Context-aware Gaze Prediction applied to Game Level Design, Level-of-Detail and Stereo Manipulation

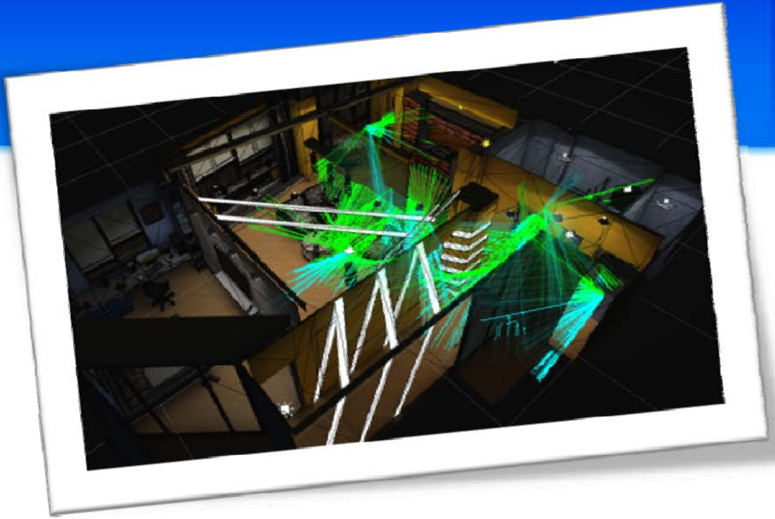
PhD Candidate
George Alex Koulieris
Computer Science Division, MUSIC Lab
Department of Electronic & Computer Engineering
Technical University of Crete

Supervisor
Assoc. Prof. Katerina Mania

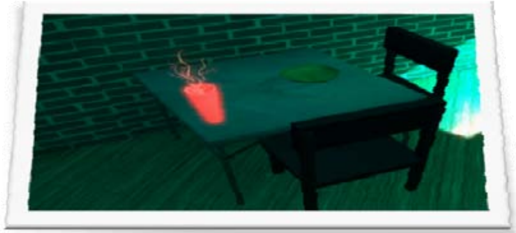
Collaborators
Dr George Drettakis, Graphics and Design with Heterogeneous Content (GRAPHDECO), INRIA Sophia Antipolis, France
Prof. Douglas W. Cunningham, Brandenburgische Technische Universitat Cottbus, Germany
September 9th, 2015

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
EUROPEAN SOCIAL FUND

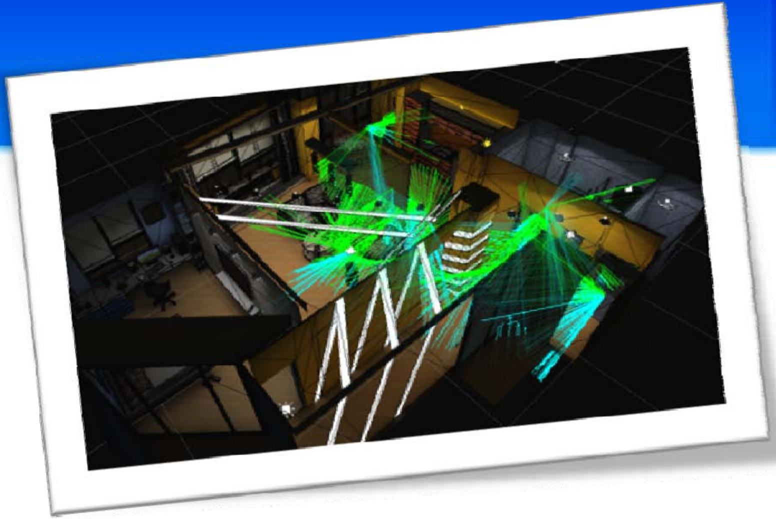
Overview



- Motivation
- Problem Statement
- Contributions
- Background
- Part 1: An Automated High-Level Saliency Predictor for Smart Game Balancing
- Part 2: C-LOD: Context-aware Material Level-of-Detail applied to Mobile Graphics
- Part 3: Gaze Prediction using Machine Learning for Dynamic Stereo Manipulation
- Summary



Overview



Motivation

Problem Statement

Contributions

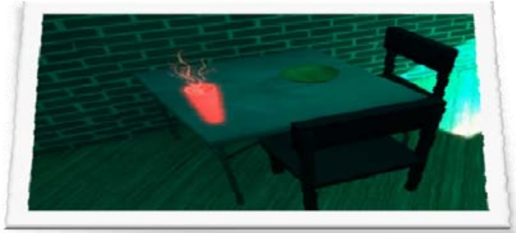
Background

Part 1: An Automated High-Level Saliency Predictor for Smart Game Balancing

Part 2: C-LOD: Context-aware Material Level-of-Detail applied to Mobile Graphics

Part 3: Gaze Prediction using Machine Learning for Dynamic Stereo Manipulation

Summary

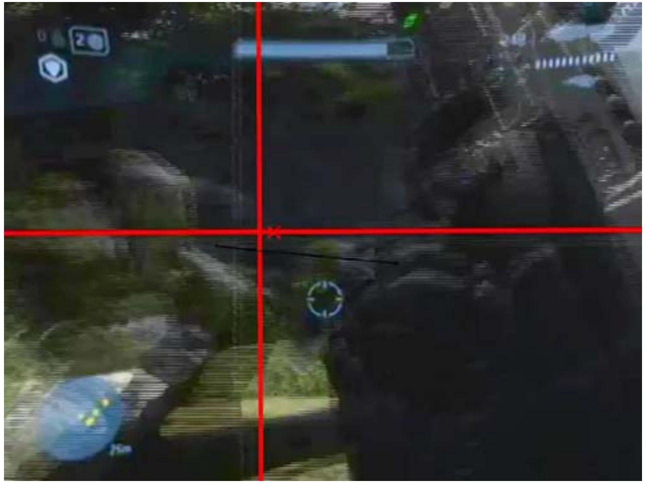


Gaze-aware applications
Motivation

Predicting gaze over a synthetic scene

Knowledge of attended regions in a synthetic scene yields three classes of applications:

- Manipulating observers' gaze to fixate on selected 3D objects based on application goals
- Saving computational resources in regions that will not be attended by reducing rendering quality
- Applying gaze-aware visual effects, e.g. stereo manipulation or tone mapping



VIDEO: ASL Gaze Tracking

Gaze-aware applications
Context

Game Level Design: Manipulating observers' gaze to fixate on selected 3D objects based on application goals

- Many game genres gameplays rely on a target detection task to solve riddles and advance gameplay
  - "Find the object"
- Object placement during level design correlates to search task completion time
  - Implicitly defining game difficulty
- **Currently, designing game levels is a tedious manual operation** (*Pagulayan et al., 2003*)
- **An accurate gaze prediction model allows us to relocate objects estimated to attract visual attention, implicitly adjusting game difficulty**

Gaze-aware applications
Context

Level-of-Detail (LOD): Saving computational resources in regions that will not be attended by reducing rendering quality

- **Higher** rendering quality on areas expected to attract visual attention
- Rendering quality **decreased** for not attended scene areas
- **Current attempts on gaze-aware LOD limited due to limited prediction accuracy** (*Cater et al., 2003*)
- **The accuracy of the gaze prediction model determines if quality reductions are perceived**

LOD on Mobile Devices

- Interest recently renewed due to explosive growth of the mobile market
- Prohibitive hardware restrictions of mobiles for complex effects

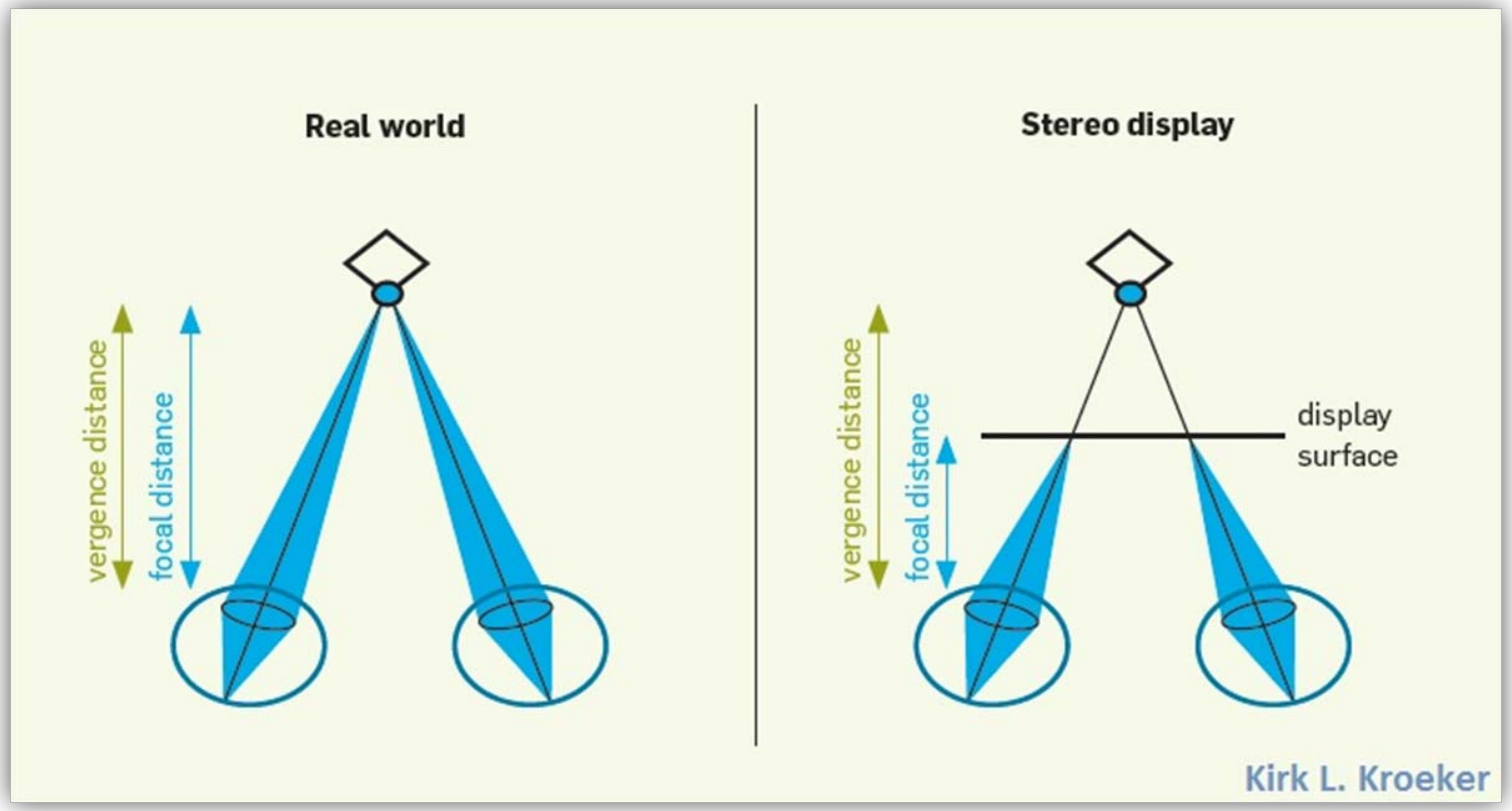
Gaze-aware applications
Context

Comfortable Stereo: Applying gaze-aware visual effects

- Recent popularity of commodity Head Mounted Displays, e.g Oculus Rift
- **Vergence-Accommodation Conflict**

Gaze-aware applications

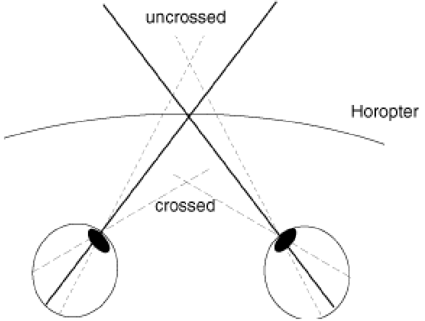
Vergence-Accommodation Conflict



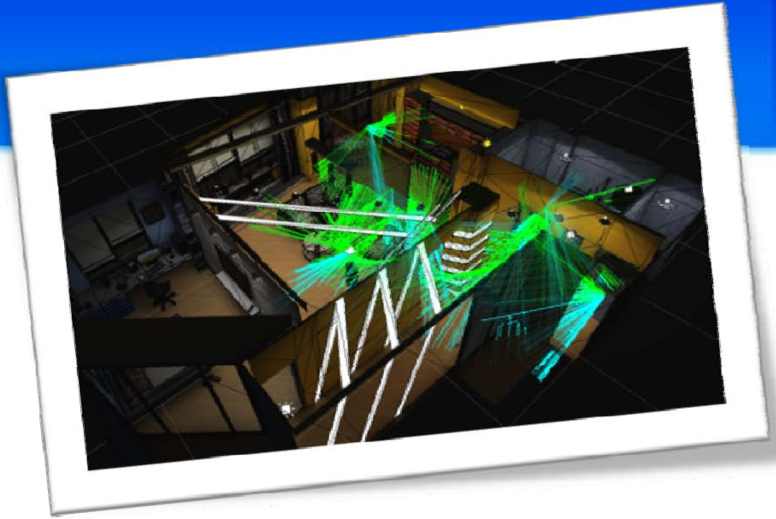
Gaze-aware applications
Context

Comfortable Stereo: Applying gaze-aware visual effects

- **Current *stereo grading* methods based on saliency estimates**
  - Are offline (*Lang et al., 2010*)
  - Apply a global scene depth compression irrespective of attended objects (*Oskam et al., 2011*)
- Need for reduced discomfort and eye-fatigue **without sacrificing depth (cardboarding)**
- **Relocate attended objects in the *disparity comfort zone*!**



