A machine learning framework for gaze guidance

Eleonora Vig, Michael Dorr, Karl R. Gegenfurtner*, and Erhardt Barth
Institute for Neuro- and Bioinformatics, University of Lübeck, Germany, http://inb.uni-luebeck.de
*Abteilung für Allgemeine Psychologie, Justus-Liebig-Universität Gießen, http://www.allpsych.uni-giessen.de
{vig|dorr|barth}@inb.uni-luebeck.de, Karl.R.Gegenfurtner@psychol.uni-giessen.de

Introduction

Motivation

What constitutes the difference between fixated and non-fixated movie patches? How can we change a patch to make it more or less salient?
We present a novel computational model of low-level saliency with dual emphasis: the same machine learning framework is used (i) for predicting saccade targets in natural dynamic scenes, and (ii) for learning how to alter the saliency level of these targets.

Gaze guidance

The objective of the GazeCom project is to improve visual communication by designing gaze-contingent interactive displays that change, in real time, the saliency distribution of the scene. To perform gaze guidance, we first predict a limited number of candidate locations that would attract the user's gaze, increase the saliency at the location that is selected for being attended and decrease saliency at all possible distractors.

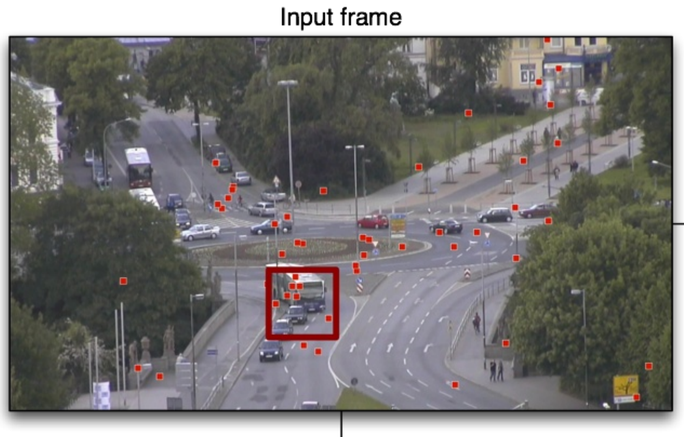
Experimental setup

A large dataset of ~40,000 saccades was obtained from 54 subjects free-viewing 18 high-resolution movie clips of outdoor scenes of ~20 sec durations each (29.97 fps, subtending 48x27 deg of visual angle). The saccade landing points were used to label image regions as attended. For the non-attended class, we shuffled the mo-vies and their scanpaths, thus eliminating the central fixation bias.

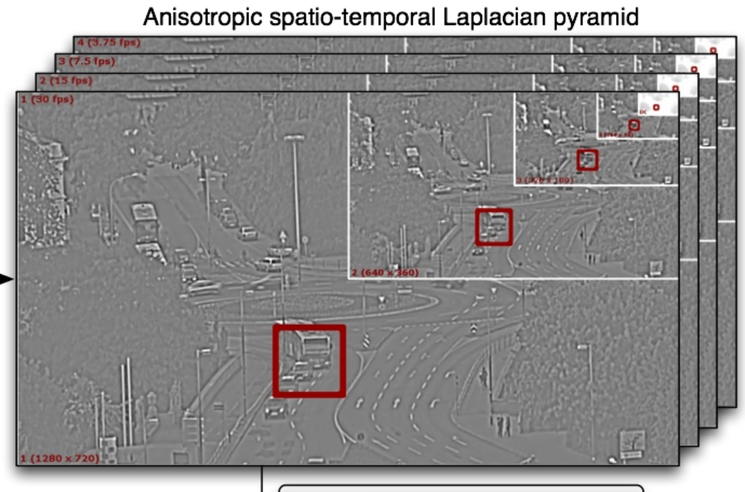
Saliency prediction and modification

Prediction phase

Input frame



Anisotropic spatio-temporal Laplacian pyramid



Eye-tracking

Attended locations
$$\begin{cases} p_1 = x_1, y_1, t_1 \\ p_2 = x_2, y_2, t_2 \\ \dots \end{cases}$$

Non-attended locations
$$\begin{cases} p'_1 = x'_1, y'_1, t'_1 \\ p'_2 = x'_2, y'_2, t'_2 \\ \dots \end{cases}$$

Spectral energy computation

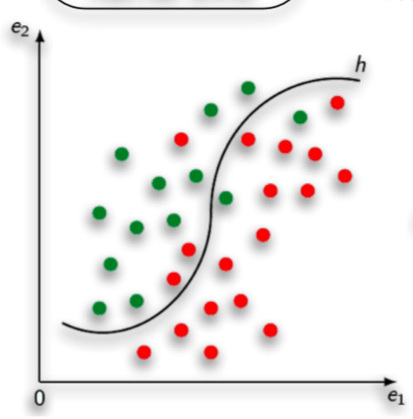
$$e_l = \sqrt{\frac{1}{W_l H_l} \sum_{i,j}^{W_l,H_l} I_l^2(i,j)}$$

