Computational models for predicting gaze direction in interactive virtual environments

Robert J. Peters and Laurent Itti

University of Southern California, Computer Science, Neuroscience

introduction

We use ***fully-computational, autonomous*** models to predict eye movements in a ***dynamic and interactive*** visual task with ***naturalistic stimuli***.

We move beyond purely stimulus-driven bottom-up models, and introduce a simple model that captures task-dependent top-down influences on eye movements.

Previous studies have either relied on qualitative/descriptive models, or have not used naturalistic interactive stimuli.

conclusions

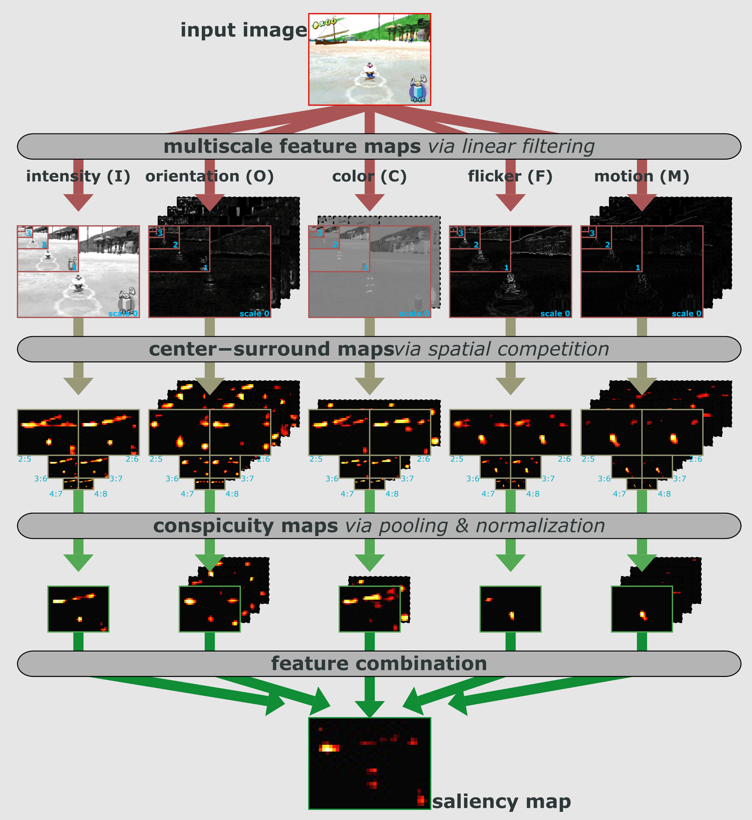
- Purely bottom-up models are able to predict eye position significantly better than chance, but not as well as in passive-viewing conditions
- Combining the bottom-up model with a simple model of top-down, task-dependent influences leads to significantly improved eye position prediction
- This kind of model could be used in autonomous machine vision situations, such as interactive virtual environments

videogame stimuli

- Nintendo GameCube games
- 24 sessions, 5 minutes each
- eye movements recorded during game play

