**Overview**

The objects that people choose to look at while viewing a scene provide an abundance of information about how a scene is ultimately understood.

**Eye fixation patterns**
- "where" the important things are in an image.

**Description**
- "what" in an image, which parts of an image are important to the viewer.

**Goal**
- Conduct combined behavioral and computer vision experiments to better understand the relationships between:
  - the objects that are detected in an image,
  - the eye movements that people make while viewing that image, and
  - the words that they produce when asked to describe it.

**Contribution**
- Comprehension of how humans view and interpret visual imagery.
- Demonstrate prototype applications for gaze-enabled detection and annotation by integrating gaze cues with the outputs of current visual recognition systems.

**Datasets**

**PASCAL VOC**
- 1,000 images, eye movements from 3 observers, 5 natural language descriptions per image from different observers.
- Task: free-viewing for 3 seconds, memory test
- 20 object categories
- Descriptions were collected by [1].

**SUN09 dataset**
- 104 images, eye movements from 8 observers, each of whom provided a scene description (additional descriptions were also obtained from other observers).
- Task: free-viewing for 5 seconds, with each viewing followed by scene description.
- 8 scene categories, 22 object categories.
- Descriptions were collected by [1].

**Example**

Fixed objects: bottle, person.
Described objects: bottle, person

---

**Experiments & Analyses**

**Do people look at objects?**
- 76.3% of the objects (for which we had detectors) were fixated in PASCAL, 65.6% for SUN09.

**What object categories did people tend to look at?**
- People are more likely to look at people, other animals, televisions, and vehicles.
- People are less likely to look at chairs, bottles, potted plants, drawers, and rugs.
- Animate objects are much more likely to be fixated than inanimate objects.

**What objects do people describe?**
- Relative to the ground truth number of objects in each image, 85.4% of the PASCAL and 58.7% of the SUN09 objects were described.
- Animate objects are much more likely to be described than inanimate objects.

**Applications**
- Combine gaze and automated object detection methods to create a collaborative system for detection and annotation.
- This required training SVM-based classifiers to discriminate true positive from false positive errors (using the deformable part model from [2], with default thresholds), then biasing these results with the gaze inputs.

**What is the relationship between gaze and description?**
- What percentage of the rectangular object detection window is part of the segmented object, and what percentage of fixations in that window were on the segmented object.

**What objects are detected in an image?**
- 91.5% of the PASCAL and 58.7% of the SUN09 objects were described.

**What objects do people describe?**
- 85.4% of the PASCAL and 58.7% of the SUN09 objects were described.

---

**Probability of being fixated when present for various object categories**

| Category | P (fixated | present) |
|----------|-------------|
| Animate  | 0.636       |
| Inanimate| 0.495       |

**Probability of being fixated when present for Animate versus Inanimate objects**

- **Animate**
  - 80% of the objects were fixated.
  - 59% of the objects were described.

- **Inanimate**
  - 49% of the objects were fixated.
  - 57% of the objects were described.

---

**Where do people fixate on these objects?**

- **All**
  - 68.4% of area
  - 69.0% of fixations
- **Person**
  - 52.7% of area
  - 58.8% of fixations
- **Chair**
  - 57.5% of area
  - 59.1% of fixations
- **Painting**
  - 91.1% of area
  - 91.5% of fixations

**Probability of being fixated when present for various object categories**

- **P (described | present)**
  - PASCAL: 0.866
  - SUN09: 0.737

**Probability of being described when present for various object categories**

- **P (described | fixated)**
  - PASCAL: 0.952
  - SUN09: 0.725

---

**Acknowledgements**

This work was supported in part by NSF Awards IIS-1161876, IIS-1054133, IIS-1111047, IIS-0959979 and the SUBSAMPLE Project of the DIGITEO Institute, France. We thank J. Maxfield and J. Weis for data preprocessing and useful discussions.

---

**Try our SBU Gaze-Detection-Description Dataset**

http://www.cs.stonybrook.edu/~jai/gaze.html