Overview



Motivation

Problem Statement

Contributions

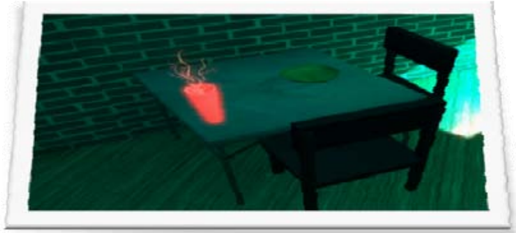
Background

Part 1: An Automated High-Level Saliency Predictor for Smart Game Balancing

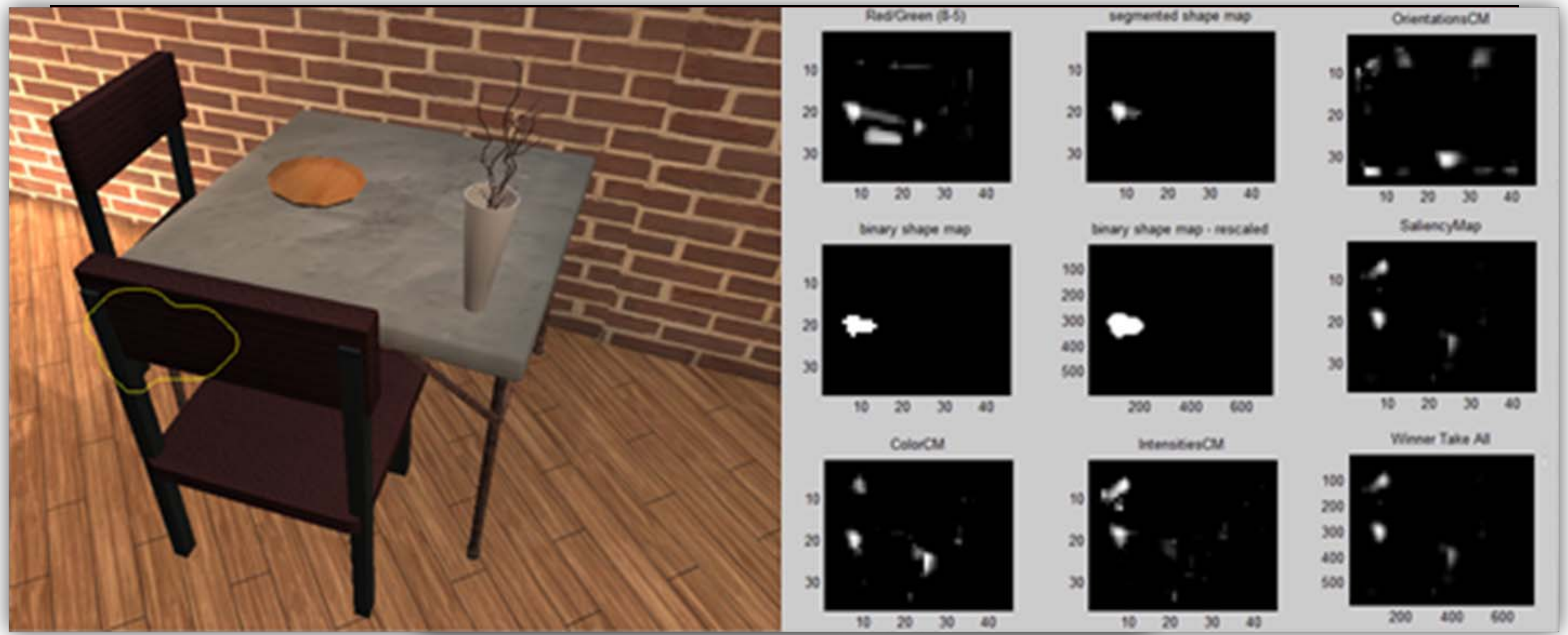
Part 2: C-LOD: Context-aware Material Level-of-Detail applied to Mobile Graphics

Part 3: Gaze Prediction using Machine Learning for Dynamic Stereo Manipulation

Summary



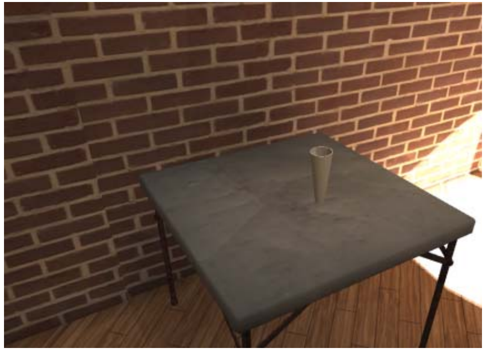
Gaze-aware applications
Problem Statement



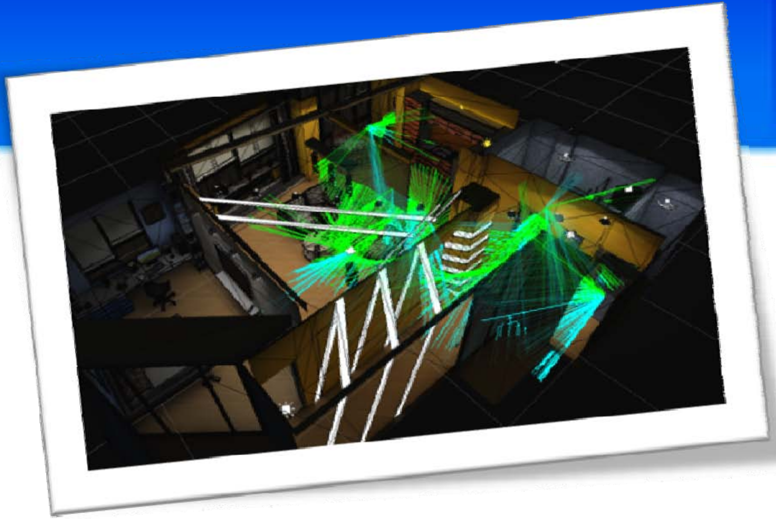
Gaze-aware applications
Problem Statement

However, existing low level, image-based gaze prediction models often fail to predict gaze

- High-level properties such as **scene semantics, attention to objects** or **task** affect gaze fixations
- **Object-context semantic relationships are NOT TAKEN into account by existing low level visual attention models**
- It is a **challenging** problem to quantify qualitative object-context relationships



Overview



Motivation

Problem Statement

Contributions

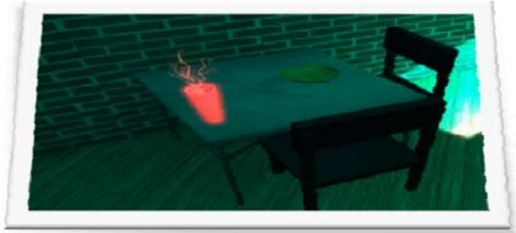
Background

Part 1: An Automated High-Level Saliency Predictor for Smart Game Balancing

Part 2: C-LOD: Context-aware Material Level-of-Detail applied to Mobile Graphics

Part 3: Gaze Prediction using Machine Learning for Dynamic Stereo Manipulation

Summary



Contributions

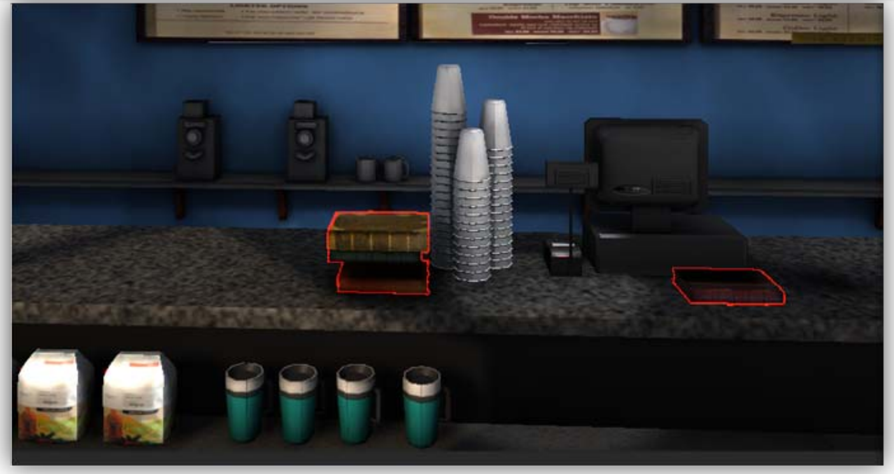
Contribution 1

1. A physically plausible High Level Saliency Model (HLSM)

Encodes six hypotheses from perception and cognitive science into mathematical equations that describe

- Semantic inter-object relationships (e.g. contextual validity, object-context consistency)
- Intra-object positional properties (e.g, object rotation)
- Object topology in terms of inter-object distances and placement (e.g. object isolation)

We developed a computational model that can estimate fixation guidance based on context



Contributions

Contribution 2

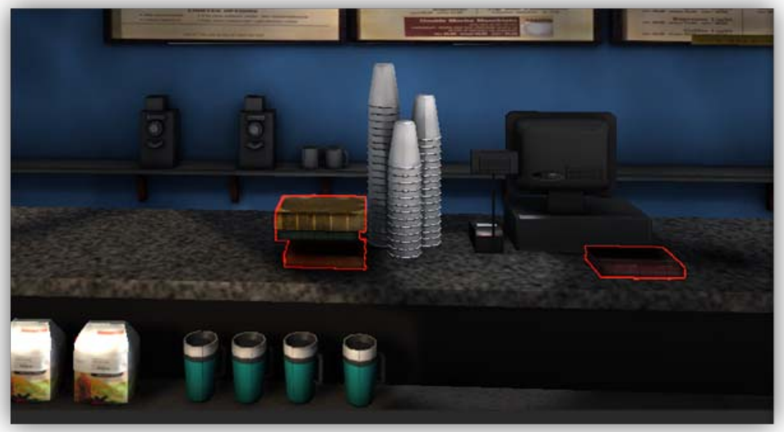
2. A Machine Learning (ML) based predictor

Our Decision Forests-based predictor

- Is automatic avoiding the need for embedding contextual information to objects
- Yields high prediction success rates, **learning** from ground truth eye tracking data
- Supports object motion in contrast to previous task-based and high level approaches

Contributions

Two models of visual attention



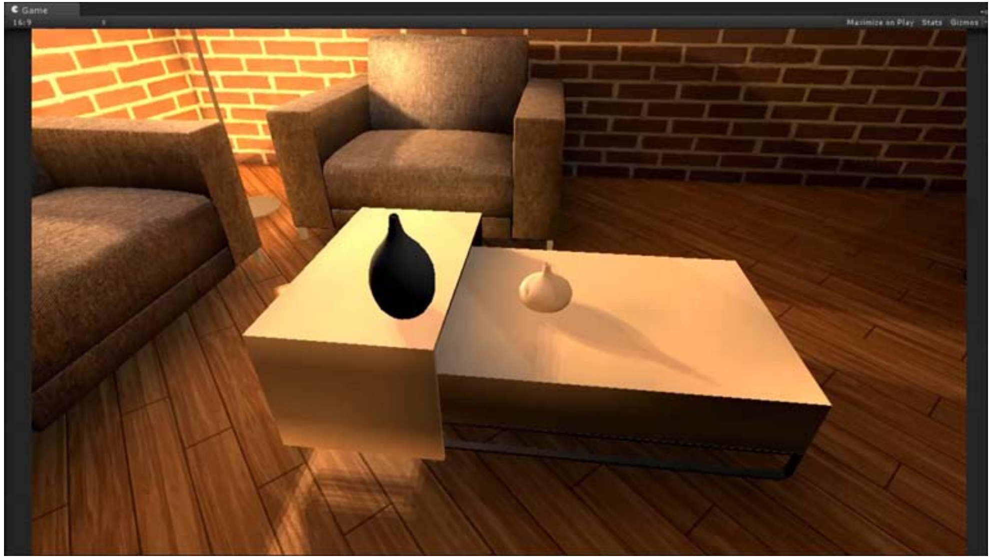
Each model specializes to a different class of applications

- The HLSM does not need eye tracking during model formation; needs semantic object tagging
- The ML-based model does not require object tagging; needs an eye tracker during model formation

Contributions
Contribution 3

3. A game balancing paradigm based on attention

- We develop a system based on the High Level Saliency Model to automatically predict gaze in real-time
- We validate the system's efficacy in adjusting game difficulty by altering object location based on predicted gaze