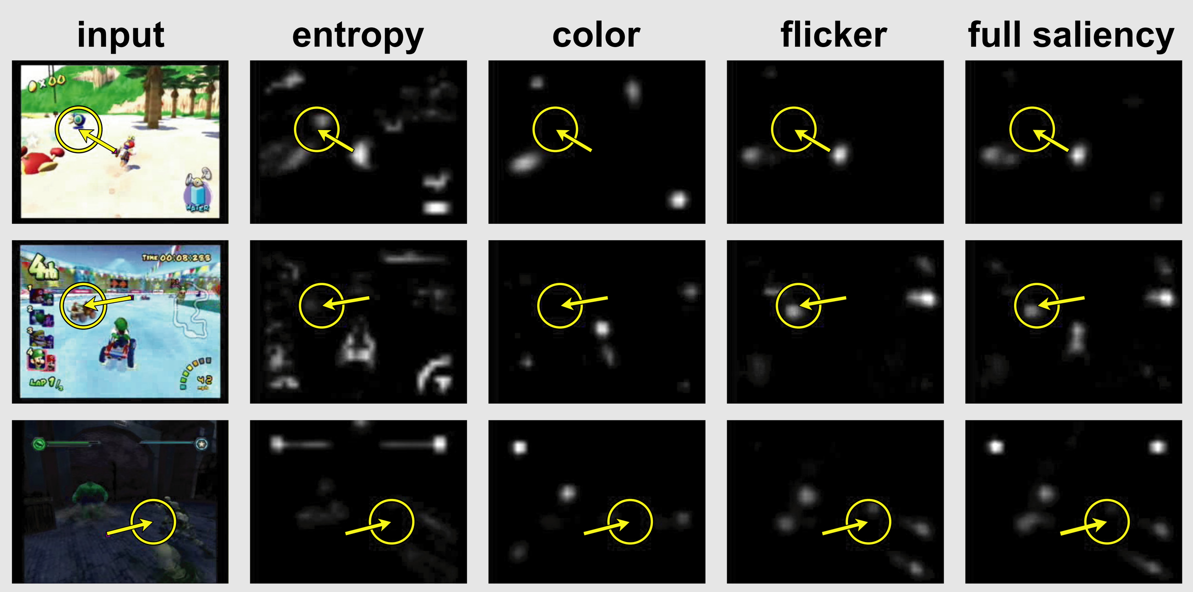
bottom-up model

- input processed through as many as 5 multi-scale feature channels
- different versions of the model include different combinations of features
- ***salience is a global property***: outliers detected through coarse global competition



bottom-up predictions

- sample frames illustrate the output of the models
- for illustration, these frames are selected at the time when a saccade is just beginning
- yellow arrows indicate the saccade path

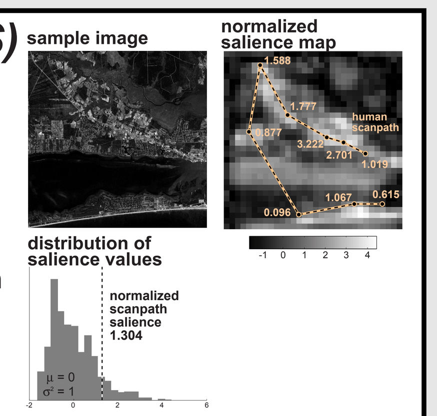


metrics: model scoring

- two different metrics used to test how well the models predict human eye movements

normalized scanpath salience (NSS)

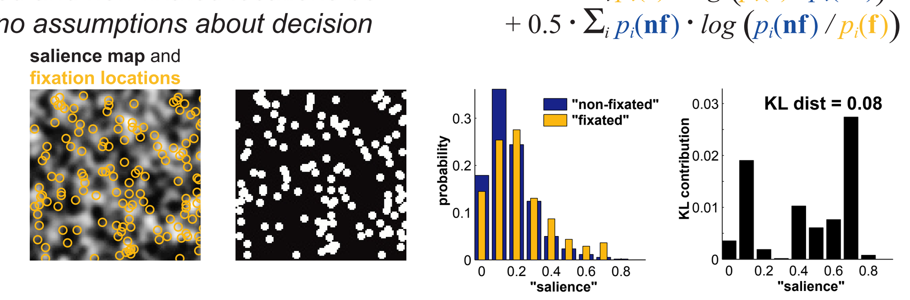
- each salience map (or control map) is normalized to have mean=0 and stdev=1
- human scanpath is overlaid on normalized salience map
- normalized salience value is extracted at each fixation location
- these values are summed to give the normalized scanpath salience (NSS)
- NSS can be compared with the distribution of random salience values



Kullback-Leibler (KL) distance

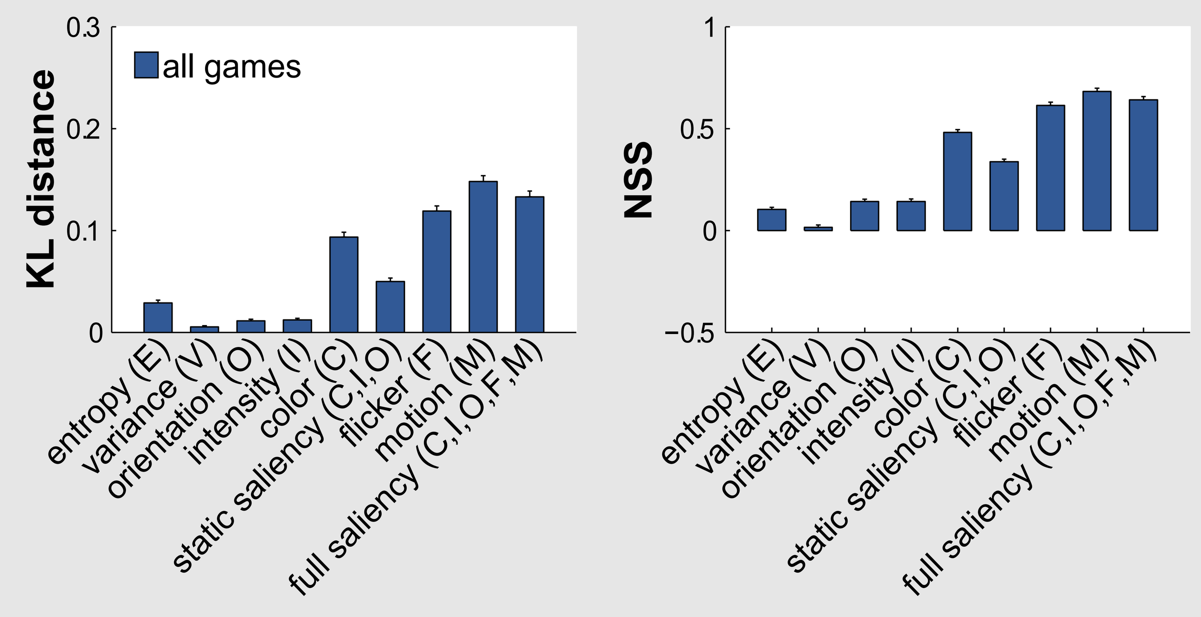
- how well can fixated and non-fixated locations be distinguished, *with no assumptions about decision criteria?*

