Eurographics Symposium on Rendering 2014
Wojciech Jarosz and Pieter Peers
(Guest Editors)

Volume 33 (2014), Number 4

C-LOD: Context-aware Material Level-of-Detail applied to Mobile Graphics

G. A. Koulieris[1], G. Drettakis[2], D. Cunningham[3] K. Mania[1]

[1]Technical University of Crete, Greece [2]Inria† [3]Brandenburg Technical University Cottbus, Germany

Abstract

Attention-based Level-Of-Detail (LOD) managers downgrade the quality of areas that are expected to go unnoticed by an observer to economize on computational resources. The perceptibility of lowered visual fidelity is determined by the accuracy of the attention model that assigns quality levels. Most previous attention based LOD managers do not take into account saliency provoked by context, failing to provide consistently accurate attention predictions. In this work, we extend a recent high level saliency model with four additional components yielding more accurate predictions: an object-intrinsic factor accounting for canonical form of objects, an object-context factor for contextual isolation of objects, a feature uniqueness term that accounts for the number of salient features in an image, and a temporal context that generates recurring fixations for objects inconsistent with the context. We conduct a perceptual experiment to acquire the weighting factors to initialize our model. We design C-LOD, a LOD manager that maintains a constant frame rate on mobile devices by dynamically re-adjusting material quality on secondary visual features of non-attended objects. In a proof of concept study we establish that by incorporating C-LOD, complex effects such as parallax occlusion mapping usually omitted in mobile devices can now be employed, without overloading GPU capability and, at the same time, conserving battery power.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Picture/Image Generation—Graphics systems and interfaces

1. Introduction

LOD algorithms render with higher visual fidelity those regions of a synthetic image that are expected to receive attention, allowing more efficient distribution of the limited resources of a graphics subsystem. The interest in efficient LOD management has been recently renewed due to the explosive growth of the mobile market, which is extremely diverse in terms of computing power. Hardware restrictions of mobile devices prohibit the use of complex effects that demand multiple texture fetches or intense Arithmetic Logic Unit (ALU) operations [ÇPAM08]. An application's expressive power is thus sacrificed in portable devices as content is displayed at degraded levels of detail or quality.

LOD managers have been empowered with perceptual principles in the past to optimize the distribution of computational time and maximize the perceived quality of a rendered scene [Lue03]. Computation time can be minimized and the quality of an effect downgraded, based on evidence determining that a user is not attending a scene area. A focused distribution of available resources only to *attended* areas allows for higher and more stable frame rates.

Recently, attention assumptions directed by scene context information were utilized to extend a physiologically plausible model of visual attention [Eck98] and use it for game balancing [KDCM14]. The High Level Saliency Model (HLSM) [KDCM14] incorporates two insights observed in object-context hierarchies. These insights are the *object singletoness hypothesis*, i.e., attention is attracted to physically isolated objects [TG02], and the *scene schema hypothesis* indicating that scenes are comprised of consistent objects expected in a specific context and inconsistent, thus out-of-context objects that are salient [BT81] (Figure 1).

There are, however, other high level saliency phenomena that affect attention which are not captured by this model. We therefore extend the model to include four new factors that allow us to process additional attentional phenomena and predict attention deployment with higher accuracy. In

† Inria authors are at REVES/Inria Sophia-Antipolis Méditerranée.

particular we introduce an object-intrinsic factor accounting for the fact that an object pops out if it is rotated in a way that violates its expected posture. The expected posture is known as *canonical form* or canonical orientation [BPL07] (Figure 1). We also add an *object-context factor* for contextual isolation of objects, a *feature uniqueness term* that accounts for the number of salient features in an image, and a *temporal context* that generates recurring fixations for objects inconsistent with the context or in a non-canonical form.

We then incorporate this new model of attention into a perceptually optimized renderer for mobile platforms that takes into account the dependence of attention deployment on scene context and object placement. This saves computational time by automatically and seamlessly removing perceptually non-important details. Integration of a contextual attention model in a LOD manager enables the usage of – otherwise omitted – complex effects in low-power devices by applying them sparingly only in regions that are expected to be attended and improves battery life by reducing GPU utilization.

We make four primary contributions:

- We extend the High Level Saliency Model [KDCM14] by introducing the four additional factors described above. These additions yield more accurate predictions of attention than previous work.
- We acquire the parameters to initialize our model's canonical form and perception of contextual singletoness in a perceptual experiment. The experimental design controls for attentional effects from low level features such as luminance or contrast, allowing us to examine the unique contribution of context.
- We develop a novel LOD manager that speeds up rendering in mobile devices based on attention predictions as derived by our model. A proof-of concept implementation selects an appropriate LOD in real-time for subsurface scattering, complex refraction and bump mapping algorithms.
- We demonstrate the accuracy of our implementation by comparing its performance to actual eye-tracking data. We also acquire mobile GPU performance statistics to ensure model effectiveness and quantify battery performance gain when limiting GPU utilization.

2. Related Work

LOD Modern video games consist of various interconnected software components such as a graphics engine and an audio engine that share hardware resources. LOD methods are essential to improve the interactivity and responsiveness of graphics systems by distributing resources to the image regions that are expected to be attended [Lue03]. Traditional LOD approaches reduce polygon count by selecting an appropriate instance of polygonal complexity for each model depending on its importance [Lue03]. Object importance can be determined by attention deployment over the scene or perceptually motivated criteria such as the projected screen size of the object, eccentricity and velocity of objects [Cla76].

