Eye Gaze Estimation

Kevan Yuen (kcyuen [at] eng [dot] ucsd [dot] edu)
Electrical and Computer Engineering
Outline

- Eye Center Localization (Isophotes) [3]
- Head Pose Estimation (Cylindrical Head Model) [10]
- Combining Eye Center Localization and Cylindrical Head Model [17]
- Visual Gaze Estimation [24]

Note: Topics start at the slide numbers labeled in brackets []
Eye Center Localization

- Method described by reference [44] ("Accurate Eye Center Location and Tracking Using Isophote Curvature" [Valenti, Gevers])

- Eyes are characterized by radially symmetric brightness patterns

- Isophotes of an image are curves connecting points of equal intensity
Eye Center Localization
Isophotes (Background)

- Eyes are characterized by radially symmetric brightness patterns

- Isophotes of an image are curves connecting points of equal intensity
Eye Center Localization
Isophotes (Background)

Toy Image

Isophote curvature at edges
(canny filter used for presentation purposes)

Images from “Accurate Eye Center Location…” [Valenti, Gevers]
Eye Center Localization
Isophotes (Background)

- The curvature (or reciprocal of the radius)

\[
\kappa = - \frac{L_y^2 L_{xx} - 2L_x L_{xy} L_y + L_x^2 L_{yy}}{(L_x^2 + L_y^2)^{3/2}}
\]

- The unit-normal of each pixel is given by:

\[
\hat{w} = \frac{\{L_x, L_y\}}{\sqrt{L_x^2 + L_y^2}}
\]

- Taking the unit normal and multiplying by radius gives a rough estimate of the center

\[
D(x, y) = -\frac{\{L_x, L_y\}(L_x^2 + L_y^2)}{L_y^2 L_{xx} - 2L_x L_{xy} L_y + L_x^2 L_{yy}}
\]

Equations from “Accurate Eye Center Location…” [Valenti, Gevers]
Eye Center Localization
Isophotes (Background)

- $D(x,y)$ gives the offset of the estimated center from the pixel at $(x,y)$

- Weighting contribution (at each pixel) by curvedness:

  $$\text{curvedness} = \sqrt{L_{xx}^2 + 2L_{xy}^2 + L_{yy}^2}$$

Equations from “Accurate Eye Center Location…” [Valenti, Gevers]
Eye Center Localization
Isophotes (Background)

Toy Image  (Unit normals)*(radius)  Weighted Accumulator

Images from “Accurate Eye Center Location…” [Valenti, Gevers]
Eye Center Localization
Isophotes (Application)

- Only works well when face is frontal.
- Works off the assumption that eye structure is symmetric, otherwise accuracy drops

Images from “Combining Head Pose…” [Valenti, Sebe, Gevers]
Head Pose Estimation

- Method described by reference [49] (“Robust Full-Motion Recovery of Head by Dynamic Templates and Re-registration Techniques” [Xiao, Kanade, Cohn])

- Provides information on how to normalize the image of the eye such that it will be symmetric for analysis

- Cylindrical Head Model (CHM)
Head Pose Estimation
Cylindrical Head Model (Math Background)

\( P_2 \) (maps 3D to 2D)
\( M \)
\( P_1 \) (maps 2D to cylinder)

\[ u' = F(u, \mu) \]
\[ u = (u, v) \]

Images from “Combining Head Pose…” [Valenti, Sebe, Gevers]
We want to solve for the matrix $M$ which rotates/translates the cylindrical head model:

$$
M = \begin{bmatrix}
1 & -\omega_z & \omega_y & t_x \\
\omega_z & 1 & -\omega_x & t_y \\
-\omega_y & \omega_x & 1 & t_z \\
0 & 0 & 0 & 1
\end{bmatrix}
$$

Which can be parameterized by 6 parameters:

$$
\mu = [\omega_x, \omega_y, \omega_z, t_x, t_y, t_z]
$$

Equations from “Robust Full-Motion Recovery…” [Xiao, Kanade, Cohn]
The assumption is that:

\[ I(F(u, \mu), t + 1) = I(u, t) \]

So we want to minimize the error:

\[
\min \quad E(\mu) = \sum_{u \in \Omega} (I(F(u, \mu), t + 1) - I(u, t))^2
\]

For which the solution is solved as:

\[
\mu = -\left( \sum_{\Omega} \begin{pmatrix} I_u F_\mu \end{pmatrix}^T \begin{pmatrix} I_u F_\mu \end{pmatrix} \right)^{-1} \sum_{\Omega} \begin{pmatrix} I_t (I_u F_\mu)^T \end{pmatrix}
\]

Equations from “Robust Full-Motion Recovery…” [Xiao, Kanade, Cohn]
Head Pose Estimation
Cylindrical Head Model (Math Background)

- The transformation from $u$ to $u'$ can be solved as:

$$F(u, \mu) = \begin{bmatrix} x - y\omega_z + z\omega_y + t_x \\ x\omega_z + y - z\omega_x + t_y \end{bmatrix} \frac{f_L}{-x\omega_y + y\omega_x + z + t_z}$$

- By taking the jacobian with respect to $\mu$, and setting $\mu = [0, \ldots, 0]$ as the initialization:

$$F_{\mu} \bigg|_{\mu=0} = \begin{bmatrix} -xy & x^2 + z^2 & -yz & z & 0 & -x \\ -\left(y^2 + z^2\right) & xy & xz & 0 & z & -y \end{bmatrix} \cdot \frac{f_L}{z^2}(t)$$