VIDEO

Contributions

Contribution 4

4. A Level-of-Detail method based on attention for mobile devices

- We present a perceptually optimized renderer for mobile platforms that reduces computation complexity and maintains a more stable frame rate by automatically removing perceptually non-important details
- Enables the usage of - otherwise omitted - complex effects such as subsurface scattering & complex refraction in low-power devices by applying them sparingly only in regions that are expected to be attended

VIDEO

Contributions

Contribution 5

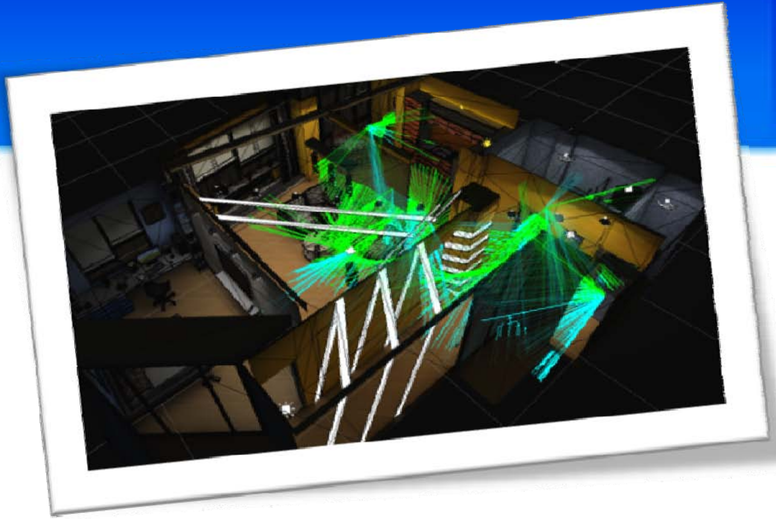
5. A gaze-based stereo manipulation method

Our gaze-aware, localized stereo grading approach

- Provides a greater sense of scene depths than previous global depth compressing methods
- Maintains the same levels of viewing comfort when compared to the state-of-the-art

VIDEO

Overview

Motivation

Problem Statement

Contributions

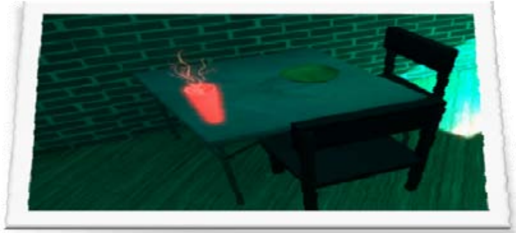
Background

Part 1: An Automated High-Level Saliency Predictor for Smart Game Balancing

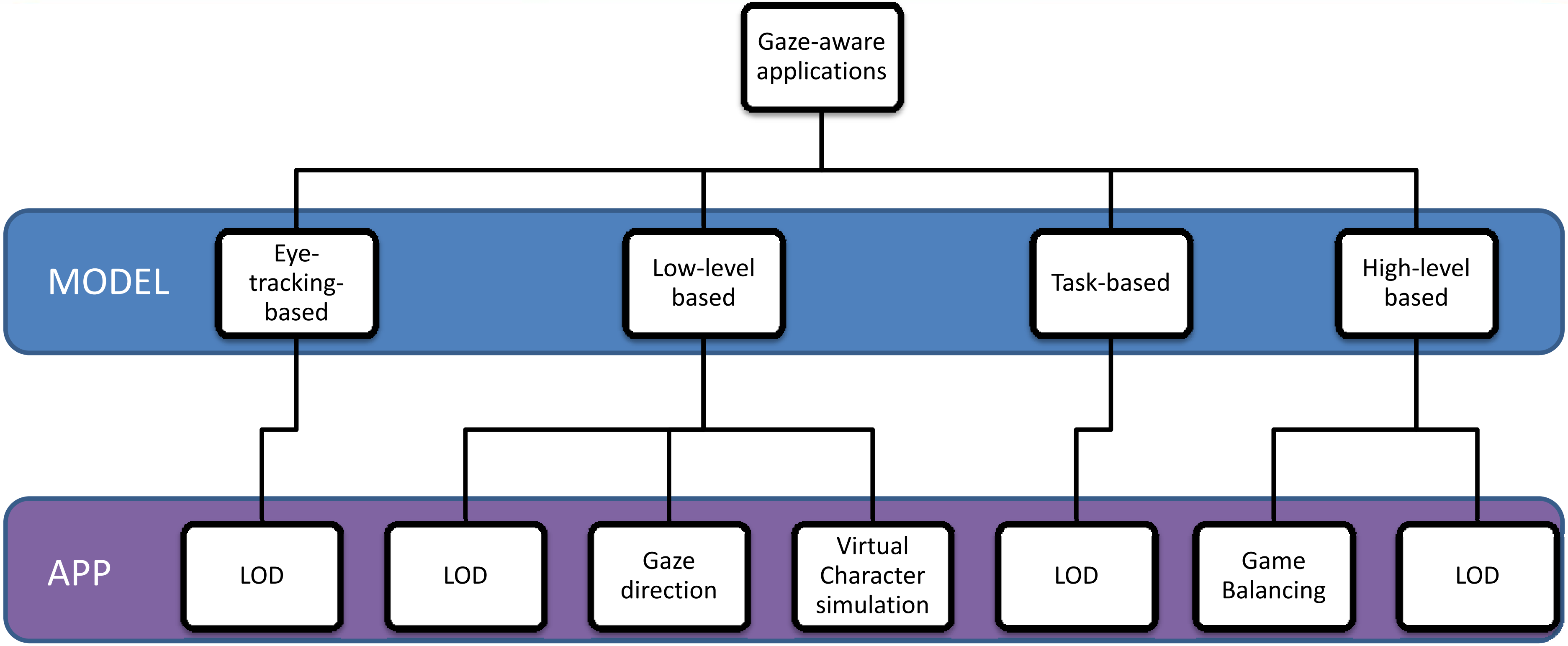
Part 2: C-LOD: Context-aware Material Level-of-Detail applied to Mobile Graphics

Part 3: Gaze Prediction using Machine Learning for Dynamic Stereo Manipulation

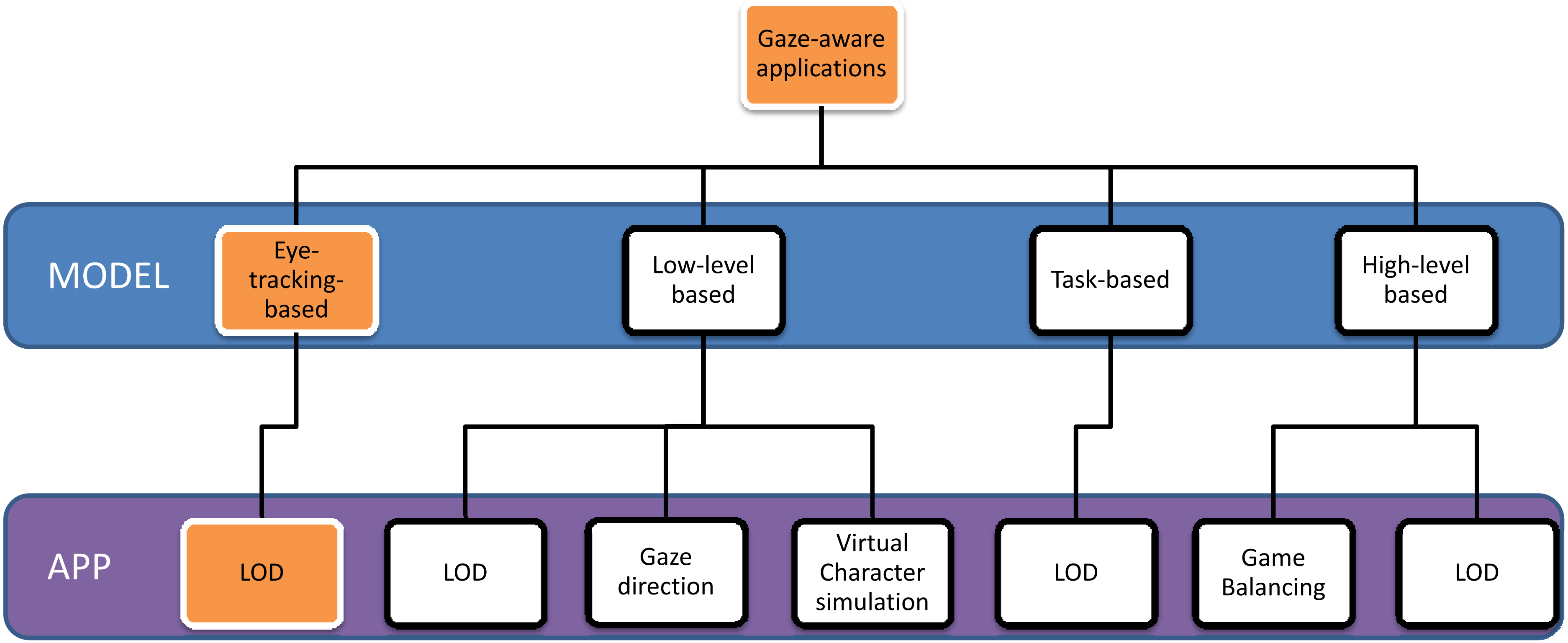
Summary



Outline


Gaze-aware applications
MODEL
Eye-tracking-based
Low-level based
Task-based
High-level based
APP
LOD
LOD
Gaze direction
Virtual Character simulation
LOD
Game Balancing
LOD

Outline

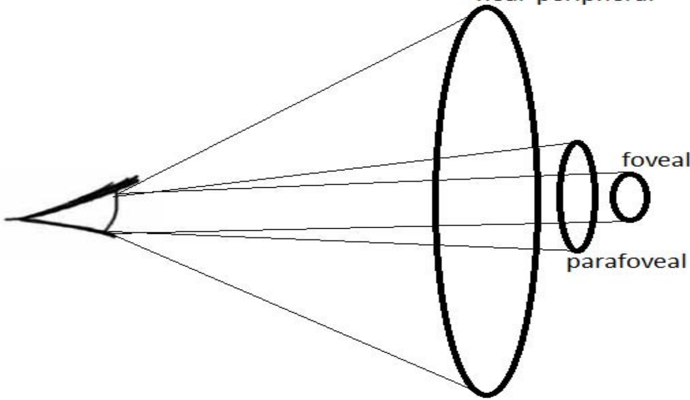

Gaze-aware applications
MODEL
Eye-tracking-based
Low-level based
Task-based
High-level based
APP
LOD
LOD
Gaze direction
Virtual Character simulation
LOD
Game Balancing
LOD

Eye-tracking based gaze prediction

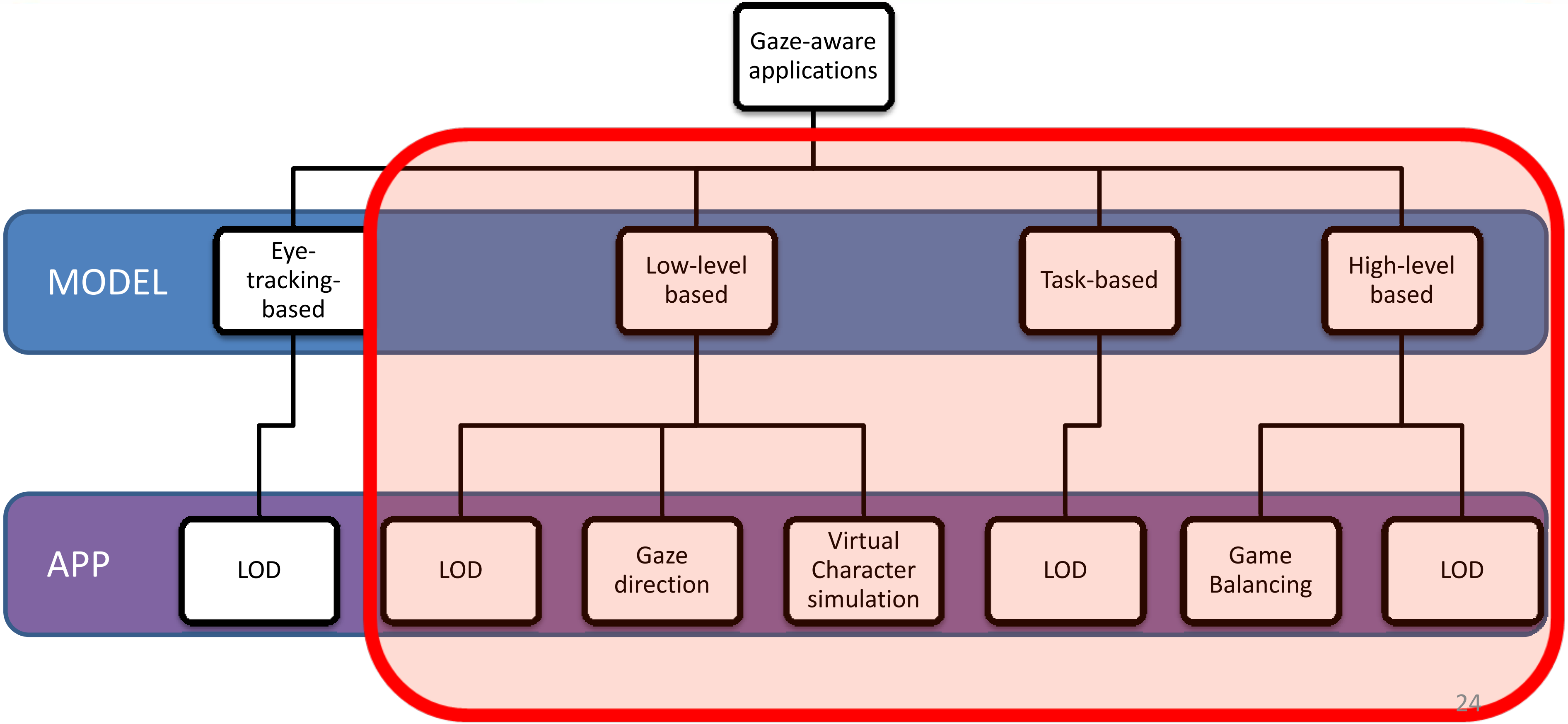
Level-of-Detail

However!