Polygonal counts are usually low in mobile devices and mobile GPUs are fill-rate bound deeming polygonal complexity LOD algorithms ineffective [ÇPAM08]. Shaders reproduce high quality visual details by exchanging polygonal complexity for additional ALU operations and heavy texture memory accesses. As computation power in mobile GPUs increases faster than memory bandwidth [Owe05] our LOD manager significantly reduces texture fetches.

Attention based LOD Gaze [LM00] and task [CCW03] based LOD managers render the 2 degree fovea region in high quality i.e. the high-resolution part of the visual field and the periphery of vision with less detail. However, LOD management based on gaze encounters difficulties to maintain display updates without artifacts after fast eye saccades. Driving LOD based on pre-defined task areas is limited since it is impossible to quantify the nearly infinite number of potential tasks.

Since low level image features such as luminance, contrast and motion are known to attract attention [IKN*98], objects saliency models based on low-level features combined with task relevant information have been employed in order to drive LOD [LKC09, HLRC*10]. Since high-level, cognitive phenomena also affect attention, low-level saliency models sometimes fail to predict fixations, especially when an observer manipulates interactive scenes [SSWR08]. Here, we develop and employ a sophisticated, multi-factor, context-based, attention predictor for interactive environments that takes into account contextual information about a scene to optimize LOD for mobile platforms.

Scene Semantics Object perception in natural scenes relies on the integration of pre-existing knowledge with recently acquired from attentional processing [HWJH99, Ren00]. In a schema-based LOD framework consistent objects are rendered with lower quality without affecting information uptake [ZMM09].

The High Level Saliency Model (HLSM) of [KDCM14] is based on the Differential-Weighting Model (DWM) that simulates attentional processing using Gaussian combination rules [Eck98]. The HLSM takes into account the fact that high level features of an image such as physical object isolation and object-context consistency (Figure 1) affect attention. The model describes two *High Level Saliency* sensory units that react to the existence of high level features in the Field-Of-View (FOV). The areas that are most consistent to the feature to which a unit is sensitive capture that unit's attention. For example, an object placed in an unexpected location will elicit a very strong response in a unit that observes object-context consistency. The model predicts saccadic targets by combining the estimated firing rate for both its units in a winner-take-all network.

Each unit's firing behaviour is encoded in Bayesian Priors and is affected by physiological noise having a Gaus-

Figure 1: *From left to right: A remote control is inconsistent with the sink context. The flowerpot is physically isolated. The tablet is contextually isolated. The chair is in a non-canonical form.*

sian distribution with a mean d'_j and a standard deviation σ [TMD83]. For each sensory unit j the model calculates for each pixel (x,y) of an image f its probability to be attended as the likelihood ratio LR [GS*66] of two noisy responses: the likelihood $l_{j,x,y,f}|s$ of observing a feature when it is present and the likelihood $l_{j,x,y,f}|n$ of erroneously observing a feature when it is not present.

$$LR_{j,x,y,f} = \frac{l_{j,x,y,f}(\lambda_{j,x,y,f}|s)}{l_{j,x,y,f}(\lambda_{j,x,y,f}|n)} = exp\left(\frac{\lambda_{j,x,y,f}d'_j - 0.5d'^2_j}{\sigma^2}\right) \quad (1)$$

The posterior probability of each image pixel to be attended is estimated as the weighted average of two units defined, namely $LR_{physical}$ and $LR_{semantic}$ encoding physical object isolation and semantic consistency. The averaging weights were derived from the results of a perceptual study.

Since the plot of many game genres is based on recognizing and acquiring objects in cluttered environments, [KDCM14] hypothesized and successfully verified that the manipulation of object placement to alter object-context consistency as well as the relocation of physically isolated objects can implicitly adjust game level difficulty. The performance of this model was not evaluated via eye tracking.

3. Overview

Gaze allocation is influenced by several other *high level factors* in cluttered environments. Not taking these factors into account deprives the model of important contextual information that would otherwise predict attention with higher accuracy. We set three criteria to be satisfied when introducing a high level component in our saliency model. The components (i) should affect attention as documented in cognitive psychology literature, (ii) should be measurable and their parameters quantifiable (iii) should be observed in a video game. We introduce four additional components. First, we subdivide the physically compound state defined in [KDCM14] by introducing two sub-states based on findings from psychological research (e.g., [Kof35]). Specifically, we hypothesize that a physically compound object can either be *contextually compound* or *contextually isolated*. Objects belonging in a set are contextually compound. An object positioned in-between a set of similar objects but dissimilar from those in the set, is hypothesized to pop out even when not salient in terms of e.g. color [Kof35]. For example, a tablet computer placed in-between magazines is salient (Figure 1). Second, we integrate an object-intrinsic assumption. The three-quarters object view, that makes a large number of surfaces visible is considered to be an object's canonical form [BTBV99,SLF*11]. The amount of angular deviation from this standard posture affects the object's saliency [BPL07] (Figure 1). Objects whose orientation is non-canonical are common in games e.g. dead characters or overturned vehicles. Third, we account for object coherence in time. An attended location is usually prevented from being attended again [PC84], an observation that has been used for LOD management [LDC06]. However, there is strong evidence that recurring fixations are generated for objects that are inconsistent with the context or for objects that are in a non-canonical form [BPL07, HWJH99]. Finally, we complement our model by accounting for the biologically motivated feature uniqueness property. A single salient feature in an image pops-out more intensely than when several salient features exist [IKN*98, FRC10].

