MOTIVATION

Where do we look while consuming video information?

- Normal viewing
- Many participants
- Various conditions
GAZE TRACKING

Commonly used solution – gaze tracking hardware

Pros
• Accurate
• Continuous

Cons
• Expensive setup
• Many participants?
• Controlled viewing
SELF-REPORTING GAZE LOCALIZATION

- Eliminate specialized hardware
- Only computer screen is utilized
- Gaze data for one frame
- Natural viewing
- Internet-based deployment
- Easily scalable
METHOD OVERVIEW

• Instruct the participant
• Compulsory tutorial
• Play the video for $t_v$ seconds
• Display the character chart for $t_c$ seconds
• Ask for best seen combination
METHOD DETAILS

- After the last video frame show character chart
  - Short display time prevents “explorations”
- User types the characters from memory
  - User notified on error
  - Errors are discarded from the dataset
- Estimate accuracy based on tutorial
TUTORIAL

- Used for
  - Practical explanation
  - Participant screening

- Participants have to complete this stage
  - Define the pass conditions
  - Reject the participants that fail
TUTORIAL STRUCTURE

- Same idea as video experiment
  - Animated symbol instead of video
  - Known ground truth
- Flexibility in success definition
Welcome to the video watching experiment.

Please sit about 1 meter (3 feet) away from the screen. Maximize the browser window, bring it to the front of all other windows, and scroll down so that you can see the entire box below these instructions.

We will first give you a short tutorial: When you click "Ready", you will see a small color "X" moving around the screen, and after a few seconds it will disappear and be replaced very briefly with scattered random letters and numbers, such as "A10" or "G94". Follow the box with your eyes and when you see the scattered letters, remember which combination you see most clearly at the point where the box disappeared. When prompted, type the letter and the two numbers you saw best in the box that appears below.

You must succeed twice in order to continue to the videos.
Thank you for your participation.
EXPERIMENTAL VALIDATION
THE DATA

• DIEM (Dynamic Images and Eye Movements) dataset
  • 84 HD videos of different length
  • Various styles (trailers, sport, ads, etc.)
  • Mostly professionally created
  • Gaze-tracked
• Pick 12 frames
• About 50 recordings per frame

http://thediemproject.wordpress.com/
MASSIVE DATA COLLECTION

- Designed to run in browser
  - Based on widely supported Adobe Flash
  - Deployed in the internet
- Requires a minimum 1024x768 screen
- Participants from Amazon mechanical turk
  - 15 cents per session of 6-10 videos
  - About $2.50 per frame of interest
  - One tutorial session
  - User can repeat the session with other videos
TUTORIAL IMPORTANCE

- The tutorial is compulsory
  - Performed till 2 successes
  - Reject after 10 tries
- Increase results quality
  - 95% of the videos
HOW TO COMPARE

Our data

DIEM data

Video watching experiment

Kernel density estimation (KDE)

Chi-square distance
KERNEL DENSITY ESTIMATION

- Non-parametric method
- Considers the data noise
- Places a kernel at every point
  - We use Gaussians
- The result is probability at each pixel
OVERALL COMPARISON

DIEM – gaze tracking

Our data

Uniform distribution

\( \chi^2 \) distance

Experiment #
VISUAL COMPARISON
CHARACTER DENSITY

- From very dense to very sparse
- 150 samples per density value
- Find acceptable approve radius
CHART DURATION

- From 0.1 sec to 1.5 sec
- About 100 sample per duration value
- Find acceptable approve radius
CONCLUSIONS

- Turkers can do a good job if “trained”
- Our method achieves results close enough to gaze tracking

<table>
<thead>
<tr>
<th></th>
<th>Our method</th>
<th>Gaze tracking</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dense sampling in time</td>
<td>Expensive</td>
<td>Built-in</td>
</tr>
<tr>
<td>Method changes gaze</td>
<td>Possibly</td>
<td>Possibly</td>
</tr>
<tr>
<td>Special hardware</td>
<td>None</td>
<td>Required</td>
</tr>
<tr>
<td>Many participants</td>
<td>Easy</td>
<td>Harder</td>
</tr>
<tr>
<td>New frame from new video</td>
<td>Cheap</td>
<td>Costly</td>
</tr>
</tbody>
</table>
LIMITATIONS

• Spatial frame coverage (Poisson-disk)
• Accuracy
FUTURE WORK

• Different planes of analysis
  • Different cultures / countries
  • Different viewing conditions
  • Different screen sizes
• User interface optimization
TAKE HOME

- Character chart is better than pointing
- Per session tutorial vs. qualification
Thank you for your attention