$$KL = 0.5 \cdot \Sigma_i\, p_i(\mathbf{f}) \cdot \log\left(p_i(\mathbf{f}) / p_i(\mathbf{nf})\right) + 0.5 \cdot \Sigma_i\, p_i(\mathbf{nf}) \cdot \log\left(p_i(\mathbf{nf}) / p_i(\mathbf{f})\right)$$



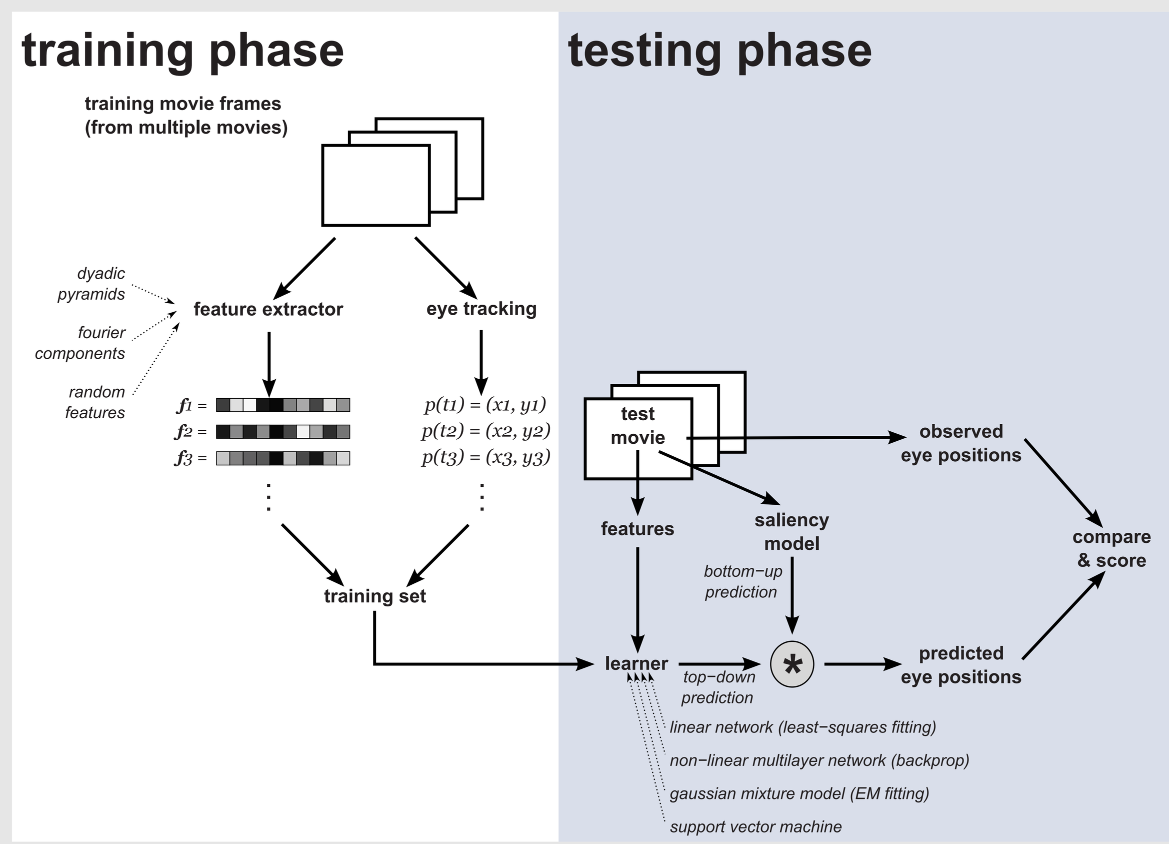
bottom-up results

- models score significantly above chance
- dynamic, global features work best
- but, scores in this *interactive task* are lower than in previous *passive-viewing* experiments



top-down model

- model learns to associate ***image "gist" signatures*** with corresponding ***eye position density maps***
- in testing, a leave-one-out approach is used: for each test clip, the remaining 23 clips are used for training
- therefore, the model must be able to ***generalize across game types*** in order to successfully predict eye positions