Our LOD manager adjusts LOD only during player motion. Pop-out artifacts [Lue03] are eliminated by exploiting the observer insensitivity to perceive changes occuring during a brief interruption known as the Change Blindness phenomenon [SL97]. We evaluate our model via eye-tracking [Duc07].

4. Attention Model

We extend the HLSM [KDCM14] with four additional components: (i) an object-context singletoness factor that traces contextual object isolation, (ii) an object-intrinsic cognitive factor, termed canonical form of objects [BPL07], (iii) a biologically motivated feature uniqueness factor [FRC10] and (iv) a factor for temporal object coherence. The first two new factors (contextual isolation and canonical form) are incorporated through two new high level sensory units. To account for feature uniqueness, equations determine the number of local maxima found in the probability output of a sensory unit. That is, the more maxima there are, the less unique a feature is. For example, if there is only a single

violation of canonical form, its uniqueness weight is high. If several violations exist, all violations are less unique. Recurring fixations to areas containing canonical form violations or schema inconsistencies are generated by multiplying a unit's current output with a number of logarithmically attenuated previous outputs.

The output of all units is multiplied with a feature uniqueness weight:

$$w_{unit}^{unq} = \frac{1}{|\vee|P_{unit,x,y,f}} \quad (2)$$

x, y denotes image location, f denotes frame number, $|\vee|$ the number of posterior probability local maxima (Figure 2).

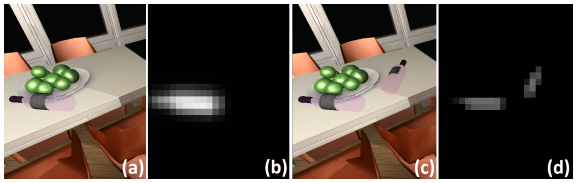


Figure 2: *A single violation of canonical form in the FOV (a) provokes a response in the canonical form sensory unit (b). When more violations of canonical form exist (c) the sensory unit's output is attenuated (d).*

The output of the schema consistency unit and the canonical form unit (Equation 1) are also multiplied with a temporal context weight:

$$w_{unit,x,y,f}^{tmp} = \prod_{f=1}^{F} P_{unit,x,y,f} e^{-af} \quad (3)$$

F the number of previous frames examined, a is a user-defined attenuation factor (Figure 3).

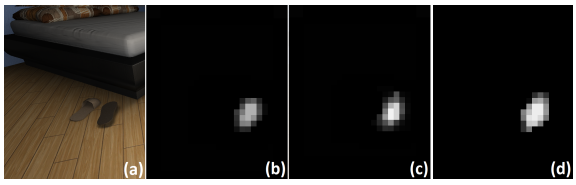


Figure 3: *The slipper on the right is in a non-canonical form (a). The output of the canonical form unit is shown in the current frame (b), in a subsequent frame (c) and in a third frame after the first (d). The increasing probability will generate recurring fixations for our model.*

The posterior probability $\mathbf{P}_{x,y,f}$ that an observer attends an image location, as part of our enhanced model, is linearly estimated [FRC10] from both the semantic consistency *(sem)* and physical isolation *(phy)* units of [KDCM14] combined with our contextual isolation *(cnt)* and canonical form *(cfr)* units, updated for feature uniqueness and temporal context:

$$\mathbf{P}_{x,y,f} = w_{sem} w_{sem}^{unq} w_{sem,x,y,f}^{tmp} \mathbf{P}_{\mathbf{sem},x,y,f} + w_{phy} w_{phy}^{unq} \mathbf{P}_{\mathbf{phy},x,y,f} + w_{cnt} w_{cnt}^{unq} \mathbf{P}_{\mathbf{cnt},x,y,f} + w_{cfr} w_{cfr}^{unq} w_{cfr,x,y,f}^{tmp} \mathbf{P}_{\mathbf{cfr},x,y,f} \quad (4)$$

In Section 5 the contribution weights w_{sem} and w_{phy} that were estimated in [KDCM14] are adapted to our model and the weights w_{cnt} and w_{cfr} are estimated based on our experimental data.

5. Perceptual Study

We conducted a perceptual experiment using a *Search* task to be comparable to [KDCM14]. We thus: (i) examine the effect of violations of canonical form and contextual singletoness on visual attention and (ii) obtain contribution weights of each factor for our model.

Stimuli We factorially combined the two factors to control the spatial arrangement of three objects (a tablet computer, a pair of spectacles and a remote control; see Figure 4) in four virtual environments. The four scenes were contextually compound/canonical, contextually compound/non-canonical, contextually singleton/canonical, or contextually singleton/non-canonical (Figure 5). All objects were consistent with the scenes and were physically compound. The Saliency Toolbox [WK06] (Figure 6) was used to ensure that the three objects had a minimum low-level saliency.

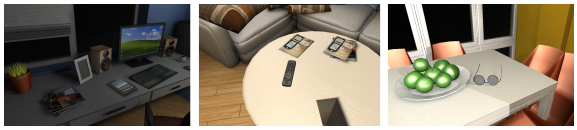


Figure 4: *The subjects searched for three objects, a tablet computer, a remote control and a pair of spectacles.*

Participants Fourty-eight people participated (8 female, mean age 23) in the experiment, with 12 people being assigned to each of the 4 conditions.

