`Imaging System for the Measurement of Body and Head Motion and Gaze Visualization for a Video-Based Eye Tracker`

Constantin Rothkopf
Advisor: Dr. Jeff Pelz
Overview

• A very brief review of eye movements
• The main idea: measuring head movements
• The device:
  – mechanical setup
  – calibration
  – data acquisition
  – data processing
• Application of the device: Classification of eye movements
• Results
The human image detector has a highly anisotropic design: high visual acuity is only available in a small central region of the retina, the fovea.
Stare at the ‘x’
Do not move your eyes!
Eye movements

• The solution to the acuity problem: eye movements
  – the eyes are moved to 'point to' objects or regions in the scene that require high acuity
  – eye movements are also made toward task-relevant targets even when high spatial resolution is not required

• Humans typically carry out more than 100,000 eye movements a day, which reach velocities of up to 600 degrees per second
Types of eye movements

Fixations
- head still, eyes fixating

Saccades
- head could be moving, eyes moving to new target

Smooth pursuits
- head still, eyes following target

VORs
- head moving, eyes pointing on target
Eye tracker

Dual Purkinje Tracker

Limbus Tracker
The RIT-Wearable-Eye-Tracker 1
The RIT-Wearable-Eye-Tracker 2

- Monochrome CMOS eye camera
- Calibration laser
- IR illuminator/optics module
- Folding mirror
- Hot mirror
- Color CMOS scene camera
Video based eye tracker

- An infrared diode illuminates the eye
- The retina reflects the infrared light very well
- The eye camera of the eye tracker records the position of the pupil together with the specular reflection
- Using image processing the position of the eye is calculated
Impact of eye movement research

• Understanding human perception:
  – Where and when do humans focus visual attention?
  – How does the brain process visual information?
  – How is the visual scene represented in the brain?

• Biologically inspired data processing:
  – Robot vision
  – Avatar vision, Animat vision

• Engineering applications:
  – Gaze position dependent information display
  – ‘Smart media’ systems, augmented reality

• Marketing
change blindness example by Jason Babcock
Most of the classification is carried out by trained experts who compare the captured video sequences from the eye camera and the scene camera of the eye tracker.

During the last 30 years a number of algorithmic approaches for the classification have been developed.

These algorithms were developed under the assumption that the head is fixed.

No algorithms were developed for the classification of smooth pursuits and VORs.
The idea

- In natural tasks the subject interacts purposefully with a changing environment: How do we use vision?
- A lightweight video-based eye tracker allows the subject to move their head and body in an almost natural way
- By measuring the head movements it is possible to:
  - distinguish smooth pursuits and VORs
  - measure the gaze direction relative to the environment
  - study head and eye coordination
• Is it possible to extract the head motion from the image sequence from the scene camera of the eye tracker?

• Three commonly used methods for ego-motion estimation:
  – epipolar geometry
  – optical flow
  – global estimation of transformation (automated image registration)

• The global estimation method was implemented using an image sequence from the scene camera
The problem: rotation and translation can result in ambiguous patterns because the point of expansion may be located outside of the field of view.

- Low accuracy
• Spherical projection of environment reduce ambiguity of image transformation under translation and rotation significantly
• Both the point of expansion and the point of contraction are visible on the sphere
Omnidirectional vision sensors

- Baker and Nayar (1999) have shown that an omnidirectional vision sensor with a single viewpoint can be constructed using a hyperbolic mirror and a perspective camera or with a parabolic mirror and an orthographic camera.

- If the system has a single viewpoint, it is possible to map the captured omnidirectional image to a planar perspective image.