- Latency
- Rarely available for common applications



Outline

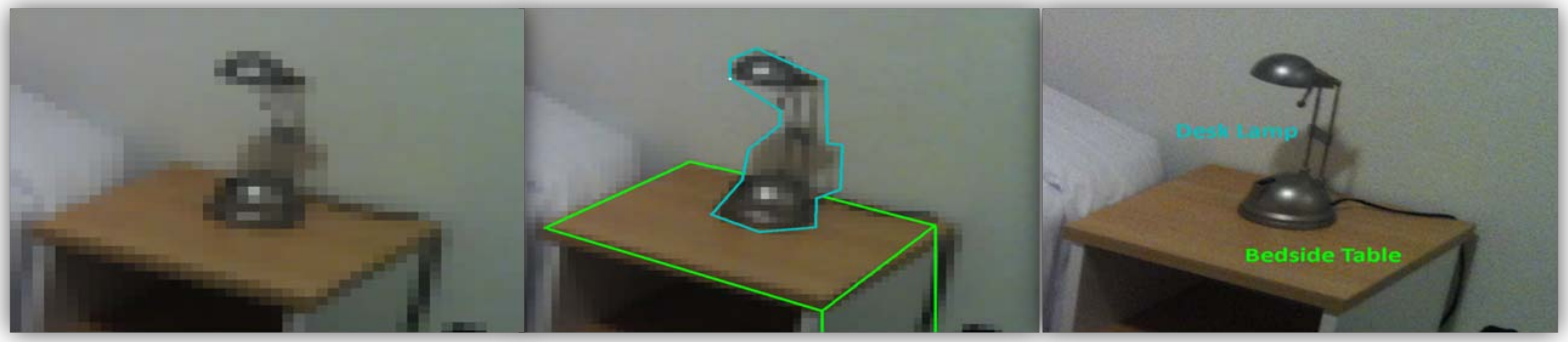

Gaze-aware applications
MODEL
APP
Eye-tracking-based
Low-level based
Task-based
High-level based
LOD
LOD
Gaze direction
Virtual Character simulation
LOD
Game Balancing
LOD

Visual Perception

Levels of abstraction

Active extraction and manipulation of environmental information ***Marr et al., 1982, Shipley et al., 2001***

- Low-level processes extract image regularities e.g. edges, color
- Mid-level processes combine regularities to form features e.g. object shapes
- High-level processes map mid-level features to meaning and semantics



Focal Attention

The selection mechanism of the brain

Mid- and high-level processes have limited resources

- Focal attention selects a few low-level features that are likely to be important
- Low-level features e.g. edges may attract focal attention almost reflex-like

Mid- & high-level features and goal-oriented properties can direct Focal Attention

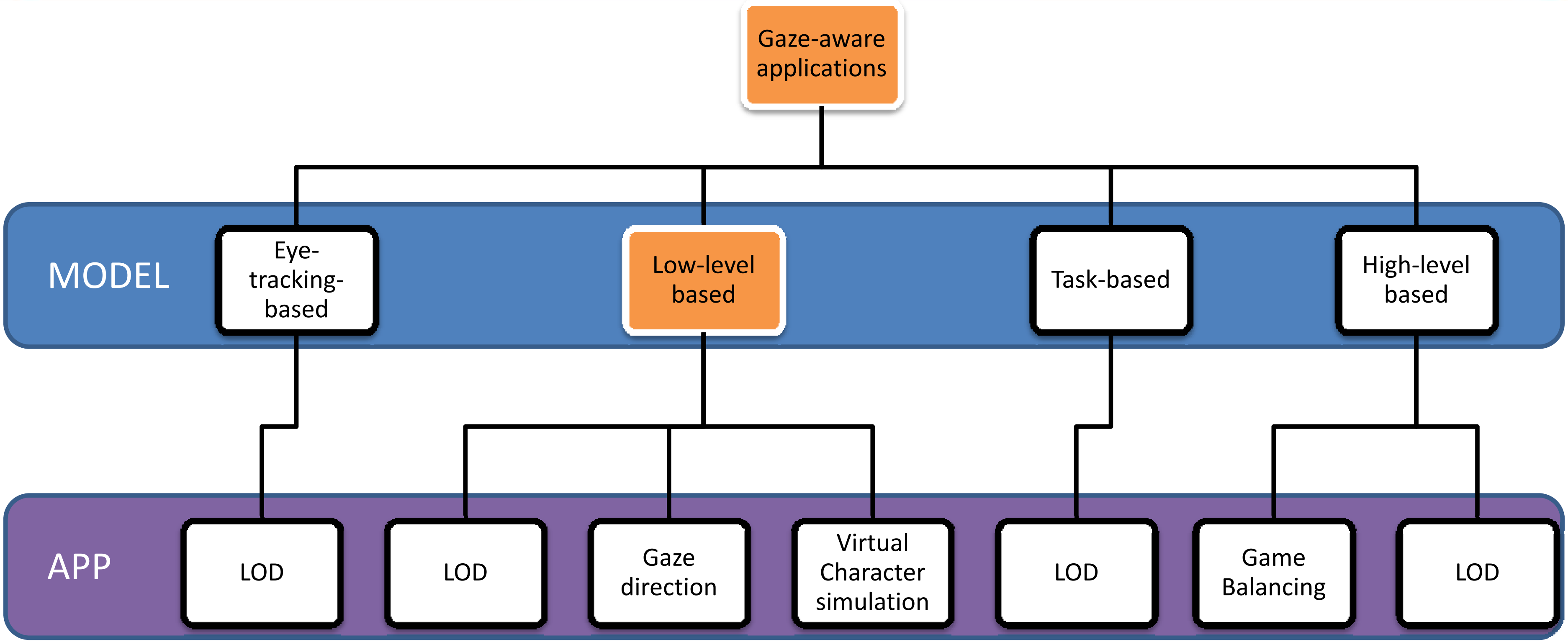
The contextual validity of an object's location affects visual search

✓ When looking for a chimney, usually we direct our gaze first to the rooftops

How are these features combined?

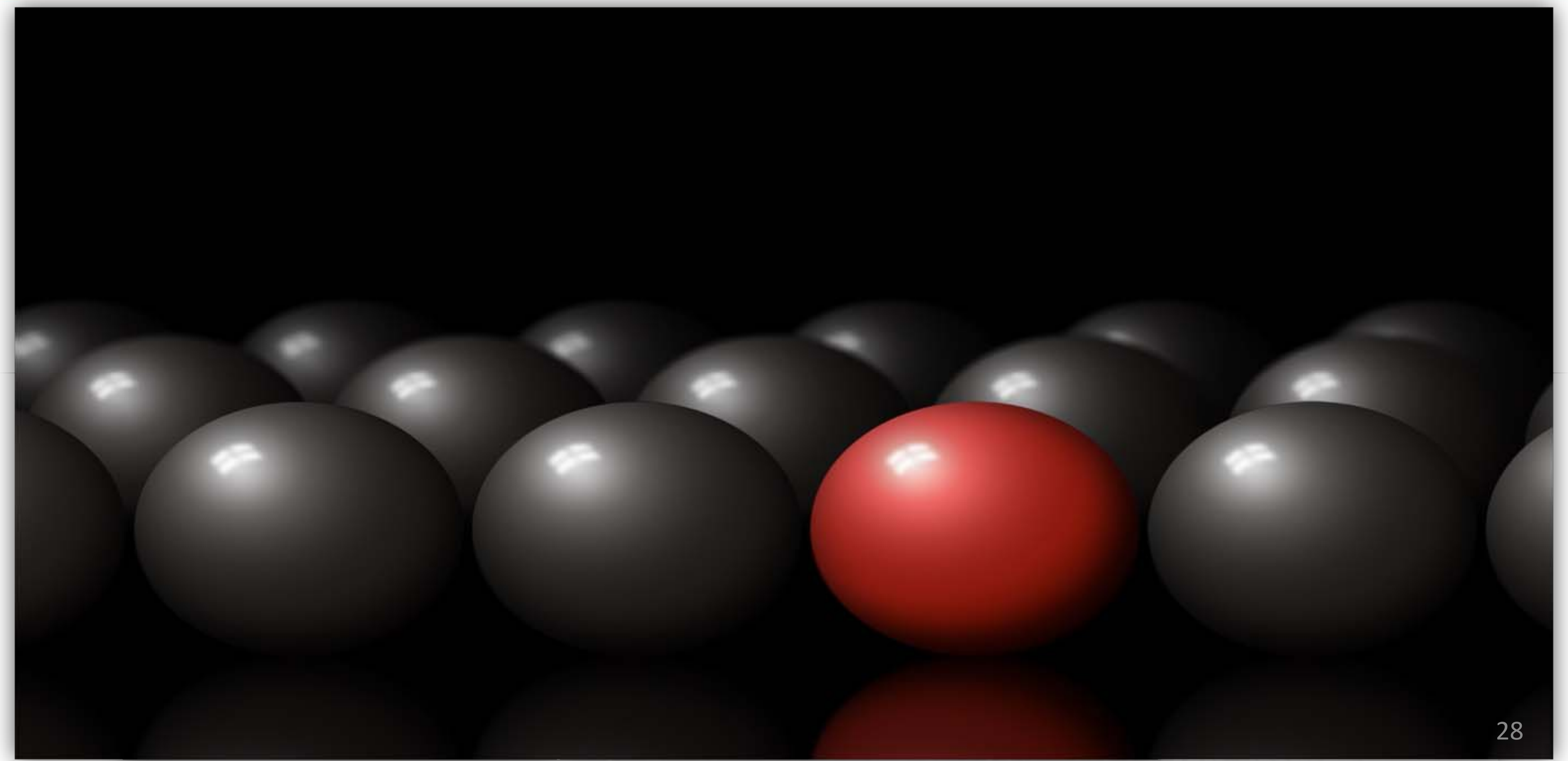
The relative contribution of low-, mid-, and high-level factors on attention is unknown

Outline


Gaze-aware applications
MODEL
Eye-tracking-based
Low-level based
Task-based
High-level based
APP
LOD
LOD
Gaze direction
Virtual Character simulation
LOD
Game Balancing
LOD