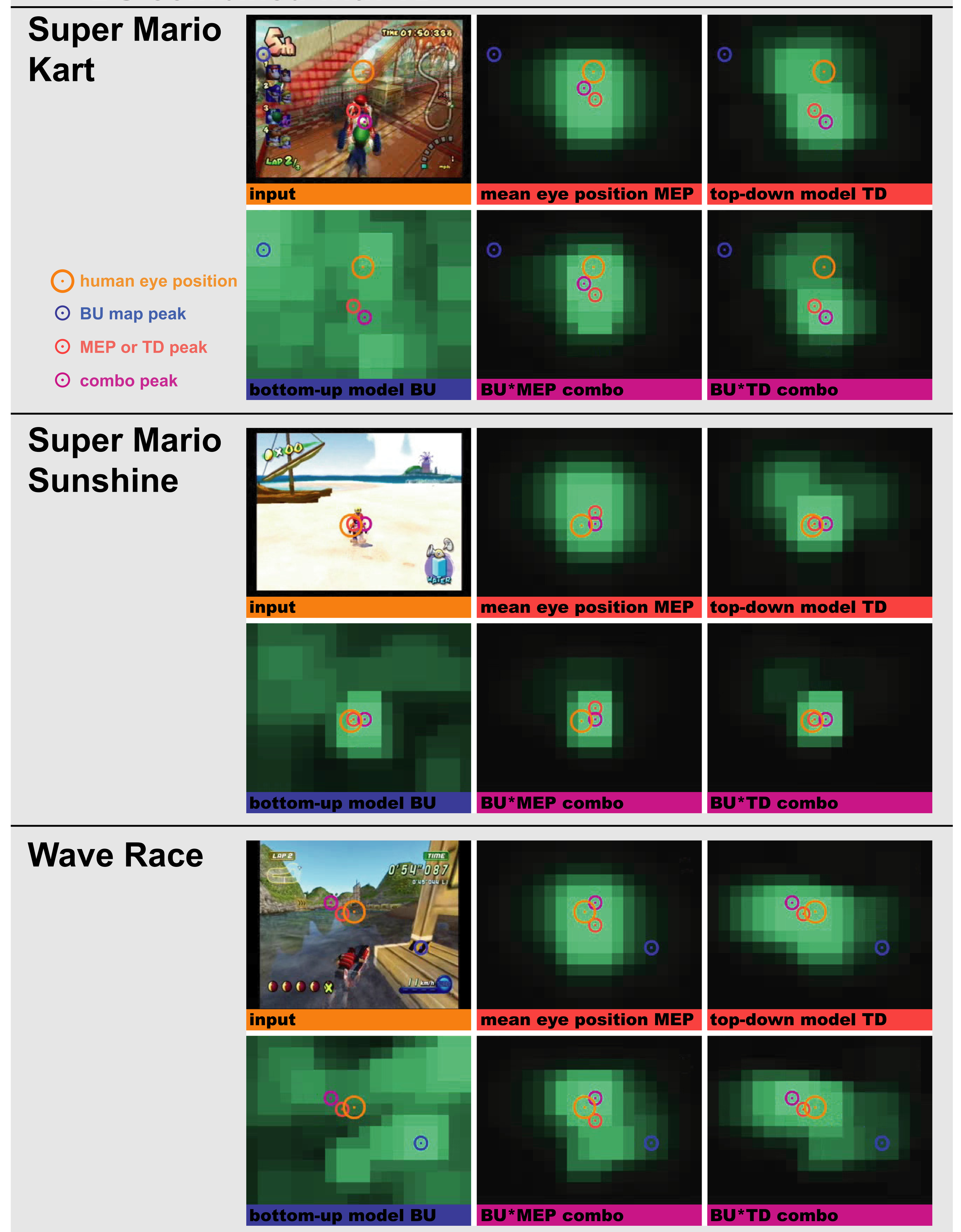


dyadic pyramid features

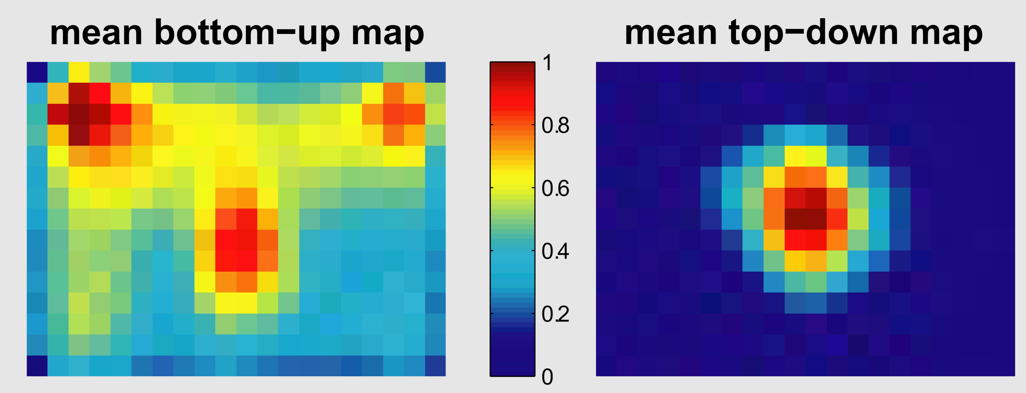
- 7 pyramids
  - 1 luminance
  - 2 color
  - 4 orientation
- 2 scales per pyramid
  - coarse
  - fine
- 32 features per scale
  - 4x4 array of local mean
  - 4x4 array of local var.

top-down predictions

- sample frames illustrating output of
  - **bottom-up (BU)** model alone
  - **mean eye position (MEP)** (a control condition)
  - **BU combined with MEP**
  - **top-down model (TD)** based on pyramid features
  - **BU combined with TD**