Apparatus The stimuli were displayed on a NVisor^{TM} SX111 Head Mounted Display (HMD), which has stereo SXGA resolution and a FOV of 102 degrees horizontal by 64 degrees vertical. Participants moved through the virtual environment using a game-pad for translation and an InterSense^{TM} InertiaCube3^{TM} 3DoF head tracker for rotation. Navigation was restricted to -70/70 degrees vertically. Eye tracking information was recorded using a twin-CCD binocular eye-tracker by Arrington Research^{TM}, which was attached to the HMD. The eye tracker was updated at a frequency of 30Hz.

Procedure Participants sat on a swivel chair and were familiarized with the setup in a training session. They were then requested to navigate around the scene in order to find and collect all three objects. Task accuracy, completion time, and eye-tracking data were recorded.

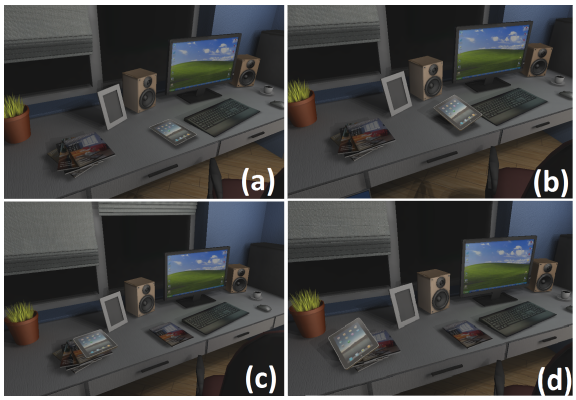


Figure 5: *The tablet is in a (a) contextually compound canonical form (tablet and keyboard), (b) contextually compound non-canonical form (slanted), (c) contextually singleton canonical form and (d) contextually singleton non-canonical form.*

Results Task accuracy was always 100%. On average, participants needed 167.788, 255.386, 82.189, and 195.985 seconds for the compound/canonical, compound/non-canonical, singleton/canonical, and singleton/non-canonical conditions, respectively. Task completion times were analyzed with a linear Hierarchical Multiple Regression analysis (HMR) with contextual singletoness being entered at stage one and canonical form at stage two. HMR fits a linear model to the data, with one term for each factor. The weight associated with each term is related to the correlation coefficient between the dependent variable (here, completion time) and the different factors. This effectively describes how well changes in the measured data can be explained or predicted by changes in the factors. Contextual singletoness contributed significantly to the regression model, $F(1,46) = 16.83, p < .001$ and accounted for 26.79% of the variation in task completion time. Introducing canonical form explained an additional 51.68%, $F(2,45) = 82.03, p < .001$, for a total explained variance of 78.47%. The coefficients for the two factors can be seen in Table 1. Predictions for a condition can be obtained by combining the intercept (i.e., performance in the Compound/Canonical condition) with the appropriate modifiers (i.e., the non-canonical form and/or singleton terms; see Table 1). The predictions of the model are consistent with the actual recorded completion times.

An analysis of the eye-tracking Regions-Of-Interest (ROIs) showed that attention is indeed attracted both to contextually singleton objects and to objects in a non-canonical form.

Discussion The canonical form and contextual isolation of objects play a significant role in attention deployment. In particular, in the non-canonical form conditions the objects were actively observed despite the fact that their recognition was extremely slow when compared to the canonical form condition. This is apparently in contradiction with [KDCM14] who found that actively attended salient objects

Coefficients	Estimate Time	p-value
Intercept	161.238	< 0.0001
+Non-Canonical Form term	100.697	< 0.0001
+Singleton Placement term	-72.500	< 0.0001

Table 1: *The regressions coefficients for each factor.*

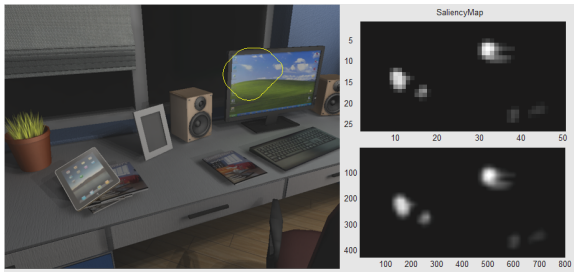


Figure 6: *The yellow contour delineates the most salient region of the image as predicted by the Saliency Toolbox [WK06]. Our hypothesis is that the tablet in a non-canonical form is the most salient object in this image.*

are easy to find. Thus, when managing LOD, an object in non-canonical form is salient and should always be rendered in high quality.

Weight Generation In their paper [KDCM14] derived the model weights from the correlation coefficients by dividing the amount of variance that a factor explained by the total explained variance. Since a single, between-participants experiment using a factorial combination of all levels of all four factors does not exist (it would require a prohibitively large number of participants), it is not possible to determine the *relative* amount of variance each factor explains. Fortunately, there is an alternative: the regression coefficients explicitly correlate changes in a factor with changes in completion time. Thus, it should be possible to get similar weights directly from the completion times. To make the completion times in two experiments comparable, we use the single condition that is the same in both experiments (physically compound/consistent in [KDCM14] and contextually compound/canonical here) to normalize the weights.