- System used for this project:
  - hyperbolic mirror from Accowle Ltd., $a=8.37\text{mm}$, $b=12.25\text{mm}$, $r_{\text{top}}=26\text{mm}$
  - standard NTSC miniature video camera with 360 lines resolution.
Mechanical setup of the device

ASL 501 eye tracker

ISCAN eye tracker
Camera calibration

- The camera was calibrated using the Matlab calibration toolbox
- Pinhole-camera model:
  \[
  q = \begin{bmatrix}
  f & k_u & 0 & q_{u0} \\
  0 & f & k_v & q_{v0} \\
  0 & 0 & 1 & \end{bmatrix}
  \begin{bmatrix}
  x/z \\
  y/z \\
  1 \\
  \end{bmatrix}
  = Ku
  \]

  \(q\): pixel coordinates, \(u\): normalized image coordinates

- Camera model including radial and tangential distortion:
Geometry of catadioptric system

- Hyperbolic mirror:
  \[
  \frac{(z+e)^2}{a^2} - \frac{x^2+y^2}{b^2} = 1, \quad e = \sqrt{a^2+b^2}
  \]

- Intersection between ray from world point \( X \) through focal point and mirror:
  \[
  f(\mathbf{v}) = \frac{b^2(e\mathbf{v}_3 + a\|\mathbf{v}\|)}{b^2\mathbf{v}_3^2 - a^2\mathbf{v}_1^2 - a^2\mathbf{v}_2^2}
  \]

- Tracing ray back from image pixel \( q \) to the mirror:
  \[
  X_h = f(K^{-1}q) \cdot K^{-1}q + t_c
  \]
  with \( K \), the camera calibration matrix and \( t_c=2e \)
After camera calibration the catadioptric system has to be calibrated.

The diameter of the mirror top rim as imaged by the calibrated camera can be expressed as:

\[
r_{\text{pix}} = \frac{r_{\text{top}}}{z_{\text{top}} + 2e}
\]
Remapping the image

Image captured by catadioptric system

remapped to $\theta,\phi$ space, 512x512
Rotation estimation 1

- Three approaches:
  - Epipolar geometry (Svoboda, Pajdla, Hlavac 1998)
  - Optical flow (Gluckman, Nayar 1998)
  - Spherical harmonics (Makadia & Daniilidis 2003)

- When using epipolar geometry a feature tracker is needed to track corresponding points between images

- Optical flow methods have been used with higher resolution cameras and are reported to be very susceptible to noise, including occlusion

- The method using spherical harmonics decomposition has been shown to work well because of its global character
Rotation estimation 2

- Spherical harmonics basis-functions on the sphere:

\[ Y^m_l(\theta, \phi) = \sqrt{\frac{(2l+1)(l-m)!}{4(l+m)!}} P^m_l(\cos \theta) e^{im\phi} \]

- with the Associated Legendre Functions:

\[ P^m_l(x) = (-1)^m (1 - x^2)^{\frac{m}{2}} \frac{d^m}{dx^m} P_l(x) \]

- and the Legendre polynomials:

\[ P_l(x) = \frac{1}{2^l l!} \frac{d^l}{dx^l} [x^2 - 1]^l \]

- The spherical harmonics build an orthonormal set of basis functions on the sphere

Real part of spherical harmonics:
- vertically: \(0 \leq l \leq 4\)
- horizontally: \(-l \leq m \leq +l\)
Rotation estimation 3

• The discrete spherical harmonics coefficients:

\[ \hat{f}(l,m) = \frac{\sqrt{\frac{B}{2B-1}}}{\sum_{j=0}^{2B-1} \sum_{i=0}^{2B-1}} a_j f(\bar{e}_j, \bar{o}_i) Y_{ij}^m(\bar{e}_j, \bar{o}_i) \]

• A rotation of a function on the sphere can be parameterized with ZYZ Euler angles:

\[ g(\bar{\alpha}, \bar{\beta}, \bar{\gamma}) = R_z(\bar{\alpha})R_y(\bar{\beta})R_z(\bar{\gamma}) \]

