Hooligan Detection:
The Effects of Spatial and Temporal Expert Knowledge

Viviane Zwanger, Dominik M. Endres, Martin A. Giese
Section for Computational Sensomotorics., Dept. of Cognitive Neurology, University Clinic, CIN, HIH, University of Tübingen.
viviane.Zwanger@cin.uni-tuebingen.de, dominik.endres@klinikum.uni-tuebingen.de, martin.giese@uni-tuebingen.de

Introduction

Surveillance of large crowds is an important task for ensuring the security of visitors at mass events, like soccer matches. We have shown [4] that after removal of global (spatial and temporal) context, the looking strategies of experts and naïve observers are quite similar when searching for security-relevant events. However, both detection and event classification rates of the experts are higher than those of the naïve observers. To test the influence of global context on fixation strategies under more natural conditions we investigated the eye movements of two experts and 20 naïve observers for characteristic phases taken from real soccer matches. The gaze patterns from both groups, watching for security-relevant events, were compared to saliency maps derived from low-level features [2]. We found significant differences between experts and naive participants, indicating that experts use specific scene-specific priors, which depend on the spatial layout of the arena and different temporal phases during the game. Furthermore, spatial expert priors became effective only after a few seconds of watching a scene, while initial looking behaviour was driven by low-level saliency.

Previous work [4]

The Tübingen hooligan simulator

The green circle (not visible in the experiment) indicates the position of the security-relevant event. **Ask me for demo!**

- **Goal**: quantification of differences in visual search for security-relevant events between experts and naïve observers.
- We staged relevant critical and non-critical events in a football-stadium-like setting with human actors.
- Controlled event statistics, and circumvention of legal problems associated with real footage.

Main results: when spatial and temporal contexts were removed, gaze patterns of experts and naïve observers did not differ significantly, but experts achieve higher detection/classification performance.

Acknowledgements

This work was supported by EU projects FP7-ICT-215866 SEARISE, FP7-249858-TP3 TANGO, FP7-ICT-248311 AMARSi, the DFG and the CIN Tübingen. We thank PHK G. Mainda and PHK H.-J. Berg from the Düsseldorf police for their expert knowledge and their patience during the experiments.

Objectives

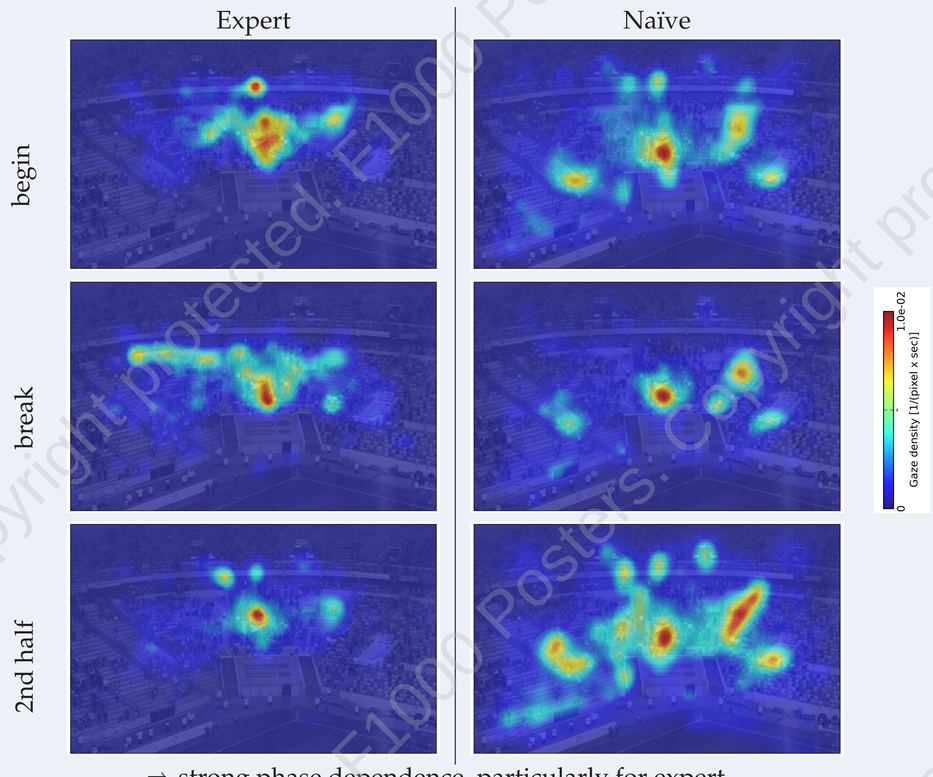
We tested the influences of

- **geometric contextual information**: entrances, fences, balustrade.
- **temporal context, 6 game phases**: begin, 1st half, break begin/end, 2nd half, exiting

on the search strategies of experts and naïve observers. Camera viewpoint and game phases were chosen following expert advice.