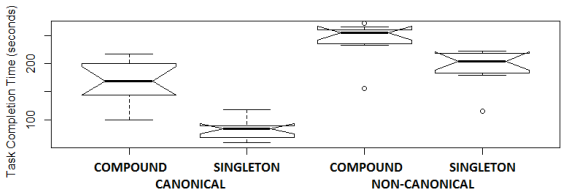


Figure 7: *Task completion time distribution of the experimental conditions. The median value for each condition is depicted by the horizontal line. The notched boxes depict the middle quartiles. The outer bars represent the extremes for each case.*

According to the regression models, the completion time should be 140.581 seconds in [KDCM14] and 161.238 here. Therefore, we divide the actual completion times in all of [KDCM14]'s conditions by 140.581, and all of our times by 161.238. For example, the actual mean of the contextually compound/canonical condition was 167.788. After normalization, it is 1.04. By comparing the relevant conditions, we can determine the relative effect of altering one factor. For example, we compare changes in contextual isolation for canonical objects ($1.04 - 0.51$) and for non-canonical objects ($1.58 - 1.22$). The average difference is 0.45. After we obtain all weights, we ensure that the weights all sum to 1. The final weights are 0.07 for schema, 0.33 for physical isolation, 0.35 for canonical form and 0.25 for contextual isolation (see supplemental document for more details).

6. LOD for Mobile Graphics

We developed a generic material LOD manager based on attention for Unity 3D™ game engine that we call C-LOD. C-LOD is a reactive fixed frame rate scheduler [Lue03] that constantly examines frame rate and attention deployment predictions using the criteria of our model. When frame rate drops below 30 frames per second on fill-rate bound mobile devices, C-LOD automatically lowers the rendering quality of objects predicted not to be attended until performance is restored (Figure 9). The highest quality possible is maintained for all attended objects.

6.1. C-LOD Effects

C-LOD can manage any effect that has at least two levels of detail. For this proof-of-concept implementation we selected three complex effects that are usually omitted in mobile devices as they require many texture fetches [ÇPAM08]. We used two LOD fall-backs for each effect, that require fewer texture fetches (Figure 8).

Subsurface light transport in translucent materials requires intense analytical calculations, making it impossible for mobile devices to render this effect [JMLH01]. To simulate the high quality effect, we approximated light transport using a pre-computed map of local thickness for each model calculated by inverting the normals of the model and estimating ambient occlusion with the inverted normals [BB11]. The medium LOD level substitutes the thickness map with a standard distance-attenuated diffuse lighting combined with the distance-attenuated dot product of the view vector and the inverted light vector. The low quality fall-back is an opaque Blinn-Phong specular shader.

Refraction is a computationally expensive effect for mobile devices. OpenGL ES2.0 devices do not support Multiple Render Targets (MRTs) thus existing methods that estimate refraction for both the front and back interfaces of an object are slow [Wym05]. Single interface refraction produces convincing results. Single interface refraction with chromatic aberration [LKM01] was selected as the high level refraction effect. The medium effect removes chromatic aberration by exchanging the wavelength-dependent sampling of the RGB channels with a single lookup, significantly reducing texture fetches by a factor of three. The low quality effect is a uniformly distorted transparent shader.

Bump Mapping via tessellation and displacement mapping is not available on OpenGL ES2.0 devices. For high quality bump mapping we incorporated the texture-heavy Parallax Occlusion Mapping method [Tat06]. For the medium quality level effect, we employed simple parallax mapping that does not support self-shadowing [KTI*01]. The low quality is a standard normal mapped shader.

6.2. C-LOD Components

The Predictor We implemented our model in the GPU. Our system detects non-canonical object forms by examining object position in relation to the view vector. We utilize object IDs to locate contextually singleton objects. An analytical determination of feature uniqueness would require the calculation of the bi-variate partial derivative of each unit's output. Identifying local maxima in a Gaussian pyramid [ZTTS06] is slow on mobile as it uses render buffer ping ponging. We count local maxima by employing an approximation that exploits hardware's linear interpolation capabilities. We render each unit's output in a 4x4 resolution frame buffer object only once each second. By thresholding 16 texel fetches per unit buffer we count up to 16 local maxima competently. We approximate temporal context calculations by storing up to F low resolution previous frame buffer objects and combine them using hardware blending and an 1D ramp texture storing the pre-calculated logarithmically attenuated function (Equation 3). For the scene schemata and physical singleton factors we re-implemented the detectors of [KDCM14], as well as our extensions described in previous sections. We initialized our model equations using the the weights estimated in Section 5.

The Texel Engine C-LOD's Texel Engine constantly monitors object predictions derived from our attention model. A special 2D texture is updated that works as a material quality lookup table (Figure 9). The columns of the texture correspond to all object/material combinations found in a scene and each row represents a LOD for all object/material combinations. A higher row number signifies a more aggressive simplification overall. Introducing a simplification for object/material combination x in row y imposes that all subsequent rows have the same or lower quality for x. This restriction maintains visual coherence between LODs and induces the smallest possible number of quality reductions. As a result, values over the diagonal of the texture are always white signifying the highest quality possible. The system updates the texture once per second in synchronization with camera movement.

The Bootstrapper The interaction between the graphics processor, CPU and memory of a mobile device is not trivial. When bootstrapping, C-LOD performs system profiling.