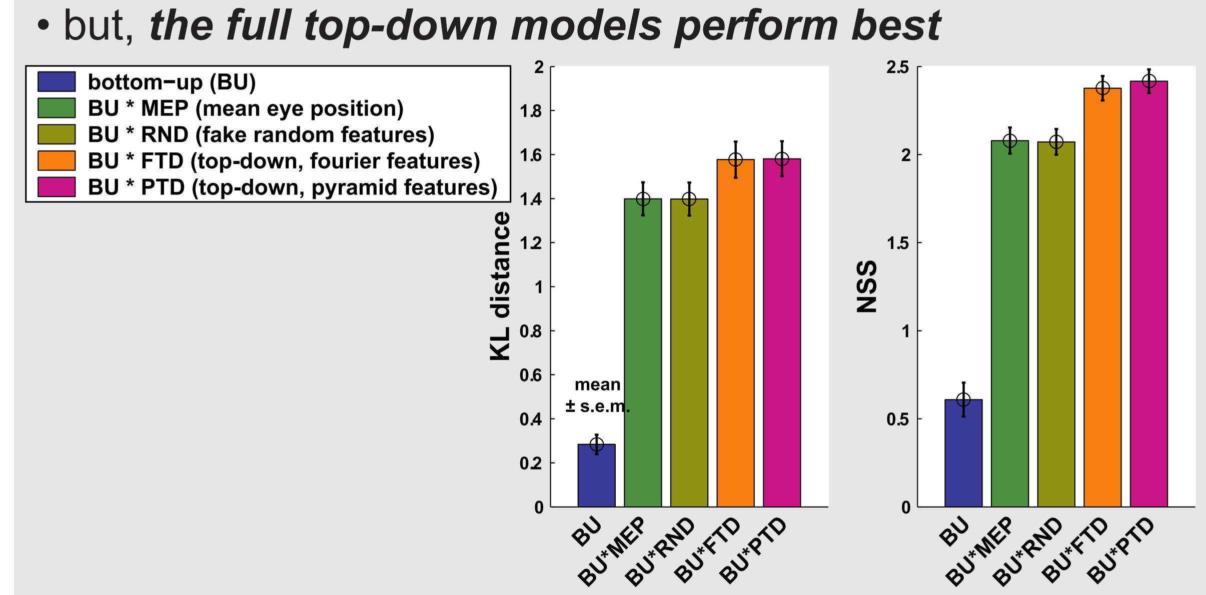


- average bottom-up and top-down maps across all frames
- bottom-up map reflects activity in the screen corners (game score, time counter, etc.) that is largely ignored by observers



top-down results

- a simple mean-eye-position control improves upon purely bottom-up performance by a factor of 3-4x
- but, ***the full top-down models perform best***



acknowledgements

Supported by the National Geospatial-Intelligence Agency (NGA) and the Intelligence Community (IC) Postdoctoral Research Program