Equations from “Robust Full-Motion Recovery…” [Xiao, Kanade, Cohn]
Head Pose Estimation
Cylindrical Head Model (Math Background)

- Then solve for $\mu$ using the initialization

$$
\mu = -\left( \sum_{\Omega} \left( I_u F_\mu \right)^T \left( I_u F_\mu \right) \right)^{-1} \sum_{\Omega} \left( I_t \left( I_u F_\mu \right)^T \right)
$$

- Using the new $\mu$, solve for an updated $F_\mu$

- Solve for $\mu$, using the new $F_\mu$ and repeat until $\mu$ converges

Equations from “Robust Full-Motion Recovery…” [Xiao, Kanade, Cohn]
Head Pose Estimation
Cylindrical Head Model (Video)

- https://www.youtube.com/watch?v=Etj_aktbnwM
The eye center localization method discussed earlier assumes frontal face for symmetry of each individual eye patch.

The Cylindrical Head Model pose tracker may erroneously converge to a local minima which may not be recoverable after that.

Combine the two methods as a check on each other.
CHM & Eye Center Localization
Initialization

- Initialize first frame as frontal face
- Initialize cylinder with parameters $\mathbf{\mu} = [0, \ldots, 0]$
- Detect the eye center locations and project onto cylinder (3-D points) for reference
Using the pose of the CHM and the 3-D eye reference points on the cylinder, a patch of pixels around the reference eye centers can be selected.

Images from “Combining Head Pose…” [Valenti, Sebe, Gevers]
CHM & Eye Center Localization
Eye Location by Pose Cues

- The patch of pixels are remapped into a normalized canonical view

- Eye center localization can then be applied to the posed normalized eye regions

- Assuming CHM is accurate, then these new eye centers are considered optimal

Images from “Combining Head Pose…” [Valenti, Sebe, Gevers]
CHM & Eye Center Localization
Eye Location by Pose Cues

- Yellow +: Reference point from which block of pixels are selected from
- Purple Dot: The tracked eye center without pose-normalization
- Blue X: The tracked eye center with pose-normalization

Images from “Combining Head Pose…” [Valenti, Sebe, Gevers]
With the more accurate pose-normalized eye locations estimated

A pose vector is calculated based off the pose-normalized eye locations

If the distance between the pose vector (from eye) and the pose vector (from CHM) is greater than a certain threshold, then the vectors are averaged to adjust the CHM
CHM & Eye Center Localization
Pose Estimation by Eye Location Cues

Images from “Combining Head Pose…” [Valenti, Sebe, Gevers]
Visual Gaze Estimation

- Visual field of view defined only by head pose (CHM)

- Point of interest defined by eyes (eye center localization) which cannot be outside the visual field of view
Visual Gaze Estimation Calibration

**Target Plane**: computer screen with known calibration points

**Calibration Plane**: points on target plane projected to Calibration Plane

Information about eye displacement is measured for each calibration point

Images from “Combining Head Pose…” [Valenti, Sebe, Gevers]
Visual Gaze Estimation Application

- Visual field of view (yellow)
- Point of Interest defined by eye center localization
- From calibration data, an interpolation can be made to see where the person is gazing at on the calibration plane.
- Project that point onto the target plane shown on the right (red dot on the blue bottle)

Images from “Combining Head Pose…” [Valenti, Sebe, Gevers]
Visual Gaze Estimation Experiment: Dot Following

- Follow a moving dot on the screen in a natural way

Images from “Combining Head Pose…” [Valenti, Sebe, Gevers]
Visual Gaze Estimation
Experimental Results: Dot Following

- Mean Error: (87.18, 103.86) pixels (1.9, 2.2) degrees

- Ground Truth: collected by recording face of subject and corresponding on-screen coordinates where subjects are looking.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Pose-Retargeted</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
</tr>
<tr>
<td>1</td>
<td>75.95, 117.01</td>
</tr>
<tr>
<td>2</td>
<td>79.16, 115.87</td>
</tr>
<tr>
<td>3</td>
<td>78.73, 128.01</td>
</tr>
<tr>
<td>4</td>
<td>99.29, 115.16</td>
</tr>
<tr>
<td>5</td>
<td>85.77, 101.13</td>
</tr>
<tr>
<td>6</td>
<td>80.25, 78.27</td>
</tr>
<tr>
<td>7</td>
<td>93.81, 104.64</td>
</tr>
<tr>
<td>8</td>
<td>92.25, 92.77</td>
</tr>
<tr>
<td>9</td>
<td>92.52, 99.32</td>
</tr>
<tr>
<td>10</td>
<td>91.92, 92.21</td>
</tr>
<tr>
<td>11</td>
<td>89.36, 98.12</td>
</tr>
</tbody>
</table>

Images from “Combining Head Pose…” [Valenti, Sebe, Gevers]
References

- Main Paper:
  “Combining Head Pose and Eye Location Information for Gaze Estimation”
  [Valenti, Sebe, Gevers]

- Reference 44 of “Combining Head Pose…”:
  “Accurate Eye Center Location and Tracking Using Isophote Curvature”
  [Valenti, Gevers]

- Reference 49 of “Combining Head Pose…”:
  “Robust Full-Motion Recovery of Head by Dynamic Templates and Re-registration Techniques”
  [Xiao, Kanade, Cohn]