Figure 8: *Left to right: Subsurface scattering, refraction and bump mapping low to high quality.*

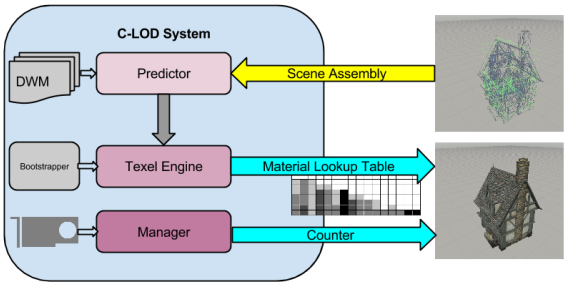


Figure 9: *The C-LOD system architecture.*

The materials managed are initially rendered at their lowest quality. Then, in rapid succession, the quality level of each object's material is increased while frame rate is monitored. This procedure determines a scale factor that controls the aggressiveness of simplifications by the Texel Engine.

The Manager A Finite State Machine (FSM) monitors frame rate during execution. When frame rate drops and motion is detected, a counter is increased. This counter is communicated to all managed materials. A re-mapped object ID of each object is appointed as the u coordinate to sample the look-up table texture and the counter variable as the v coordinate. The sampled value is communicated to the fragment shader where it controls a conditional branch that selects the appropriate LOD for the shader or acts as an iteration counter e.g., for ray marching in the parallax occlusion mapping shader. Updating the counter only when camera moves, reduces luminance offsets and flickering effects. Frame rate is constantly re-evaluated and the counter is increased/decreased to maintain the best LOD for the current conditions (Figure 9).

7. Evaluation of C-LOD

We evaluated C-LOD's efficacy both via eye tracking and by acquiring GPU performance data on a mobile device. We also measured battery performance improvement.

Model Accuracy To measure the model's accuracy in predicting attention we performed an experiment on our eye-tracked HMD set-up.

Design To empirically verify that changes in LOD were not perceived and did not affect attention deployment, we rendered a scene consisting of 50k triangles and complex

Est.	Object gazed	HQ	C-LOD	Total
R	random object	$< 5\%$	$< 5\%$	$< 5\%$
$E1$	1st prediction	40%	42.3%	41.1%
$E2$	1st or 2nd	69.9%	74.8%	72.3%
$E3$	1st or 2nd or 3d	86.9%	92.7%	89.7%

Table 2: *The ratio of frames that the attended object was predicted correctly for the high quality condition, the C-LOD managed condition and in total. E1 denotes that the gazed object matches the first prediction. E2 denotes that the gazed object matches either the first or the second predicted object. E3 denotes that the gazed object matches either the first, or the second or the third object.*

shaders twice. In the first version of the scene (HQ), all effects were set in the highest quality possible. In the second condition (C-LOD) quality was managed by our system. The rendering was performed on a high-end desktop computer to eliminate fluctuations in the frame rate that would have occured in a tablet device inadvertently affecting attention deployment. The FOV of the HMD was restricted to 40 degrees horizontally and 23 degrees vertically to simulate a 10.1" tablet held at a 30cm observer distance [SSC10]. Participants were requested to find and collect seven objects placed in consistent, inconsistent, physically isolated, contextually compound, contextually isolated locations and in a canonical/non-canonical form. In total, 22 people participated (2 female, mean age 22), with 11 people in each of the two conditions.

Results In total, 88,404 object fixations were recorded for all participants (Figure 10). Given that human attention may be directed at multiple foci [AP00], we recorded the three most prominent objects predicted to be fixated by our system for each frame of the simulation. We defined three quantitative estimators to denote the ratio of frames that gaze was allocated in an increasingly larger subset of the predicted objects, to the total number of simulation frames. A baseline R estimator was defined that selects a random object in the FOV for each frame. Both conditions yielded similar results. We summarize the estimators and their results in Table 2. In short, the addition of the C-LOD changes did not alter gaze performance, and thus were most likely not perceived by the participants.

Model Efficiency To assess the impact of C-LOD on GPU performance we reconstructed 2,947 seconds of player

motion of both experimental conditions on an Android quad-core Cortex A9 1.6GHz OpenGL ES2.0 mobile device and sampled the framerate at a 5Hz rate. A total of 17,681 frame rate samples were collected. An independent-samples t-test was conducted, revealing a significant difference between the HQ ($M = 24.05, SD = 2.92$) and C-LOD ($M = 25.6, SD = 1.33$) conditions; $t(8,418) = -44.16, p < 0.0001$. The C-LOD condition exhibits a consistently stabler frame rate and provides a slightly higher mean frame rate when compared to the HQ quality setting (Figure 11). The Android Debug Bridge (ADB) and Tracer for OpenGL tools were employed to conduct a deep frame inspection. C-LOD estimations run for 4ms on average per frame. Given the increase in mean frame rate between the two conditions it can be concluded that this cost is amortized between frames.

Battery life improvement Quering ADB indicated that the battery's average voltage drop was 21mVolts greater for the HQ condition versus the C-LOD managed condition. This indicates an increased discharge rate that was also portrayed in the total run time. Player motion data from the validation experiment were re-played in the HQ and C-LOD settings until battery run out. The C-LOD condition lasted 249 minutes; the HQ condition lasted for 233 minutes.