Attention prediction

Saliency based on Low-level image features



Focal Attention Models

Feature Integration Theory (FIT): A two staged-model

Stage 1

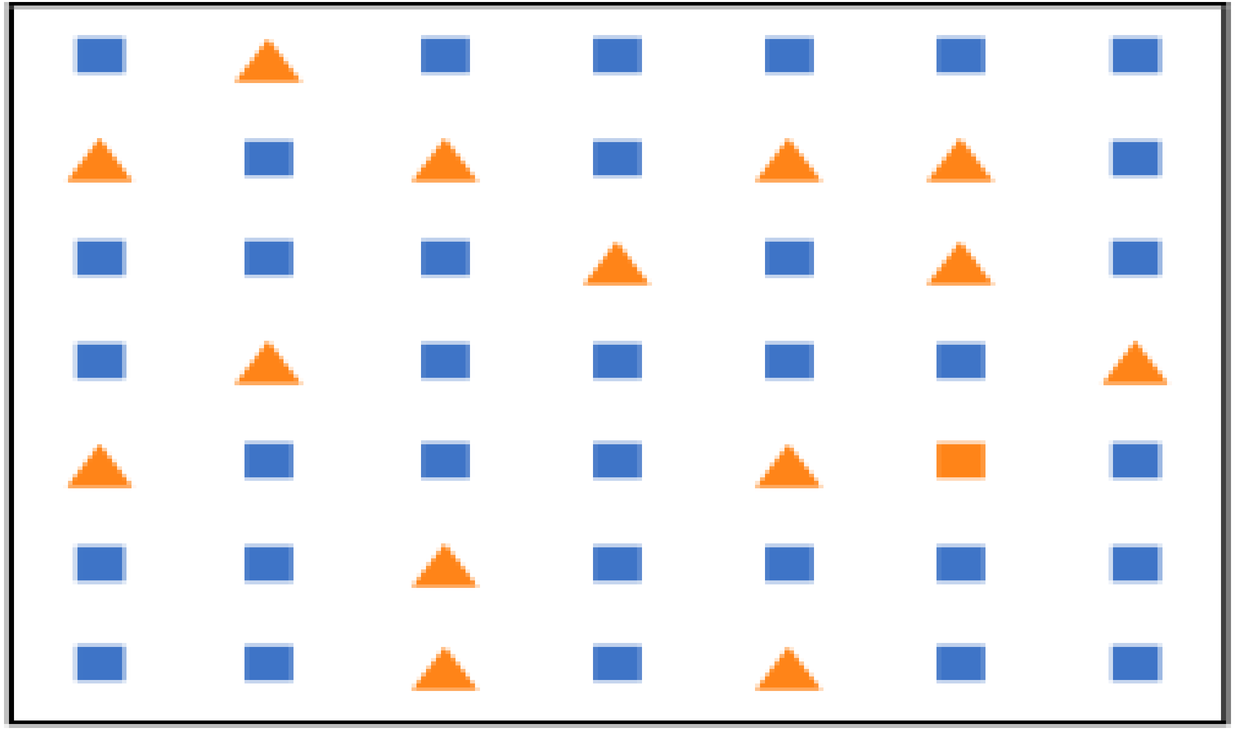
- Low-leve
- Focal att

Stage 2

Low level fe

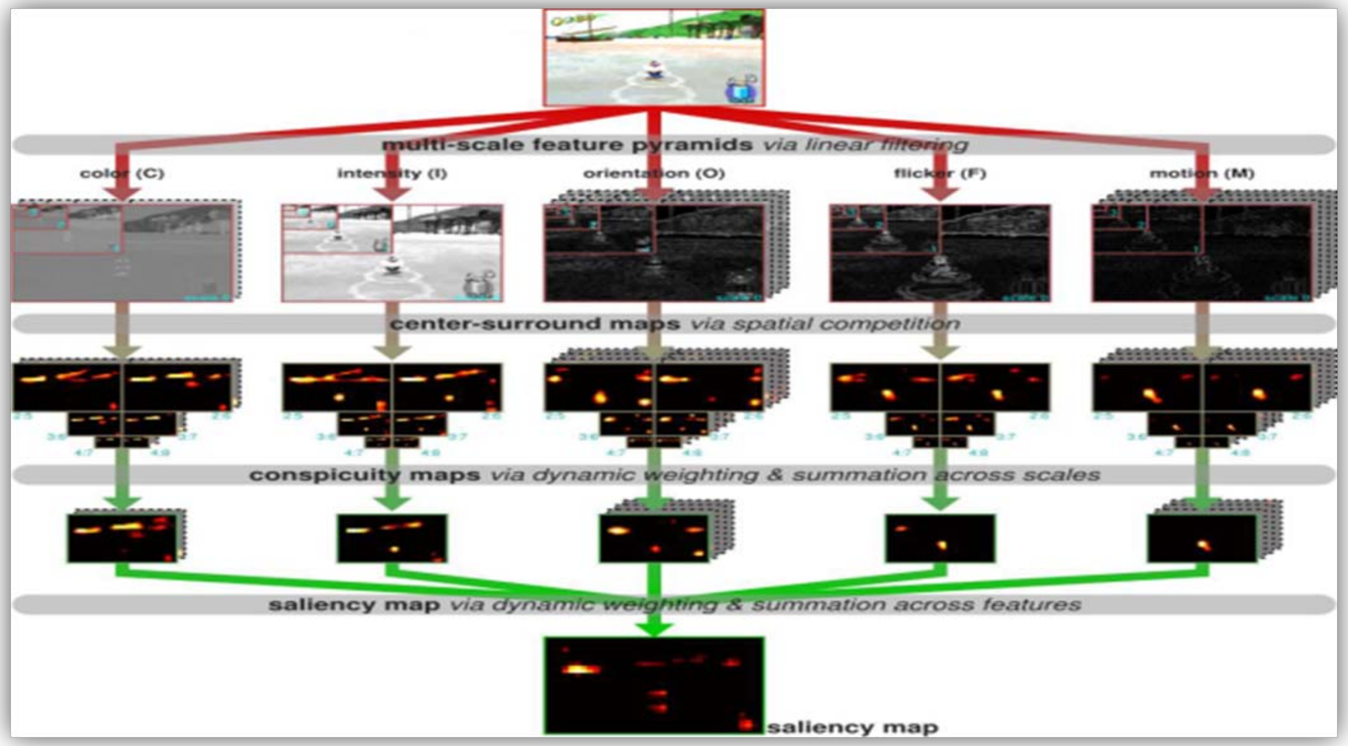
serial, one r

a slow,



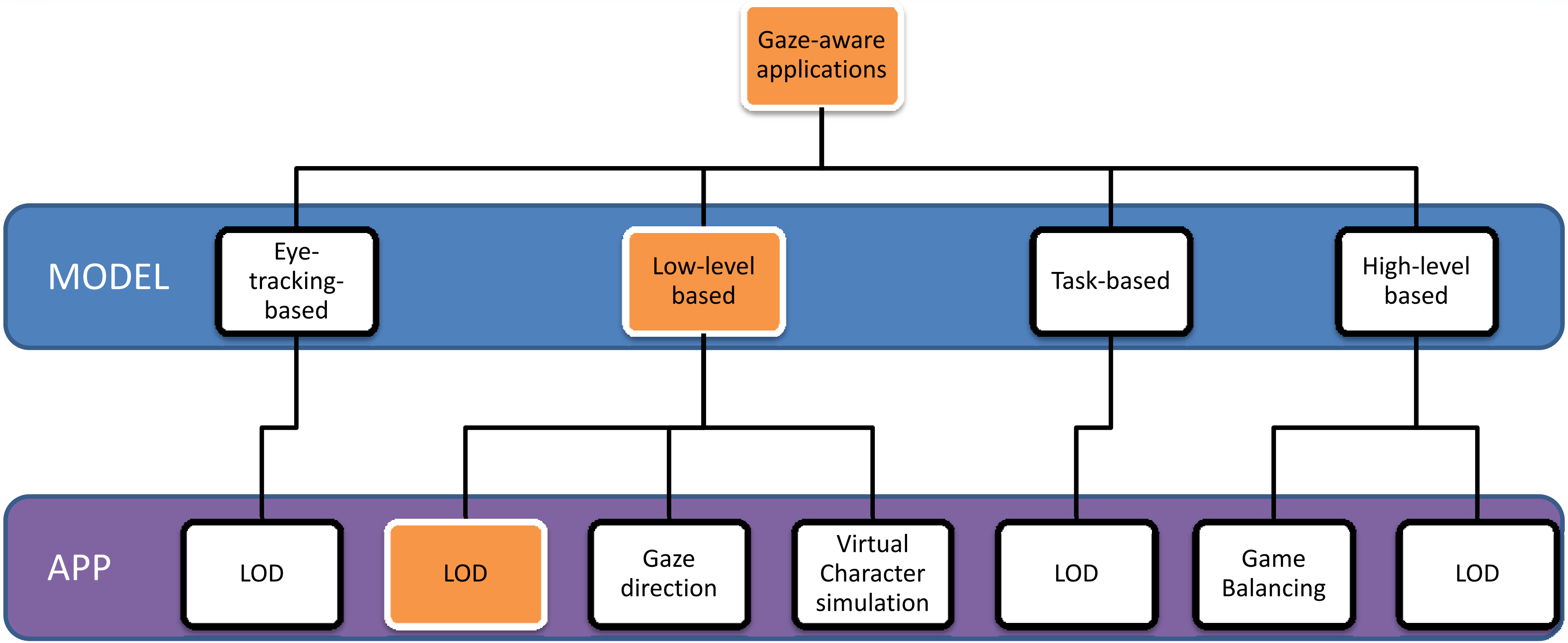
Feature Integration Theory

A computational model



(Itti et al., 1998)

Outline


Gaze-aware applications
MODEL
Eye-tracking-based
Low-level based
Task-based
High-level based
APP
LOD
LOD
Gaze direction
Virtual Character simulation
LOD
Game Balancing
LOD

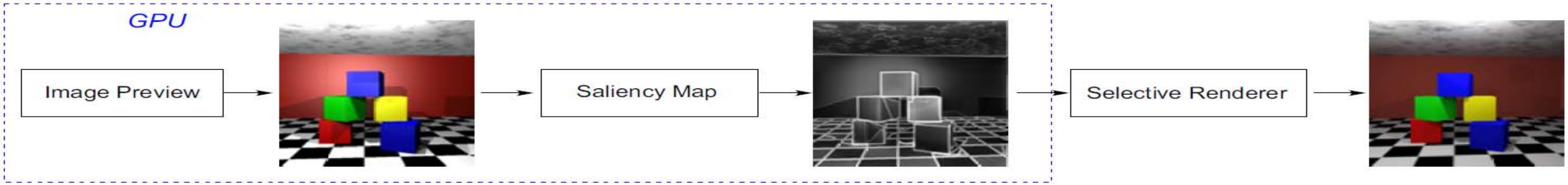
Low-level attention-aware applications

Level-of-Detail

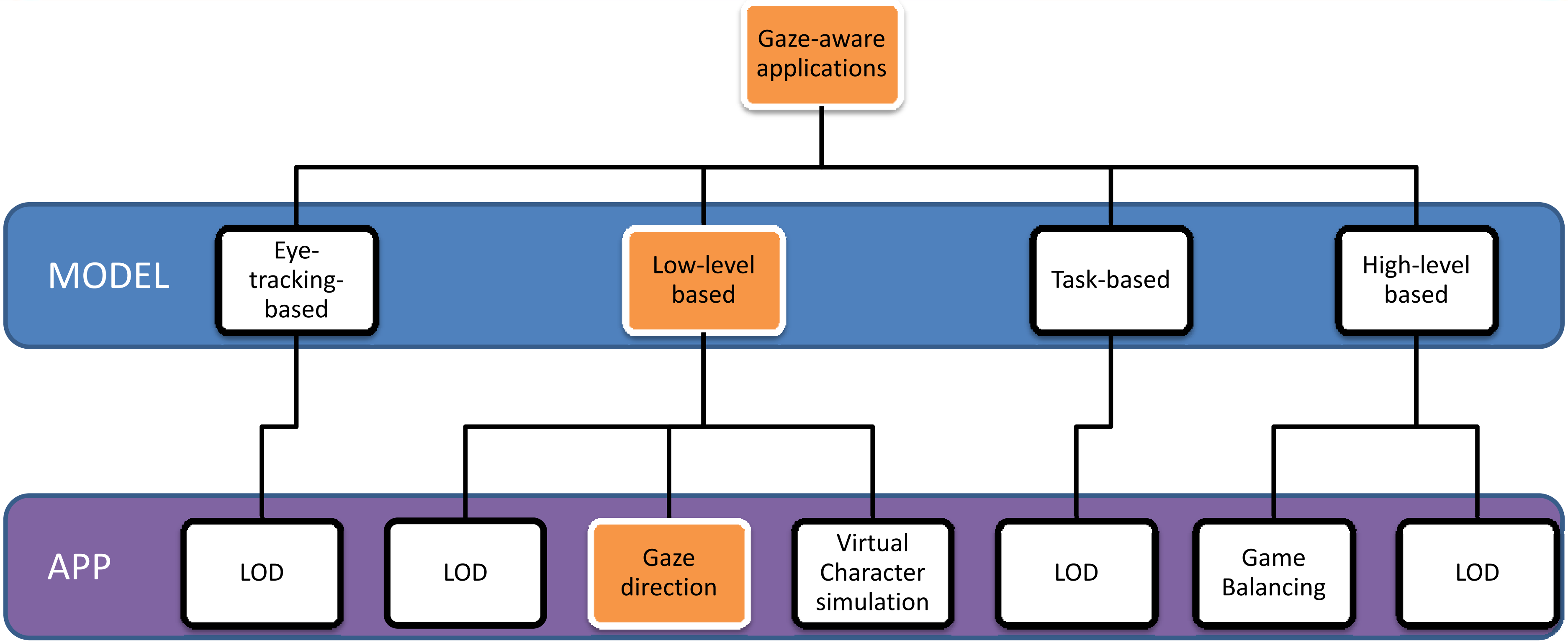
FIT-guided selective rendering often fails

Suffers from low prediction accuracy when

- A Task is being conducted
- High level semantic context properties drive attention **top-down**; e.g. when searching for an object



Outline


Gaze-aware applications
MODEL
Eye-tracking-based
Low-level based
Task-based
High-level based
APP
LOD
LOD
Gaze direction
Virtual Character simulation
LOD
Game Balancing
LOD

