Learning to Predict Where Humans Look

Tilke Judd, Krista Ehinger, Frédo Durand, Antonio Torralba
{tjudd,kehinger,fredo,torralba}@mit.edu

Introduction

Teaser

Where do you look in these images?

This is where other people looked in eye tracking tests

Applications

Looking where people look has applications in graphics, advertising design and human-computer interaction. In addition to enabling intelligently automated cropping and Harvard's and driving level of detail for non-photorealistic rendering, a good model of saliency can be used in scene rendering and enhanced image and video compression.

Previous models of saliency

Most models of saliency are based on a computational model. Typically, multiple low-level visual features such as intensity, color, orientation, and texture are extracted from the image at multiple scales and combined in a linear or non-linear fashion into a master saliency map that represents the saliency of each pixel. The most salient objects tend to have high contrast or a specific color, orientation, or texture that is different from the surrounding area. The most salient object in an image is often what humans look at first. Some examples of low-level visual features include:

1. Color: High contrast or specific colors (e.g., red, blue)
2. Orientation: Straight lines or edges
3. Texture: Patterns or structures (e.g., grass, bricks)
4. Intensity: Bright or dark regions

Problems with previous models

Current saliency models do not accurately predict human fixations. Below, we describe several challenges and limitations:

1. Low accuracy: The current models are not very accurate, especially for images with low consistency or high entropy. The models tend to overestimate the saliency of regions that lie within the salient region of the image (or when measuring human performance in this paper).
2. High false positives: The models often predict fixations in regions that are not actually salient, resulting in a large number of false positives.
3. Insufficient data: The models are trained on a limited dataset, which may not be representative of the diversity of images that humans look at.

Our Contributions

1. Create a large public database of eye tracking experiments that show where people actually look in images.
2. Create a supervised learning model of saliency that correlates with human fixations on natural images.

Our Experimental Setup

Image database

We collected a large database of 1000 natural images. This dataset is used to train our model of saliency.

Eye tracking experiment

We ran a large eye tracking experiment with 15 users and 1000 images. This is the largest eye tracking database for saliency models that we know about and have made available to the public.

Analysis of database

Size of regions of interest

In most images, humans fixate on human faces. However, when viewing the clean-up of a scene, they fixate on specific objects rather than the entire scene, suggesting a more focused area of the region of interest. On the right is a histogram of the radius of the regions of interest in pixels.

Training the model

Learning a model of saliency

Features

We collect a set of features that are believed to be predictive of where people look.

Performance Results

This ROC curve compares the performance of models trained on different sets of features. The y-axis shows the percentage of human fixations that lie inside the area of an image predicted as salient by a model.

Acknowledgements

Funding comes from NSF CAREER awards 0447561 and IIS 0747120, Microsoft Research New Faculty Fellowship, Sloan Fellowship, Royal Dutch Shell, Quanta T-Family, MIT Singapore ISAMIT 14, Kavli graduate fellowship. Thanks to Adobe, China, Barbara Hodges, Nicolas Pinto, Yann LeTallec, Nando's team for support.