Discussion Results indicate that C-LOD identifies the observed object 8 times better than a random estimator in the worst case (Table 2). For three attended objects prediction rate approaches 90%. This suggests that quality reductions go mostly unnoticed. Integrating C-LOD in a mobile 3D graphics application stabilizes frame rate without sacrificing perceived quality and boosts battery run time by 6.5% (Figure 11).

8. Conclusion

We presented an extension to the HLSM [KDCM14] by introducing four novel factors that affect attention deployment: object canonical form, contextual singletoness, feature uniqueness and temporal context. We acquired the parameters to initialize our model in a perceptual experiment. We developed a LOD manager for mobile devices that maintains a constant framerate by selecting an appropriate LOD for materials based on attention. We evaluate the performance our algorithm via eye-tracking and by acquiring GPU performance data on mobile devices, confirming that complex effects such as parallax occlusion mapping that are usually omitted in mobile devices can now be employed without exhausting GPU capability. We verified an increase in battery life due to less GPU utilization.

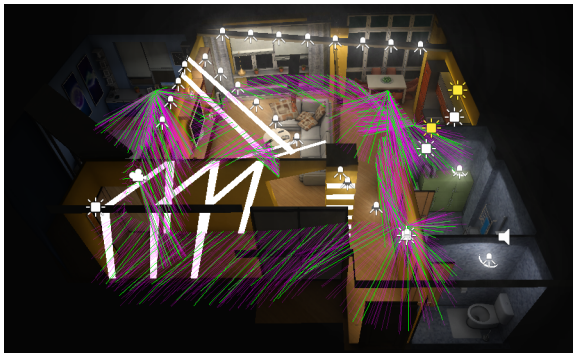


Figure 10: *Our validation tool indicates the subject's gaze point with magenta colored beams. The green beams indicate predictions by our attention model.*

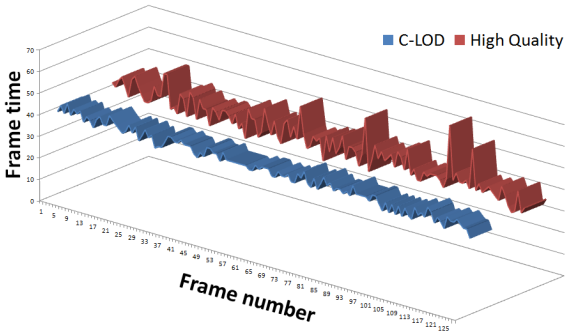
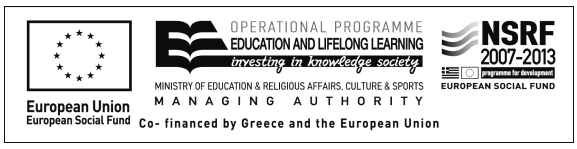


Figure 11: *Frame time for 128 random sequential frames of the HQ and C-LOD conditions. Notice the intense fluctuation of the frame time in the HQ condition when compared to the C-LOD condition.*

Future work includes extending our model with low-level factors for more accurate predictions when gross low level irregularities exist in an image. We will investigate the performance of the proposed LOD manager in dynamic scenes. An attention based cinematography system could be developed that applies post-process effects such as Depth-Of-Field based on attention.

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.



References

[AP00] AWH E., PASHLER H.: Evidence for split attentional foci. *Journal of Experimental Psychology: Human Perception and Performance 26*, 2 (2000), 834. 7

[BB11] BARRE-BRISEBOIS C.: Approximating translucency for

a fast, cheap and convincing subsurface-scattering look. In *Game Developers Conference* (2011). 6

[BPL07] BECKER M. W., PASHLER H., LUBIN J.: Object-intrinsic oddities draw early saccades. *Journal of Experimental Psychology: Human Perception and Performance 33*, 1 (2007), 20. 2, 3

[BT81] BREWER W. F., TREYENS J. C.: Role of schemata in memory for places. *Cognitive Psychology 13*, 2 (1981), 207–230. 1

[BTBV99] BLANZ V., TARR M. J., BÜLTHOFF H. H., VETTER T.: What object attributes determine canonical views? *Perception-London 28*, 5 (1999), 575–600. 3

[CCW03] CATER K., CHALMERS A., WARD G.: Detail to attention: exploiting visual tasks for selective rendering. In *Proceedings of the 14th Eurographics workshop on Rendering* (2003), Eurographics Association, pp. 270–280. 2

[Cla76] CLARK J. H.: Hierarchical geometric models for visible surface algorithms. *Communications of the ACM 19*, 10 (1976), 547–554. 2

[ÇPAM08] ÇAPIN T. K., PULLI K., AKENINE-MÖLLER T.: The state of the art in mobile graphics research. *IEEE Computer Graphics and Applications 28*, 4 (2008), 74–84. 1, 2, 6

[Duc07] DUCHOWSKI A.: *Eye tracking methodology: Theory and practice*, vol. 373. Springer, 2007. 3

[Eck98] ECKSTEIN M. P.: The lower visual search efficiency for conjunctions is due to noise and not serial attentional processing. *Psychological Science 9*, 2 (1998), 111–118. 1, 2

[FRC10] FRINTROP S., ROME E., CHRISTENSEN H. I.: Computational visual attention systems and their cognitive foundations: A survey. *ACM Transactions on Applied Perception (TAP) 7*, 1 (2010), 6. 3, 4