Training data

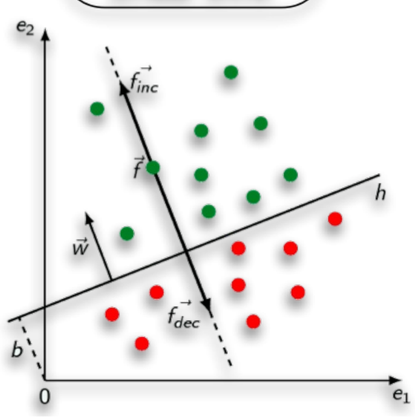
$$\vec{f_{p_1}} = \boxed{e_1\ e_2\ e_3\ e_4\ e_5\ \cdots\ e_{18}\ e_{19}\ e_{20}}$$
$$\vec{f_{p_2}} = \boxed{\cdots}$$
$$\dots$$
$$\vec{f_{p'_1}} = \boxed{\cdots}$$
$$\vec{f_{p'_2}} = \boxed{\cdots}$$
$$\dots$$

5 spatial, 4 temporal pyramid levels $I_0 \dots I_{20}$

1^{st} learner: Kernel SVM → discard the wrongly classified locations → 2^{nd} learner: Linear SVM → Transformation rules



Separating hyperplane
$h = (\vec{w}, b)$



For each candidate location r with spectral energy vector $\vec{f}$:

$$\vec{f_{inc}} = \vec{f} + \alpha \vec{w} \frac{b}{\|w\|}$$
$$\vec{f_{dec}} = \vec{f} - \alpha \vec{w} \frac{b}{\|w\|}$$

α - degree of change

Modification phase

Empirical saliency (density of saccade landing points)

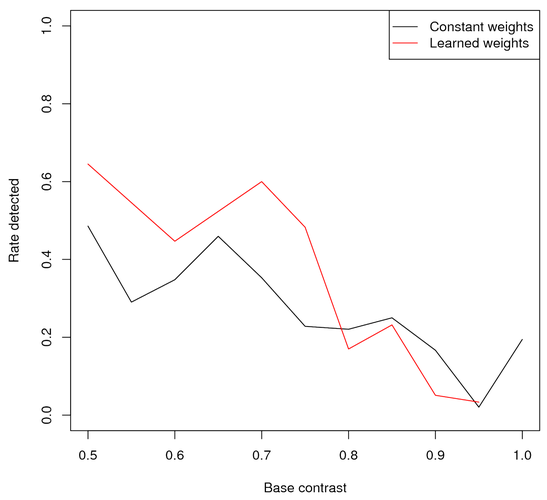
Thresholding

Pick 10 candidate locations, compute their spectral energy vector:

$$\begin{cases} r_1 = x_1, y_1, t_1 \rightarrow \vec{f_{r_1}} \\ r_2 = x_2, y_2, t_2 \rightarrow \vec{f_{r_2}} \\ \dots \\ r_9 = x_9, y_9, t_9 \rightarrow \vec{f_{r_9}} \\ r_{10} = x_{10}, y_{10}, t_{10} \rightarrow \vec{f_{r_{10}}} \end{cases}$$

To guide the observer's gaze to a specific location r, apply a transformation rule $\vec{f_{inc}}$ to increase its saliency. Apply $\vec{f_{dec}}$ to all distractors.

Results & Summary



Visibility of learned vs. constant energy modifications. We tested different global contrast levels for the underlying movie to allow for different dynamic ranges of the transformations.

- **The predictability of eye movements**
  - **is higher (AUC of 0.84) than previously reported results on both static and dynamic scenes.**
  - **is high although little information is used: only one feature (spectral energy) per movie patch and per scale.**
- **The same Machine Learning framework is used to derive transformations in the energy profiles that alter the saliency distribution of the scene.**
- **Preliminary results show that gaze-contingent energy modifications do indeed have a gaze guiding effect.**

References

[1] E. Vig, M. Dorr, and E. Barth. Efficient visual coding and the predictability of eye movements on natural movies. Spatial Vision, 2009. (in press).

[2] E. Barth, M. Dorr, M. Böhme, K. R. Gegenfurtner, and T. Martinetz. Guiding the mind's eye: improving communication and vision by external control of the scanpath. In: Human Vision and Electronic Imaging, volume 6057 of Proc. SPIE. B. E. Rogowitz, T. N. Pappas, and S. J. Daly (Eds.) (2006).

[3] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11):1254-1259 (1998).

[4] M. Dorr, and E. Barth. A gaze-contingent display for gaze guidance. Poster 103.

**Our research has received funding from the European Commission within the GazeCom project (IST-C-033816) of the FP6.
All views herein are those of the authors alone; the European Community is not liable for any use made of the information.**