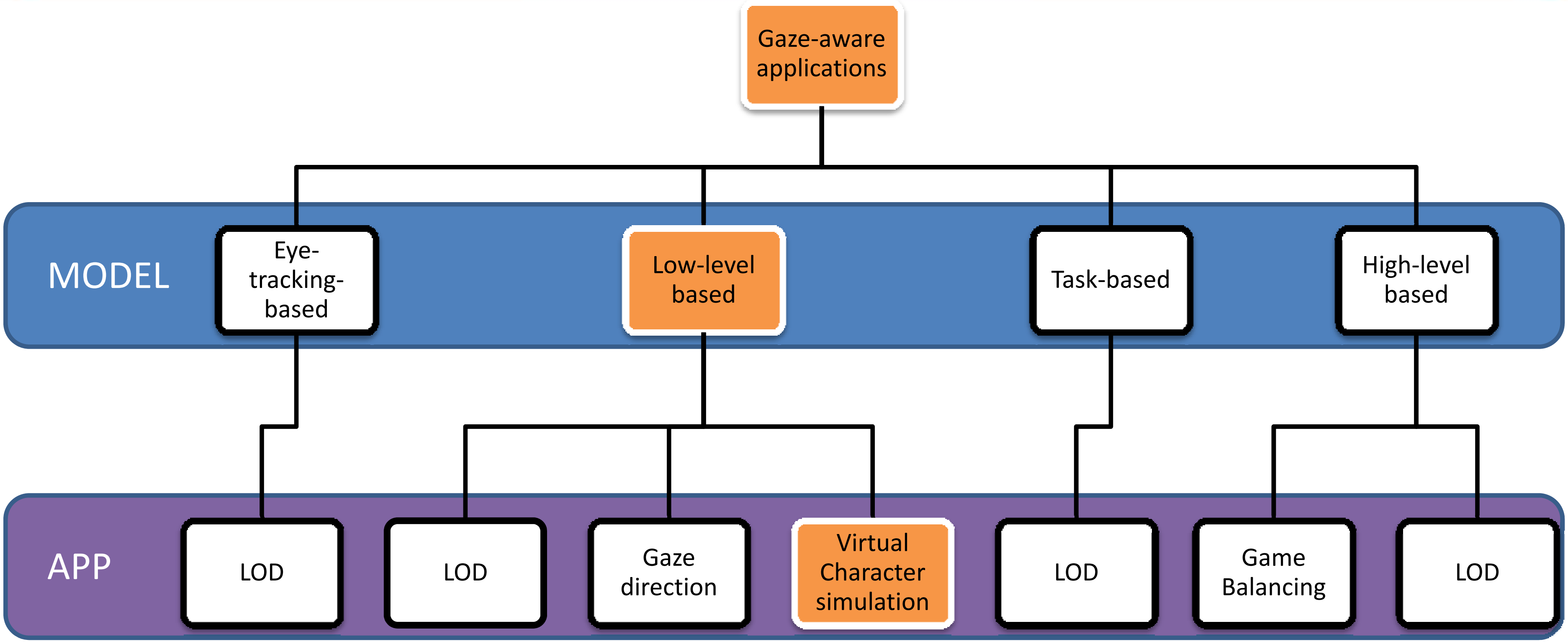
Gaze Direction

Low-level-based guiding principles

In-game advertising

- The modified item MUST share perceptual features with a target item
- Requires manual 3D-model modifications

Outline

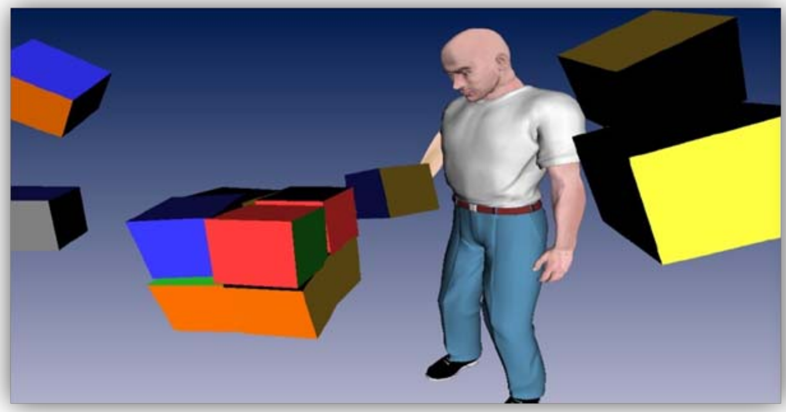

Gaze-aware applications
MODEL
Eye-tracking-based
Low-level based
Task-based
High-level based
APP
LOD
LOD
Gaze direction
Virtual Character simulation
LOD
Game Balancing
LOD

Simulating gaze behavior

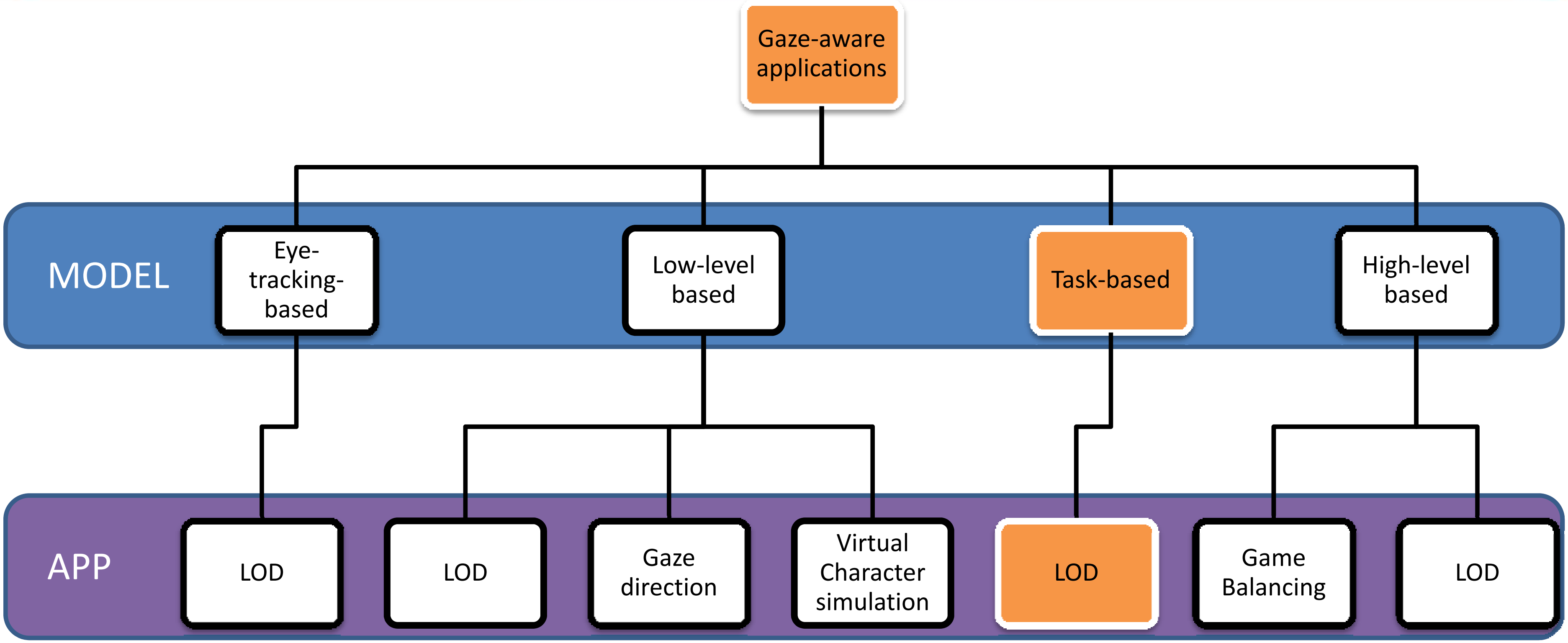
Characters and crowds

Simulating gaze behavior

- Limited to low-level salient objects
- Characters and crowds not responding naturally to tasks



Outline


Gaze-aware applications
MODEL
Eye-tracking-based
Low-level based
Task-based
High-level based
APP
LOD
LOD
Gaze direction
Virtual Character simulation
LOD
Game Balancing
LOD

Task-related Saliency

Saliency based on task demands

Task is predetermined

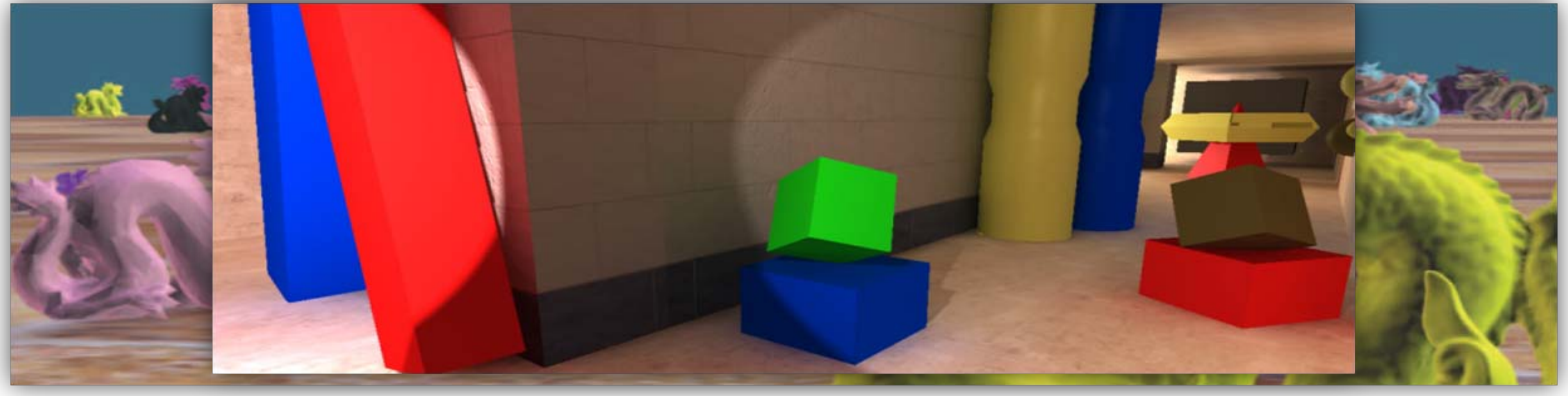
- Task has to be pre-determined
- Approach is limited since it requires labor intensive definition of task restrictions

Task-related Saliency

Combining task-based methods and low level features

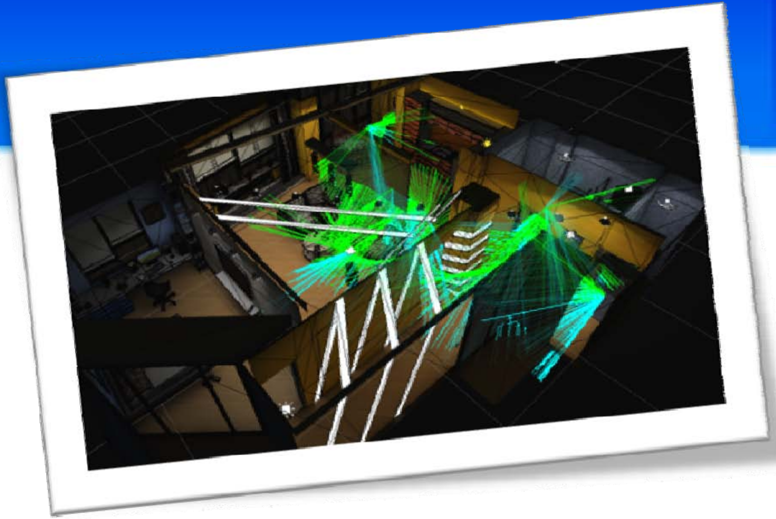
Joined low level & task-based methods

- Suffer from the issues of both models
  - Context-agnostic
  - Require manual definition of task restrictions



