Modelling User Interest using Gaze for Proactive Augmented-Reality Information Retrieval

Melih Kandemir

Helsinki Institute for Information Technology, Department of Information and Computer Science, Helsinki University of Technology, P.O. Box 5400, FI-02015 TKK, Finland
melihk@cis.hut.fi, http://www.cis.hut.fi/projects/mi/

HELSINKI UNIVERSITY OF TECHNOLOGY
Department of Information and Computer Science

Summary

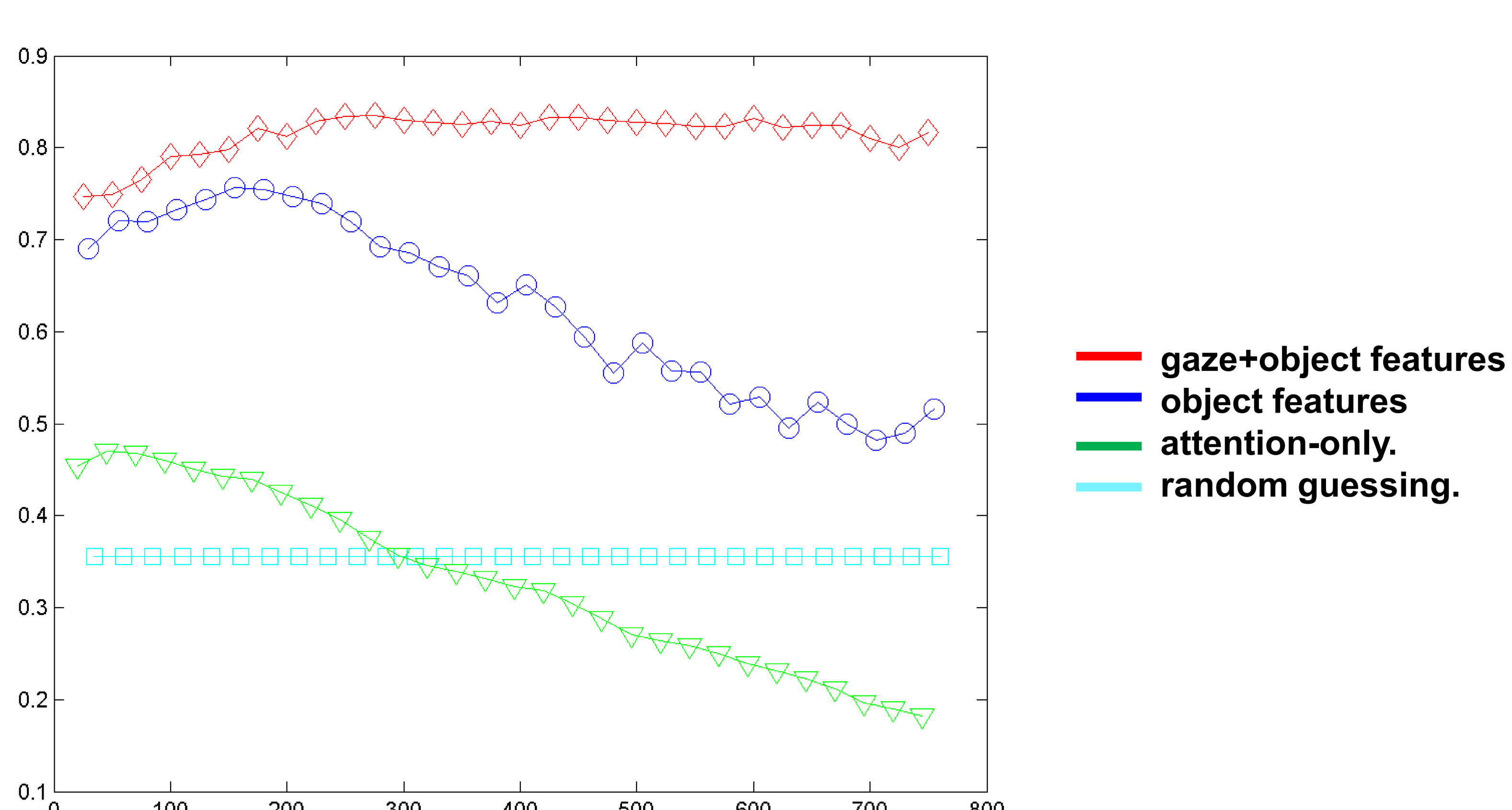
We build proactive user interfaces for ubiquitous augmented-reality information access devices. To this end, I develop statistical models that infer the relevances of real-world objects from gaze patterns in dynamic scenes.