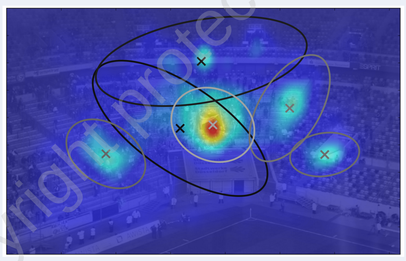
- grandstand of soccer stadium
- recordings from 3 games, 2 minutes per scene
- scenes contained mildly security-relevant event: throwing objects, climbing fence.
- **Ask me for demo!**

Gaze maps per game phase



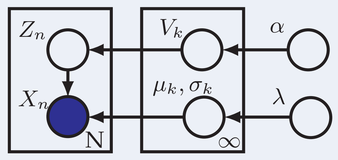
⇒ strong phase dependence, particularly for expert.

Region-of-interest evaluation



Gaze point clusters

Cluster membership map



Graphical model of the Dirichlet process mixture model [1]. Gaze points X_n are drawn out of clusters $Z_n \in \{0, \ldots, k, \ldots, \infty\}$ with probability $\propto V_k$. Cluster k has mean μ_k and covariance matrix σ_k. α controls the rate of cluster generation. **Advantage**: (semi)-automatic determination of the number of clusters through the *data*.

Method

- Cluster all gaze points with Dirichlet process mixture model [1] to determine location and number of clusters.
- Compute membership map by assigning each location to the maximally probably cluster. Count gaze points based on the membership map.
- 3-way ANOVA on gaze point count within each cluster.

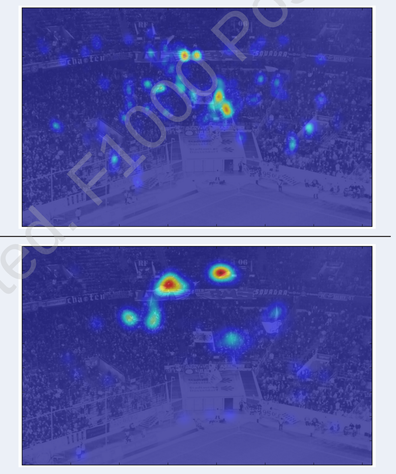
Main results: strong main effects for observer ($p < 0.035$) and game phase ($p < 0.001$ except for fence region: $p < 0.166$)
⇒ for computer vision surveillance system, integration of phase-dependent expert knowledge would be beneficial.

Comparison to AIM

Motivation

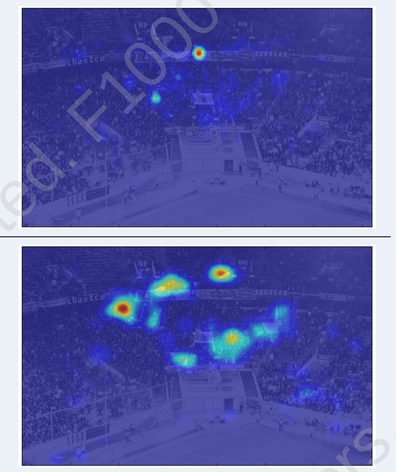
The EU-FP7 project *Smart Eyes: Attending and Recognizing Instances of Salient Events (SEARISE)* [3] aims at developing a system for the automatic detection of security-relevant events. It includes an attentional component derived from *Attention based on Information Maximization* (AIM) [2], which directs its gaze. We therefore compare the AIM saliency measure to expert gaze pattern for tuning and modulation of AIM.

First 10 seconds



⇒ in the first few seconds, both expert and AIM look at waving flags (or waving people). **Ask me for demo!**

120 second average



⇒ AIM keeps looking at waving flags, but expert concentrates on relevant locations, here: *balustrade* , which separates two inimical groups of fans.

⇒ We implemented multiplicative saliency modulation based on expert gaze maps in the SEARISE system.

- **Ask me for demo!**

References

[1] David M. Blei and Michael I. Jordan. Variational methods for the dirichlet process. In *In Proceedings of the 21st International Conference on Machine Learning*, 2004.

[2] N. Bruce and J. Tsotsos. Saliency, attention, and visual search: An information theoretic approach. *Journal of Vision*, 9(3):1–24, 2009.

[3] The SEARISE consortium. Smart eyes: Attending and recognizing instances of salient events. http://www.searise.eu, 2011.

[4] D. Endres, M. Höffken, F. Vintila, N. Bruce, J. Bouecke, P. Kornprobst, H. Neumann, and M. Giese. Hooligan detection: the effects of saliency and expert knowledge. In *Perception ECVP suppl.*, volume 39, page 193, 2010.