Thesis Contributions

Overview



Motivation

Problem Statement

Contributions

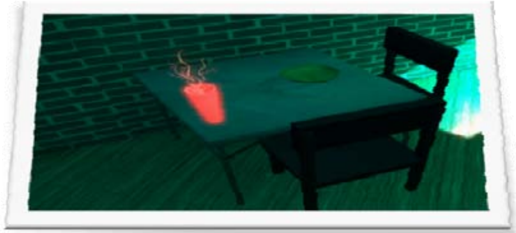
Background

Part 1: An Automated High-Level Saliency Predictor for Smart Game Balancing

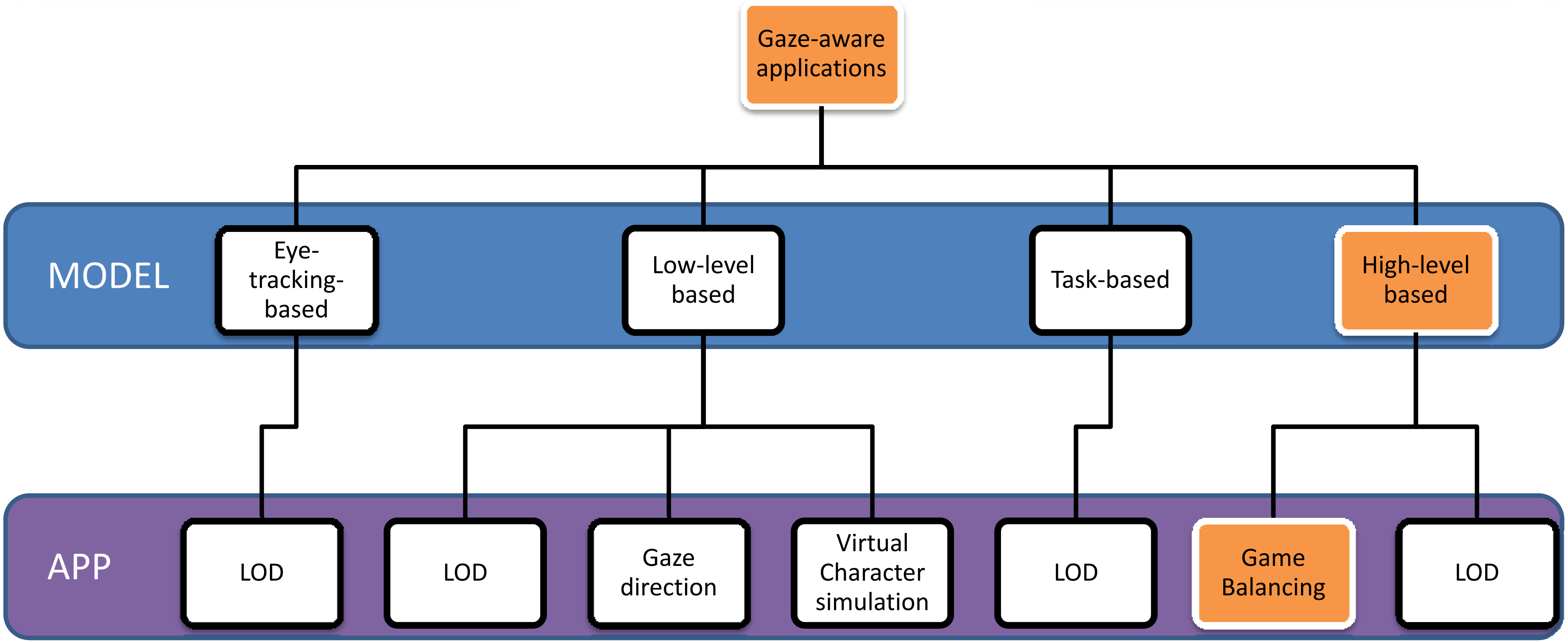
Part 2: C-LOD: Context-aware Material Level-of-Detail applied to Mobile Graphics

Part 3: Gaze Prediction using Machine Learning for Dynamic Stereo Manipulation

Summary



Outline


Gaze-aware applications
MODEL
Eye-tracking-based
Low-level based
Task-based
High-level based
APP
LOD
LOD
Gaze direction
Virtual Character simulation
LOD
Game Balancing
LOD

High Level Saliency

Mapping visual representations to meaning and semantics

A High Level Saliency Predictor

- Our goal is to define a computational model applicable to any context; a challenging task
- When attending a scene, recently acquired knowledge from attentional processing is combined with pre-existing knowledge about a context, e.g. "bedroom"

However!

Until recently a model that explicitly links in a **physiologically plausible** manner high level saliency hypotheses with attention deployment was missing

High Level Saliency

Hypotheses

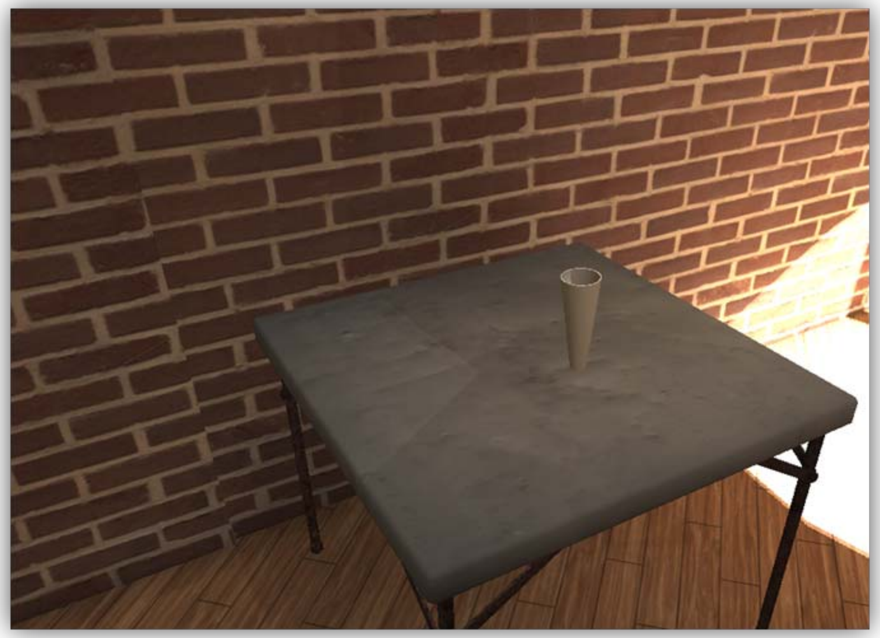
Scene Schemata ***Brewer & Treyens, 1981***

Out-of-context objects are salient

High Level Saliency

Hypotheses

Object Singletoness	***Theeuwes & Godijn, 2002***
Physically isolated objects pop out	



High Level Saliency

Scene schemata & Object singletoness

VIDEO

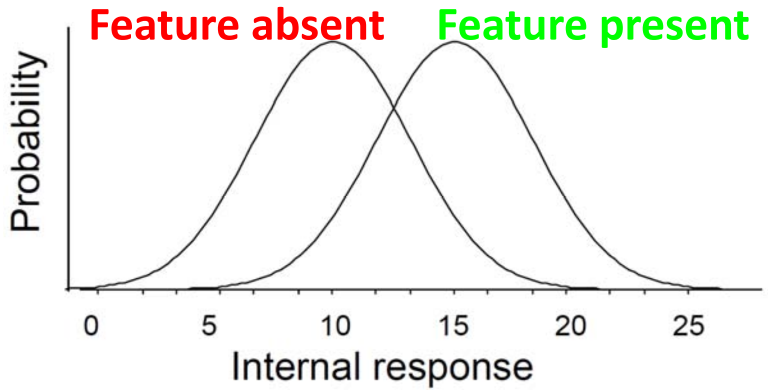
Attentional Processing

The Differential Weighting Model (DWM) *Eckstein, 1998*

A single stage model

- Attentional processing via Gaussian combination rules
- Firing behavior encoded in Bayesian priors
- For every pixel of every frame:

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



Extending the Differential Weighting Model

Defining units for scene schemata and physical isolation states

Extending the Differential Weighting Model

Extending the Differential Weighting Model

Scene schema & physical isolation sensory units

High Level Sensory Units

We extended the original DWM equations to describe two novel high-level sensory units tuned to:

- Schema inconsistencies
- Singleton state of objects

Extending the Differential Weighting Model

Scene schema & physical isolation sensory units

Averaging

We average the components using weights obtained from perceptual studies exploring the effect of High Level hypotheses on task completion time

Viewport *Likelihood ratios* *Saliency*

Perceptual Study

Obtaining experimental data during gameplay

Motivation

We conducted three experiments for three reasons:

- Effect of scene schemata and singletons in VEs unknown
- Empirical classification of objects in scene schemata
- Extension of the DWM requires weighting factors for the two hypotheses

Perceptual Study

Experiment 1

Experiment 1: Classifying objects into schemata

- Empirical classification of scene objects in relation to their surroundings
- 50 objects assessed
- 21 people rated objects using a 7-point Likert scale
  - "How likely for each item is to appear in a given scene?"

Perceptual Study

Experiments 2 & 3

Experiments 2 & 3: Determining the effect of Semantic and Physical Context on task completion time

- Effect of physical and semantic manipulations on task completion time
- Inspired by a real Adventure game
- Search task (Exp. 2): Participants knew exactly what they were searching for
- Non-Search task (Exp. 3): Participants did not know what they were searching for; free exploration of the environment

Perceptual Study

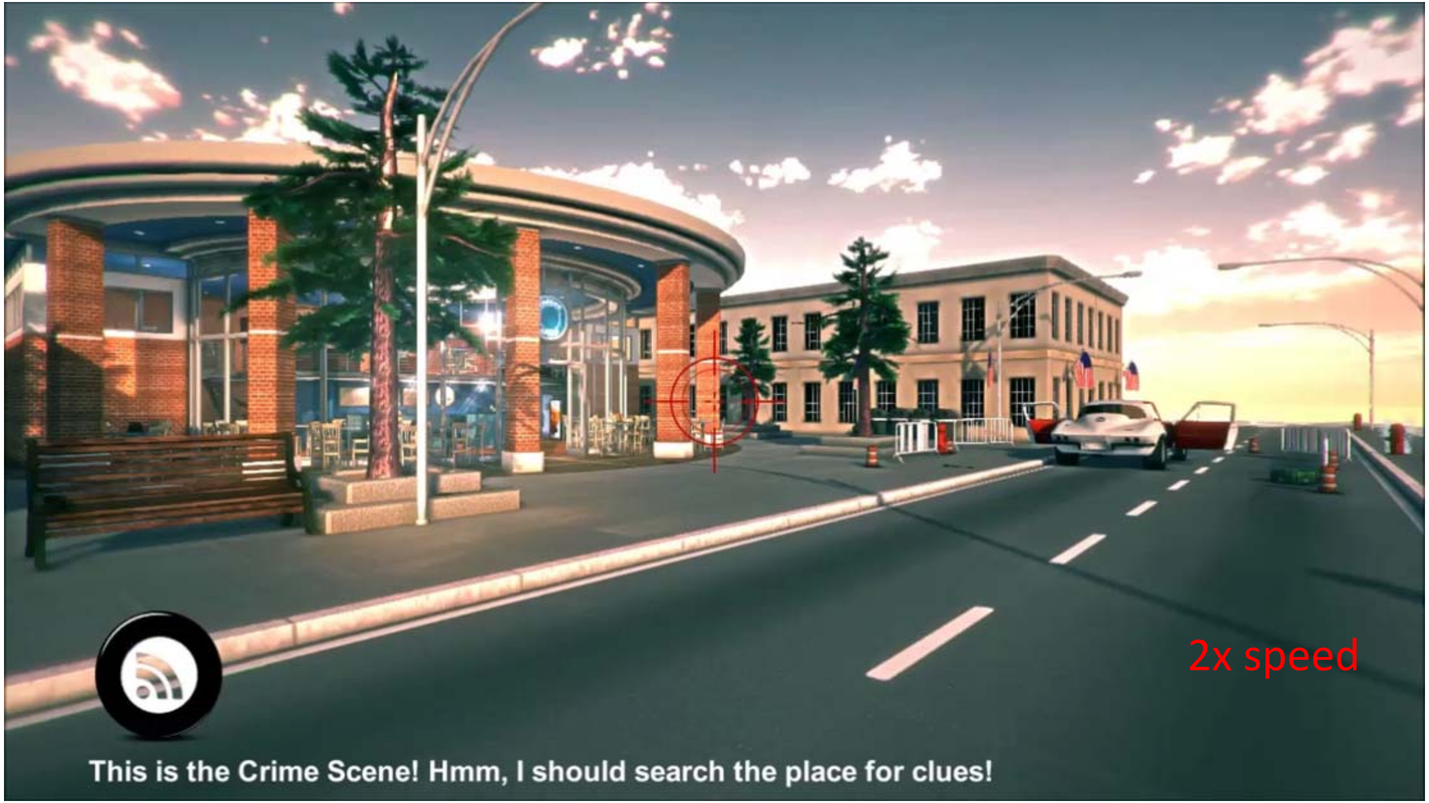
Experiments 2 & 3

Stimuli

- Factorial combination of factors - 4 conditions generated - 2 Tasks
  1. Consistent/Compound
  2. Inconsistent/Compound
  3. Consistent/Singleton
  4. Inconsistent/Singleton
- 80 people, 10/condition