• Under this rotation the new coefficients can be expressed as:

\[ \hat{f}_m = \sum_{p=-l}^{+l} \hat{f}_p U_{pm}(g), \quad \text{with} \quad U_{pm}(g(\bar{\alpha}, \bar{\beta}, \bar{\gamma})) = e^{-ip\bar{\alpha}} P_{pm}(\cos(\bar{\alpha})) e^{-im\bar{\gamma}} \]

where the \( P_{lmn} \) are generalized associated Legendre functions.
Rotation estimation 4

- The common trick: the rotation can be reparametrized as:

\[
g(\hat{\alpha}, \hat{\alpha}, \hat{\alpha}) = g_1(\hat{\alpha} + \frac{\pi}{2}, \frac{\pi}{2}, 0)g_2(\hat{\alpha} + \frac{\pi}{2}, \frac{\pi}{2}, \hat{\alpha} + \frac{\pi}{2})
\]

\[
\hat{f}_m = \sum_{p=-l}^{+l} \hat{f}_p \sum_{k=-l}^{l} U^l_{pk}(g_1(\hat{\alpha} + \frac{\pi}{2}, \frac{\pi}{2}, 0)) U^l_{km}(g_2(\hat{\alpha} + \frac{\pi}{2}, \frac{\pi}{2}, \hat{\alpha} + \frac{\pi}{2}))
\]

- The advantage is that the unknown variables appear only in the exponents:

\[
\hat{f}_m = e^{-im(\hat{\alpha} + \frac{\pi}{2})} \sum_{p=-l}^{+l} e^{-ip(\hat{\alpha} + \frac{\pi}{2})} f_p \sum_{k=-l}^{l} P^l_{pk}(0) P^l_{km}(0) e^{-ik(\hat{\alpha} + \frac{\pi}{2})} f_p
\]

- The problem to solve is then:

\[
\sum_{p=1}^{l} \sum_{m=0}^{l} \left[ e^{-im(\hat{\alpha} + \frac{\pi}{2})} \sum_{p=-l}^{+l} e^{-ip(\hat{\alpha} + \frac{\pi}{2})} f_p(t_1) \sum_{k=-l}^{l} P^l_{pk}(0) P^l_{km}(0) e^{-ik(\hat{\alpha} + \frac{\pi}{2})} f_p(t_1) - f_p(t_2) \right]^2 = 0
\]
• This equation can be minimized using Quasi-Newton Methods like Broyden’s Method

• Newton’s Method:
  \[ x_{n+1} = x_n - F'(x_n)^{-1}F(x_n) \]
  with \( x_{n+1} \): next iterate, \( x_n \): previous iterate, \( F'(x) \): the Jacobian of the function \( F(x) \)

• Broyden’s method uses a different update rule:
  \[ x_{n+1} = x_n - \bar{e}_n B_n^{-1}F(x_n), \quad \text{with} \quad B_n(x_n - x_{n-1}) = F(x_n) - F(x_{n-1}) \]

• The Matlab library implementing Broyden’s method by C.T.Kelley was used
• Three Matlab functions:
  – image sequence is remapped onto the unit sphere and stored
  – spherical harmonic coefficient are calculated from remapped image sequence and stored in a text file
  – rotation estimation from stored spherical harmonics coefficients

• The processing time is reduced by precalculating the mapping transformation and the spherical harmonics
Evaluation of rotation estimation 1

- The performance of the algorithm was first assessed with synthetic images.
- An image was warped onto the unit sphere and rotated with \( g(\alpha, \beta, \gamma) \).
- The algorithm based on the spherical harmonics decomposition was used to estimate the rotations.
## Evaluation of rotation estimation 2

**Image size: 1024x1024:**

<table>
<thead>
<tr>
<th>angle</th>
<th>( l \leq 5 )</th>
<th>( l \leq 8 )</th>
<th>( l \leq 12 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha = 10^\circ )</td>
<td>9.87°</td>
<td>9.91°</td>
<td>9.61°</td>
</tr>
<tr>
<td>( \beta = 10^\circ )</td>
<td>10.08°</td>
<td>10.12°</td>
<td>10.53°</td>
</tr>
<tr>
<td>( \gamma = 10^\circ )</td>
<td>9.92°</td>
<td>9.77°</td>
<td>9.93°</td>
</tr>
</tbody>
</table>