[GS*66] GREEN D. M., SWETS J. A., ET AL.: *Signal detection theory and psychophysics*, vol. 1. Wiley New York, 1966. 3

[HLRC*10] HILLAIRE S., LÉCUYER A., REGIA-CORTE T., COZOT R., ROYAN J., BRETON G.: 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* (2010), ACM, pp. 191–198. 2

[HWJH99] HENDERSON J. M., WEEKS JR P. A., HOLLINGWORTH A.: The effects of semantic consistency on eye movements during complex scene viewing. *Journal of experimental psychology: Human perception and performance 25*, 1 (1999), 210. 2, 3

[IKN*98] ITTI L., KOCH C., NIEBUR E., ET AL.: A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on pattern analysis and machine intelligence 20*, 11 (1998), 1254–1259. 2, 3

[JMLH01] JENSEN H. W., MARSCHNER S. R., LEVOY M., HANRAHAN P.: A practical model for subsurface light transport. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques* (2001), ACM, pp. 511–518. 6

[KDCM14] KOULIERIS G. A., DRETTAKIS G., CUNNINGHAM D., MANIA K.: *An Automated High Level Saliency Predictor for Smart Game Balancing*. Tech. rep., Technical University of Crete, 2014. 1, 2, 3, 4, 5, 6, 8

[Kof35] KOFFKA K.: *Principles of Gestalt psychology*. Harcourt, Brace and World, New York, NY., 1935. 3

[KTI*01] KANEKO T., TAKAHEI T., INAMI M., KAWAKAMI N., YANAGIDA Y., MAEDA T., TACHI S.: Detailed shape representation with parallax mapping. In *Proceedings of ICAT* (2001), vol. 2001, pp. 205–208. 6

[LDC06] LONGHURST P., DEBATTISTA K., CHALMERS A.: 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* (2006), ACM, pp. 21–29. 3

[LKC09] LEE S., KIM G. J., CHOI S.: Real-time tracking of visually attended objects in virtual environments and its application to lod. *Visualization and Computer Graphics, IEEE Transactions on 15*, 1 (2009), 6–19. 2

[LKM01] LINDHOLM E., KILGARD M. J., MORETON H.: A user-programmable vertex engine. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques* (2001), ACM, pp. 149–158. 6

[LM00] LOSCHKY L. C., MCCONKIE G. W.: User performance with gaze contingent multiresolutional displays. In *Proceedings of the 2000 symposium on Eye tracking research & applications* (2000), ACM, pp. 97–103. 2

[Lue03] LUEBKE D. P.: *Level of detail for 3D graphics*. Morgan Kaufmann, 2003. 1, 2, 3, 6

[Owe05] OWENS J.: Streaming architectures and technology trends. In *ACM SIGGRAPH 2005 Courses* (2005), ACM, p. 9. 2

[PC84] POSNER M. I., COHEN Y.: Components of visual orienting. *Attention and performance X: Control of language processes 32* (1984), 531–556. 3

[Ren00] RENSINK R. A.: The dynamic representation of scenes. *Visual cognition 7*, 1-3 (2000), 17–42. 2

[SL97] SIMONS D. J., LEVIN D. T.: Change blindness. *Trends in cognitive sciences 1*, 7 (1997), 261–267. 3

[SLF*11] SECORD A., LU J., FINKELSTEIN A., SINGH M., NEALEN A.: Perceptual models of viewpoint preference. *ACM Transactions on Graphics (TOG) 30*, 5 (2011), 109. 3

[SSC10] SLATER M., SPANLANG B., COROMINAS D.: Simulating virtual environments within virtual environments as the basis for a psychophysics of presence. *ACM Transactions on Graphics (TOG) 29*, 4 (2010), 92. 7

[SSWR08] SUNDSTEDT V., STAVRAKIS E., WIMMER M., REINHARD E.: A psychophysical study of fixation behavior in a computer game. In *Proceedings of the 5th symposium on Applied perception in graphics and visualization* (2008), ACM, pp. 43–50. 2

[Tat06] TATARCHUK N.: Dynamic parallax occlusion mapping with approximate soft shadows. In *Proceedings of the 2006 symposium on Interactive 3D graphics and games* (2006), ACM, pp. 63–69. 6

[TG02] THEEUWES J., GODIJN R.: Irrelevant singletons capture attention: Evidence from inhibition of return. *Perception & Psychophysics 64*, 5 (2002), 764–770. 1

[TMD83] TOLHURST D. J., MOVSHON J. A., DEAN A.: The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision research 23*, 8 (1983), 775–785. 3

[WK06] WALTHER D., KOCH C.: Modeling attention to salient proto-objects. *Neural Networks 19*, 9 (2006), 1395–1407. 4, 5

[Wym05] WYMAN C.: An approximate image-space approach for interactive refraction. *ACM Transactions on Graphics (TOG) 24*, 3 (2005), 1050–1053. 6

[ZMM09] ZOTOS A., MANIA K., MOURKOUSSIS N.: A schema-based selective rendering framework. In *Proceedings of the 6th Symposium on Applied Perception in Graphics and Visualization* (2009), ACM, pp. 85–92. 2

[ZTTS06] ZIEGLER G., TEVS A., THEOBALT C., SEIDEL H.-P.: On-the-fly point clouds through histogram pyramids. In *Workshop on Vision, Modeling, and Visualization (VMV 2006)* (2006), pp. 137–144. 6