Perceptual Study

Sample Search task

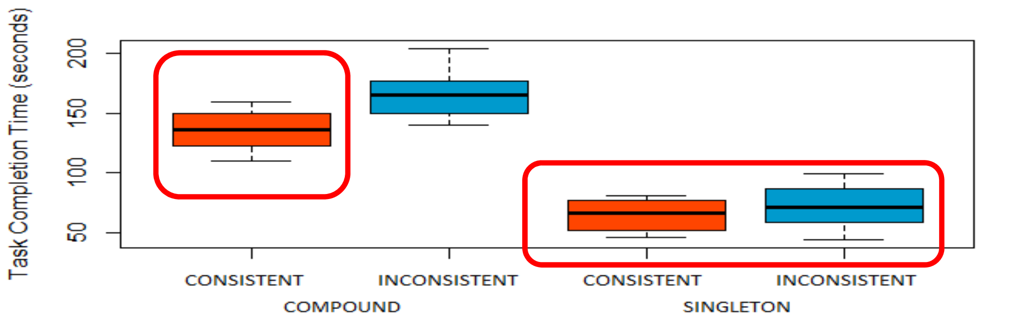


VIDEO

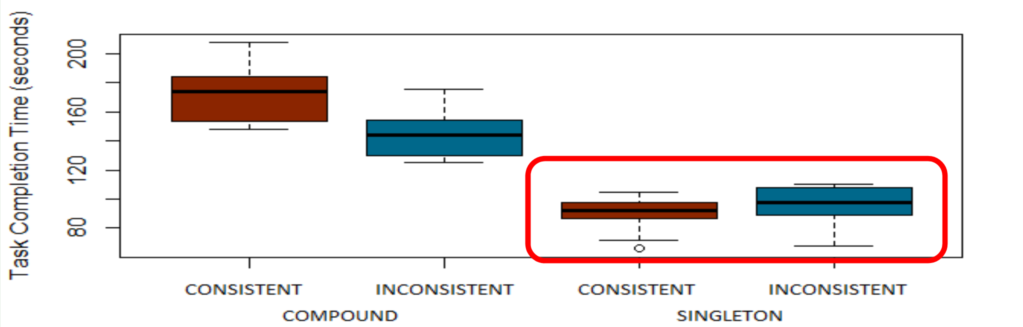
Perceptual Study

Results

Experiment 2: Task completion time distribution in a Search task



Experiment 3: Task completion time distribution in a Non-Search task

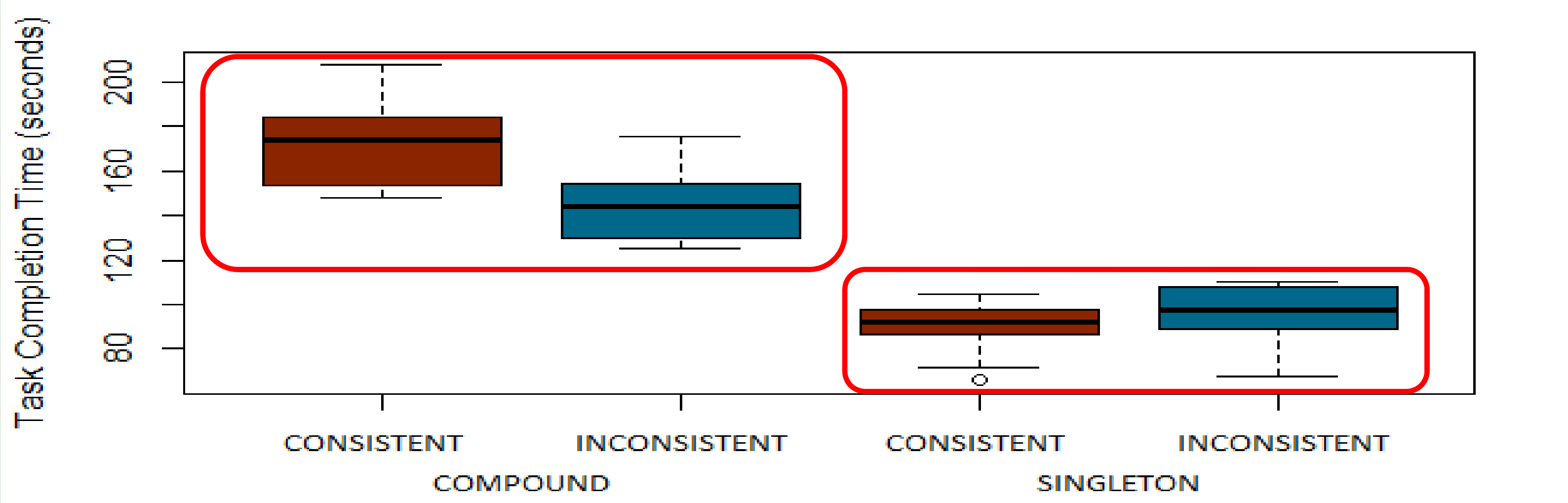


Object consistency decreases task completion time for Search tasks

Perceptual Study

Results

Experiment 3: Task completion time distribution in a Non-Search task



In a Non-Search task, consistency

- Increased task completion time for compound objects
- Decreased task completion time for singleton objects.

Perceptual Study

Weight Generation

Weight Generation via General Linear Modeling

- We subjected the completion times to a Multiple Linear Regression (MLR) analysis
- Generate a quantitative model of how the different experimental factors affect performance
  - Also indicates the relative importance of each factor

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

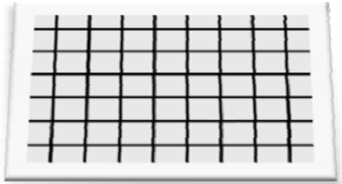
Table II The Regression Coefficients and Their Significance on the Overall Model, for the Case of a Nonsearch Task

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

HLSM: The High Level Saliency Model

GPU based implementation

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



$$\lambda_{semantic} = \frac{c-m}{c} \qquad \lambda_{physical} = \frac{1}{1-c} \times \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2}$$

Implementation - Stage 1

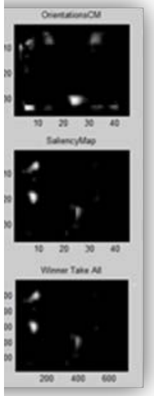
- Compares the scene schema of each object against the scene schemata of the objects surrounding it
- Estimates the number of neighbors for each examined object + available image depth information

HLSM: The High Level Saliency Model

GPU based implementation

Implementati

- Runs on a
- Highlights



High Level Saliency

HLSM: The High Level Saliency Model

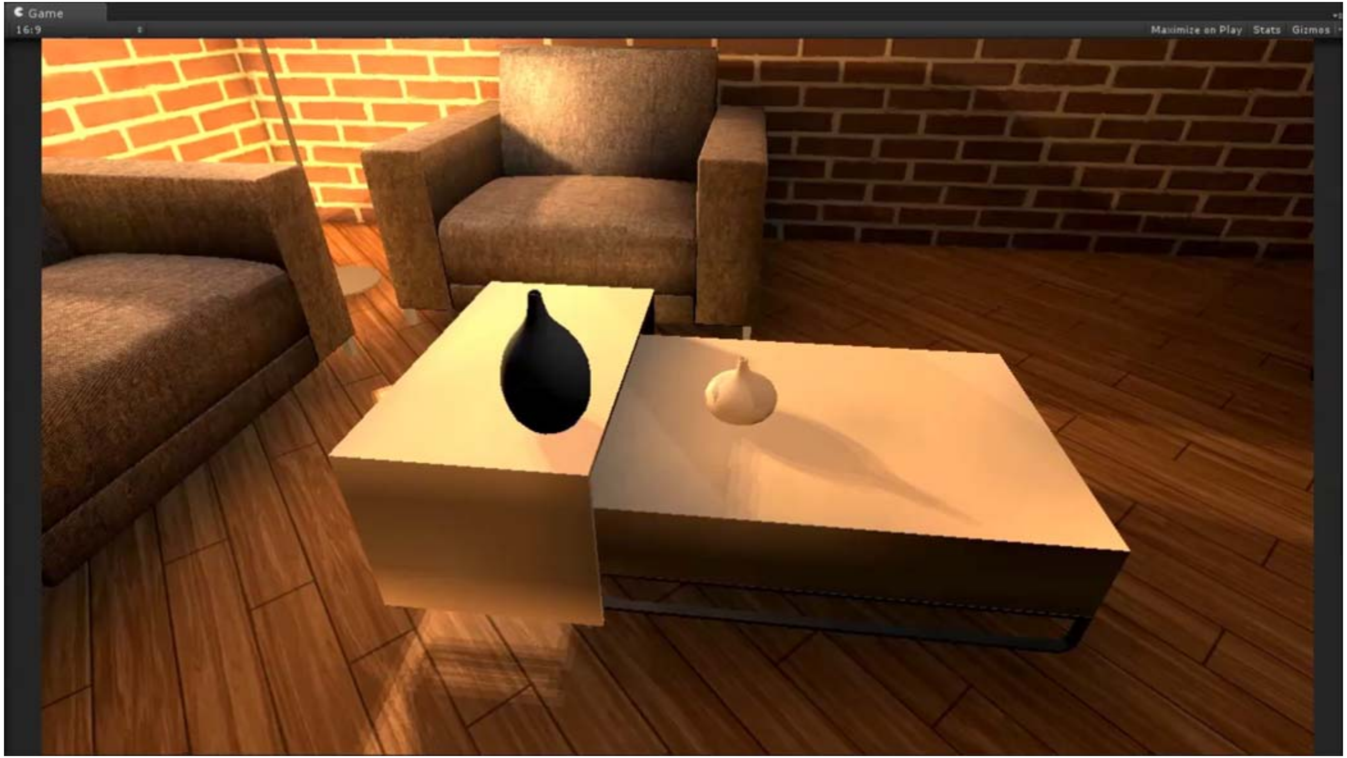
Game Balancing

Game Level Editing

- Looking for an object is a common task in (Action-)Adventure video games
- Plot-critical objects are placed in selected locations to ease or burden the player
- Maintain challenge via Game Balancing
- Aid game level designer to identify object saliency depending on location

HLSM: The High Level Saliency Model

Saliency Detection



VIDEO

Evaluation of Game Level Editing

Experiments 4 & 5: Evaluation of the implementation

Evaluation of the High Level Saliency Model

Evaluate the efficiency of our model in predicting attention deployment by

- Examining its effect on task completion time
- Acquiring eye-tracking data

Evaluation of Game Level Editing

Experiments 4 & 5: Evaluation of the implementation

Design of the Game Balancing evaluation experiment

- Game Levels were designed with the HLSM system
- Four game levels
  1. Easy/Search task
  2. Hard/Search task
  3. Easy/Non-Search task
  4. Hard/Non-Search task
- Three objects in quest for all levels
- Used the already classified objects of Experiment 1
- 40 participants, 10/condition

Evaluation of Game Level Editing

Footage

VIDEO

Evaluation of Game Level Editing

Completion time & Eye-tracking data Analysis

Completion time Analysis

- Both tasks completion times were reduced for the easy tasks vs the hard tasks
- Consistent with analysis of Experiments 2 & 3

Eye-tracking data Analysis

Clear influence of both context consistency and physical isolation in attention deployment

Evaluation of Game Level Editing

Eye-tracking Data Heat maps

Aggregated fixations over raw eye data from all participants and visual angles

Search task: *Spectacles* ***Non-Search task:*** *Items for a car trip*

Heat map analysis

- In a Search task eye gaze is directed to consistent object locations
- In a Non-Search task the eye scan pattern spans over the entire scene

Summary

Conclusion & Future Work

Conclusion

- Successful prediction of salient objects identified as non-salient in terms of low level features
- We validated our model's performance via eye-tracking
- Implicitly adjust game level difficulty by manipulating object topology

Summary

Conclusion & Future Work

Limitations

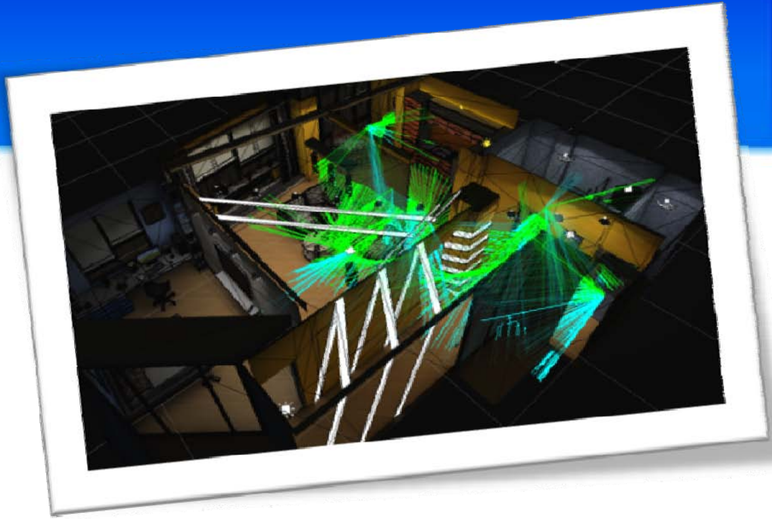
Only two contextual factors are integrated in this model; more can be found in the cognitive psychology literature

Publications

Koulieris, G.A., *Drettakis, G., Cunningham, D., Mania, K. (2014, September). An Automated High Level Saliency Predictor for Smart Game Balancing. ACM Transactions on Applied Perception (TAP) 11, 4, Article 17, 21 pages.*

Koulieris, G.A., *Drettakis, G., Cunningham, D., Mania, K. (2014, July). High level saliency prediction for smart game balancing. In ACM SIGGRAPH 2014 Talks, p. 73.*

Overview



Motivation

Problem Statement

Contributions

Background

Part 1: An Automated High-Level Saliency Predictor for Smart Game Balancing

Part 2: C-LOD: Context-aware Material Level-of-Detail applied to Mobile Graphics

Part 3: Gaze Prediction using Machine Learning for Dynamic Stereo Manipulation

Summary