**Image size: 512x512:**

<table>
<thead>
<tr>
<th>angle</th>
<th>( l \leq 5 )</th>
<th>( l \leq 8 )</th>
<th>( l \leq 12 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \gamma = 5^\circ )</td>
<td>xx°</td>
<td>xx°</td>
<td>xx°</td>
</tr>
<tr>
<td>( \gamma = 10^\circ )</td>
<td>xx°</td>
<td>xx°</td>
<td>xx°</td>
</tr>
<tr>
<td>( \gamma = 25^\circ )</td>
<td>25.15°</td>
<td>xx°</td>
<td>xx°</td>
</tr>
<tr>
<td>( \gamma = 65^\circ )</td>
<td>64.47°</td>
<td>xx°</td>
<td>xx°</td>
</tr>
</tbody>
</table>
Combining the data
Several algorithms described in the literature based on three ideas:
- eye movement velocity criterion (thresholding)
- spatial clustering (for fixed head position)
- statistical modeling (Hidden Markov Model)

Most algorithms were developed assuming fixed head positions for use in tasks like reading, equation solving, signal detection, and image quality evaluation.

No algorithms have been developed for the classifications of eye movements including fixations, saccades, smooth pursuits, and VORs.
- Discrete time: $t, t+1, t+2, \ldots$
- State sequence: $Q = \{ q_t, q_{t+1}, q_{t+2}, \ldots \}$
- Observation sequence: $O = \{ O_t, O_{t+1}, O_{t+2}, \ldots \}$
- Transition probabilities: $a_{ij} = P( q_{t+1} = S_j | q_t = S_i )$
- Probabilities of observation variable: $P_j( O_t ) \sim \text{Normal}( \mu_j, \sigma_j^2 )$
Salvucci (1999) used a 2-state HMM for the classification of fixations and saccades.

- The observation variable is the eye-in-head velocity.
- The distributions of these velocities are modeled with Gaussian distributions.
- The parameters of the transition probabilities $a_{ij}$ and the parameters of the velocity distribution $\mu_j$, $\sigma_j$ are estimated from the data with the Baum-Welsh algorithm.
• Discrete time: \( t, t+1, t+2, \ldots \)
• State sequence: \( Q = \{ q_t, q_{t+1}, q_{t+2}, \ldots \} \)
• Observation sequence: \( O = \{ O_t, O_{t+1}, O_{t+2}, \ldots \} \)
  \( O_t = [v_{\text{eye}}, v_{\text{head}}]^T \)
• Transition probabilities: \( a_{ij} = P(q_{t+1} = S_j | q_t = S_i), \ 0 \leq i,j \leq 3 \)
• Probabilities of observation variable: \( P_j(O_t) \sim \text{Normal}(\mu_j, \Sigma_j) \)
• A total of 3 minutes of smooth pursuits of various velocities and VORs were recorded from one subject
• The head movements were measured using a magnetic Fastrack system
• The algorithm was able to classify the occurring fixations, saccades with 100% accuracy
• 65% of the smooth pursuits and 100% of the VORs were classified correctly
Work to do...

• Validation of classification results through comparison with expert classification
• Develop a system with a higher resolution camera for better accuracy
• Implementing the nonlinear minimization in C
... for a great undergraduate experience at the Center for Imaging Science

to Dr. Jeff Pelz for the introduction to research in visual perception

to Dr. Elliott Horch for being a model advisor and great teacher

to Dr. Roger Easton for the introduction to the joy of linear systems

to Dr. Jon Arney for reminding me

to Jason Babcock for the
Questions!