Introduction

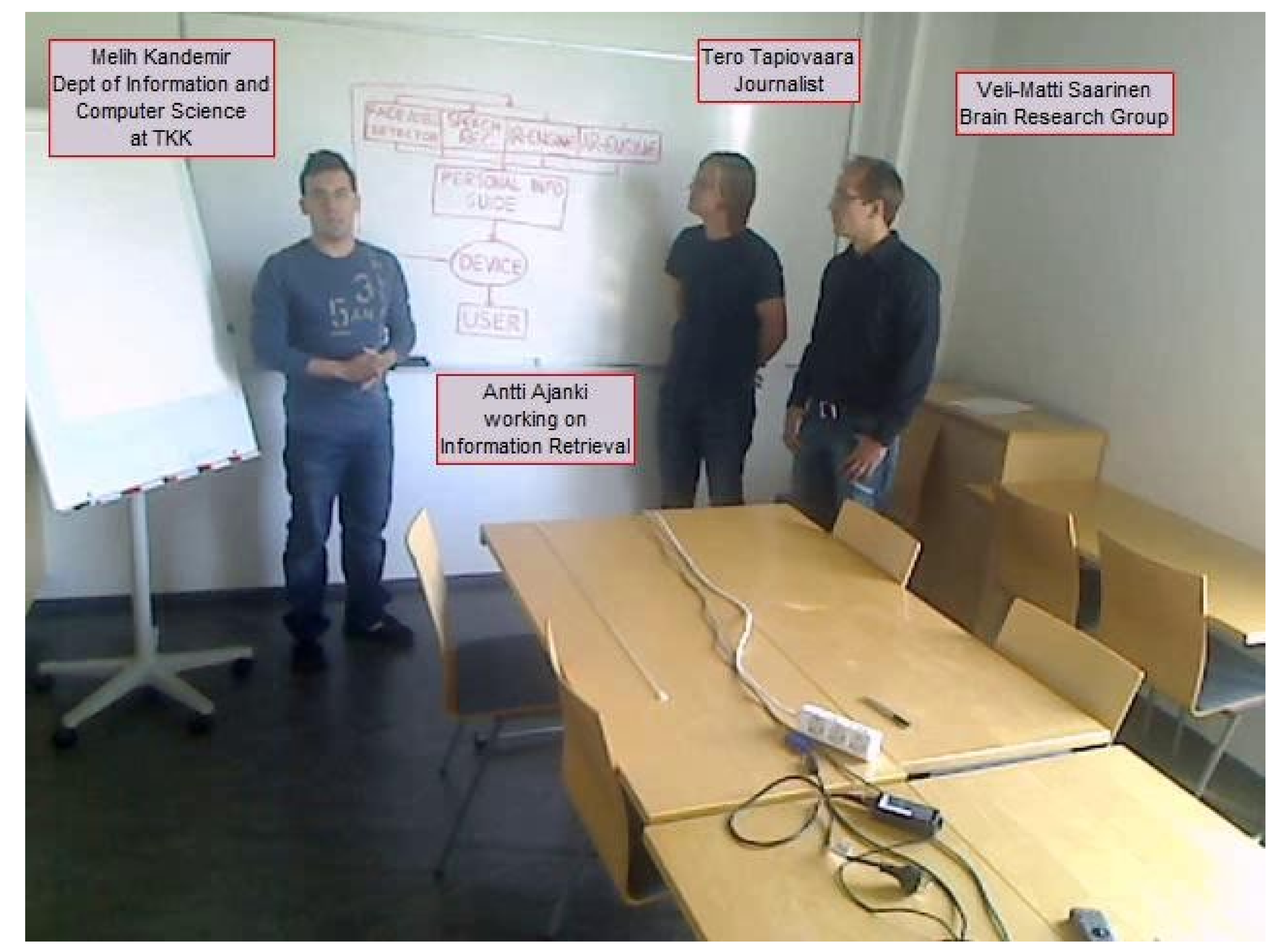
- As eye-glasses with data augmentation and gaze tracking capabilities are recently available, it is now possible to develop ubiquitous augmented-reality information access devices.
- Ubiquitous AR: real-world objects are augmented with additional information.
- A selection mechanism is required when there are multiple objects in the scene.
- Explicit solutions require full user attention and cause Midas touch effect: *each glance activates an action whether it is intended or not.*
- Imagine a real-world Google, where the scene objects are information sources and gaze patterns serve as implicit queries.
- I develop statistical models that provide an implicit mechanism for ranking the scene objects by inferring their relevances from the gaze patterns.

Inferring Object Relevances

- We tested the feasibility of inferring object relevances from gaze patterns [1].
- Gaze data are collected from subjects watching the video. Subjects ranked the objects by relevance in a subset of frames afterwards.
- We extracted gaze and object features of the recent past for the labelled subset of frames.
- *A sample gaze feature* : mean distance of fixations to the center of an object.
- *A sample object feature :* mean area of the bounding box of an object.
- A standard ordinal logistic regression model gives significantly better accuracy when gaze features are also used together with object features.
- Currently, we are working on a multi-task learning scheme, where learning separate relevance assignment models for objects are regarded as relevant subtasks.



Proportion of correct predictions as a function of window size.



A screenshot from a hypothetical ubiquitous augmented-reality information access device.

Model Generalisation

- In the offline setting, we will take a learning-to-learn approach [2]: the model will learn to infer the parameters of a previously unseen object using the parameters of seen objects.
- In the online setting, we will formulate the problem in re-inforcement learning [3] terms by defining:
  - a cost function that measures the success of the ranks given by the current model.
  - a list of actions for adjusting model parameters (e.g. stepwise increment/decrement)
  - a statistical model for inferring the ground truth feedback from gaze.

The ubiquitous augmented-reality device we are using.

Utilising psychophysics

- We will perform psychophysical experiments to discover the gaze and video features that are most correlated to object relevance.
- We will develop a model that infers the cognitive task of the user given the visual properties of her fixation locations.
- We will extend the previously developed relevance inference models to be dependent on the cognitive task.

References

[1] Meilh Kandemir, Veli-Matti Saarinen, Samuel Kaski, Inferring Object Relevance from Gaze in Dynamic Scenes, Submitted to ETRA 2010.

[2] Kai Puolamäki, Antti Ajanki, and Samuel Kaski, Learning to learn implicit queries from gaze patterns, ICML '08: Proceedings of the 25th international conference on Machine learning (New York, NY, USA), ACM, 2008, pp. 760–767.

[3] Aaron Wilson, Alan Fern, Soumya Ray, and Prasad Tadepalli, Multi-task reinforcement learning: a hierarchical bayesian approach, ICML '07: Proceedings of the 24th international conference on Machine learning (New York, NY, USA), ACM, 2007, pp. 1015–1022.