. 1987. Shifts in selective visual attention: towards the underlying neural circuitry. In *Matters of Intelligence*. Springer, 115–141.

KOULIERIS, G. A., DRETTAKIS, G., CUNNINGHAM, D., AND MANIA, K. 2014. C-lod: Context-aware material level-of-detail applied to mobile graphics. *Computer Graphics Forum (Proceedings of the Eurographics Symposium on Rendering 2014, Lyon, France) 33,* 4.

LEE, S., KIM, G. J., AND CHOI, S. 2009. Real-time tracking of visually attended objects in virtual environments and its application to lod. *Visualization and Computer Graphics, IEEE Transactions on 15,* 1, 6–19.

LONGHURST, P., DEBATTISTA, K., AND CHALMERS, A. 2006. A gpu based saliency map for high-fidelity selective rendering. In *Proceedings of the 4th international conference on Computer graphics, virtual reality, visualisation and interaction in Africa*. ACM, 21–29.

MANIA, K., ROBINSON, A., AND BRANDT, K. R. 2005. The effect of memory schemas on object recognition in virtual environments. *Presence: Teleoperators and Virtual Environments 14,* 5, 606–615.

MARKOU, M. AND SINGH, S. 2003. Novelty detection: a review part 2, neural network based approaches. *Signal processing 83,* 12, 2499–2521.

MARR, D. 1982. Vision: A computational investigation into the human representation and processing of visual information, henry holt and co. *Inc., New York, NY*.

MOEHRING, M., GLOYSTEIN, A., AND DOERNER, R. 2009. Issues with virtual space perception within reaching distance: Mitigating adverse effects on applications using hmds in the automotive industry. In *Virtual Reality Conference, 2009. VR 2009. IEEE*. IEEE, 223–226.

MOURKOUSSIS, N., RIVERA, F. M., TROSCIANKO, T., DIXON, T., HAWKES, R., AND MANIA, K. 2010. Quantifying fidelity for virtual environment simulations employing memory schema assumptions. *ACM Transactions on Applied Perception (TAP) 8,* 1, 2.

NAKAYAMA, K., SILVERMAN, G. H., ET AL. 1986. Serial and parallel processing of visual feature conjunctions. *Nature 320,* 6059, 264–265.

O'CRAVEN, K. M., DOWNING, P. E., AND KANWISHER, N. 1999. fmri evidence for objects as the units of attentional selection. *Nature 401,* 6753, 584–587.

OYEKOYA, O., STEPTOE, W., AND STEED, A. 2009. A saliency-based method of simulating visual attention in virtual scenes. In *Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology*. ACM, 199–206.

PAGULAYAN, R. J., KEEKER, K., WIXON, D., ROMERO, R. L., AND FULLER, T. 2003. User-centered design in games. *The human-computer interaction handbook: fundamentals, evolving technologies and emerging applications*, 883–906.

PALMER, S. E. 1999. *Vision science: Photons to phenomenology.* The MIT press.

RAYNER, K. 2009. Eye movements and attention in reading, scene perception, and visual search. *The quarterly journal of experimental psychology 62,* 8, 1457–1506.

RENSINK, R. A. 2000. The dynamic representation of scenes. *Visual cognition 7,* 1-3, 17–42.

SALVUCCI, D. D. AND GOLDBERG, J. H. 2000. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 symposium on Eye tracking research & applications*. ACM, 71–78.

SHIPLEY, T. F. AND KELLMAN, P. J. 2001. *From fragments to objects: Segmentation and grouping in vision*. Vol. 130. Access Online via Elsevier.

STEINMETZ, P. N., ROY, A., FITZGERALD, P., HSIAO, S., JOHNSON, K., AND NIEBUR, E. 2000. Attention modulates synchronized neuronal firing in primate somatosensory cortex. *Nature 404,* 6774, 187–190.

SUNDSTEDT, V., BERNHARD, M., STAVRAKIS, E., REINHARD, E., AND WIMMER, M. 2013. Visual attention and gaze behavior in games: An object-based approach. In *Game Analytics*. Springer, 543–583.

SUNDSTEDT, V., CHALMERS, A., CATER, K., AND DEBATTISTA, K. 2004. Top-down visual attention for efficient rendering of task related scenes. In *VMV*. 209–216.

SUNDSTEDT, V., DEBATTISTA, K., LONGHURST, P., CHALMERS, A., AND TROSCIANKO, T. 2005. Visual attention for efficient high-fidelity graphics. In *Proceedings of the 21st spring conference on Computer graphics*. ACM, 169–175.

SUNDSTEDT, V., STAVRAKIS, E., WIMMER, M., AND REINHARD, E. 2008. A psychophysical study of fixation behavior in a computer game. In *Proceedings of the 5th symposium on Applied perception in graphics and visualization*. ACM, 43–50.

SWEETSER, P. AND WYETH, P. 2005. Gameflow: a model for evaluating player enjoyment in games. *Computers in Entertainment (CIE) 3,* 3, 3–3.

THEEUWES, J. 2010. Top–down and bottom–up control of visual selection. *Acta psychologica 135,* 2, 77–99.

THEEUWES, J. AND GODIJN, R. 2002. Irrelevant singletons capture attention: Evidence from inhibition of return. *Perception & Psychophysics 64,* 5, 764–770.

TOLHURST, D. J., MOVSHON, J. A., AND DEAN, A. 1983. The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision research 23,* 8, 775–785.

TREISMAN, A. M. AND GELADE, G. 1980. A feature-integration theory of attention. *Cognitive psychology 12,* 1, 97–136.

WALTHER, D. AND KOCH, C. 2006. Modeling attention to salient proto-objects. *Neural Networks 19,* 9, 1395–1407.

WOLFE, J. M. 1994. Guided search 2.0 a revised model of visual search. *Psychonomic bulletin & review 1,* 2, 202–238.

YARBUS, A. L., HAIGH, B., AND RIGSS, L. A. 1967. *Eye movements and vision*. Vol. 2. Plenum press New York.

ZOTOS, A., MANIA, K., AND MOURKOUSSIS, N. 2009. A schema-based selective rendering framework. In *Proceedings of the 6th Symposium on Applied Perception in Graphics and Visualization*. ACM, 85–92.