Tracking Viewer's Attention in Static Scenes with Augmented Saliency

Mohammad Mushfiqur Rahman, Siavash Gorji and James J. Clark

Electrical & Computer Engineering Department, CIM Lab, McGill University

Introduction

Our research objective is concerned with attention tracking - determining where, and to what people are paying attention while viewing static scenes. In particular, we are interested in predicting the visual attention of people looking at stationary images. Hence, the purpose of this research work is to *improve* the existing system that predicts the viewer's attention in static scenes.

Applications

- Object detection & recognition
- Visual surveillance
- Human-robot interaction
- Advertising
- Robotics

Fig. 1: Visual attention pattern used for Advertisement analysis [1]

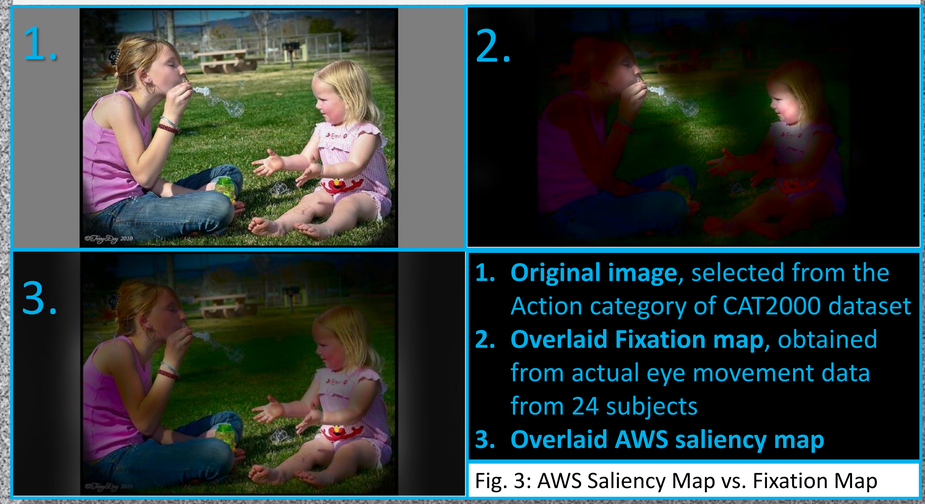
Current Technology

- The most common way of attention tracking is using the mature technology in gaze tracking.
- Some examples of devices that have been used to track eye movements are shown below in fig. 2.
- We are interested in predicting viewer's attention rather than tracking it directly.
- Attention models have commonly been validated against eye movements of human observers.

Fig. 2: Eye tracking devices [2]

Image Salience

- The current state-of-the-art of attention prediction techniques are based on computing **image salience maps**, which are descriptions of how likely regions in the image are to attract viewers' attention.
- According to a recent comparative study of the state-of-the-art in saliency models [3], the most efficient algorithm used to track the viewer's attention is the **AWS (Adaptive Whitening Saliency) Saliency Model** [4].
- The AWS saliency map gives us some good results, but still not close enough to the actual ones from eye-trackers.



1. **Original image**, selected from the Action category of CAT2000 dataset
2. **Overlaid Fixation map**, obtained from actual eye movement data from 24 subjects
3. **Overlaid AWS saliency map**

Fig. 3: AWS Saliency Map vs. Fixation Map

Augmented Saliency Model

- One area of improvement in existing models is to use high-level abstract information from the scene, e.g. detecting actors in the scene (face, body), what the actors are doing, estimating where and to which objects the actors are paying attention to.
- Here, we consider the actors' faces and their gaze direction as active manipulators of the viewer's attention.
- To achieve that we have used **Microsoft's Face API** within MATLAB environment. The Face API provides high precision face location detection that can detect up to 64 human faces in an image and extract other attributes such as face pose [5].
- With these information we could modify the saliency map to get our **Augmented Saliency Model** that would give us improved results.

GUI for Augmented Saliency

We have been developing a GUI to obtain the Augmented Saliency Map in MATLAB using Microsoft Face API.

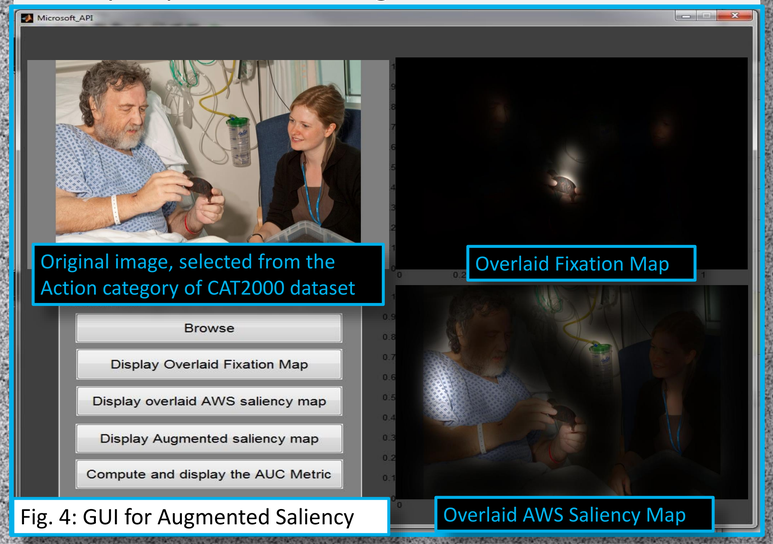


Fig. 4: GUI for Augmented Saliency

Conclusion

The GUI for Augmented Saliency Model is still under development. It is currently capable of performing the following tasks:

- ✓ Display Overlaid Fixation Map
- ✓ Display AWS Saliency Map
- ✓ Compare the AWS vs Fixation Map

Future Work

After static scenes we could move further by applying the same methodology to dynamic videos and movies.

References

1. Image from trends.e-strategyblog.com
2. Images from www.tobii.com
3. Borji, A., Sihite, D.N., & Itti, L. (2013). Quantitative Analysis of Human-Model Agreement in Visual Saliency Modeling: A Comparative Study. IEEE Trans. Image Processing, Vol. 22, No. 1.
4. Garcia-Diaz, A., Leboran, V., Fdez-Vidal, X. R., & Pardo, X. M. (2012). On the relationship between optical variability, visual saliency, and eye fixations: A computational approach. Journal of Vision, Vol. 12, No. 6, pp. 1–22.
5. http://www.projectoxford.ai/doc/face/overview