Optical flow based Head Movement and Gesture Analysis in Automotive Environment

Presentation for IEEE Intelligent Transportation Systems Conference

Presenter: Shinko Y. Cheng
Authors: Sujitha Martin, Cuong Tran, Ashish Tawari, Jade Kwan, and Mohan M. Trivedi
Date: Sept. 16th - 19th 2012
Outline

- Introduction and Motivation
  - Key Terms and Research Issues
  - Related Studies
- Optical flow based Head Movement and Gesture Analyzer (OHMeGA)
  - Concept and Algorithms
  - Noise and Other Practical Matters
- Experimental Results
- Concluding Remarks
Head pose is the 3D orientation of a head.
Head pose is the 3D orientation of a head.
Head dynamics is the motion that describes the change in head position.
Terminology

- Head pose is the 3D orientation of a head.
- Head dynamics is the motion that describes the change in head position.
- Head gesture entails how the head moved from the starting orientation to the ending orientation.
Safety of the driver and those in the vicinity is highly dependent on the driver’s awareness of the constantly changing driving environment.

Head gesture detection and analysis is a vital part of looking inside a vehicle when designing intelligent driver assistance systems (IDAS).
Motivation [video]
Continuous head pose estimation for head gesture analysis is computationally intensive.

Higher level cues
- Fixation time and location
- Rate of motion and rate of change in motion

System goals for head gesture analysis in IDAS
- Runs in real-time
- User-independent
- Simple to implement and set-up
- Robust and accurate
Related Studies

- Feature vectors like head motion histograms (from head pose) for lane change intent prediction [1].

- Head nodding frequency using head pose to determine driver vigilance [2].

- Foot gesture analysis using optical flow in prediction of driver behavior [3].


OHMeGA: Concept

- Intuitiveness: head gestures can be segmented into head motion states and no-head motion (fixation) states.
- Higher level cues: rate of head motion, rate of change in head motion, and fixation time.
OHMeGA: State Machine

- Rule based state machine
- Two types of states
  - Dynamic (motion)
  - Static (no-motion or fixation)
- Two parts
  - Horizontal motion
  - Vertical motion
- Each set of colored arrows represent the flow of one of four unique head gestures.
From a frontal facing camera, head motion in yaw and pitch rotation angle can be represented as motion in the vertical and horizontal direction of the 2D image plane.

Steps to compute global flow vector:

- Interest point detection
- Lucas-Kanade’s optical flow algorithm

\[ \mathbf{u} = -S^{-1}d \]

\[ S = \begin{bmatrix} \sum_{p_i \in \mathbb{N}} I_x(p_i)I_x(p_i) & \sum_{p_i \in \mathbb{N}} I_x(p_i)I_y(p_i) \\ \sum_{p_i \in \mathbb{N}} I_x(p_i)I_y(p_i) & \sum_{p_i \in \mathbb{N}} I_y(p_i)I_y(p_i) \end{bmatrix} \]

\[ d = \begin{bmatrix} \sum_{p_i \in \mathbb{N}} I_x(p_i)I_t(p_i) \\ \sum_{p_i \in \mathbb{N}} I_y(p_i)I_t(p_i) \end{bmatrix} \]

- Majority vote on optical flow vectors
Non-ideal conditions are:

- Finite frame rate

- Noise in the camera sensors (i.e. camera vibrations)

- Motions detected by optical flow in the image plane may not be only due to head movements (i.e. hand movements near the face)

- No-direct correspondence between head rotation in yaw (pitch) angle to horizontal (vertical) motion in the image plane
OHMeGA: State Transitions

- **Head gesture:**
  FxS $\rightarrow$ ML $\rightarrow$ FxL $\rightarrow$ MR $\rightarrow$ FxS

- **Top curve:** horizontal motion detected in the image plane with state labels

- **Bottom curve:** ground truth state labels

- Using threshold and area under the curve to handle noise
OHMeGA is evaluated on two sets of data

- **In-laboratory:**
  - 5 subjects (~600 head gestures),
  - Subjects followed instructions (i.e. “STOP” and “GO”) by pressing the brake or the accelerator pedal
  - Subjects answered “distractions” in the form of mathematical equations on the right side monitor.

- **On-road:** manually selected head gestures for preliminary evaluations
Results [Video]

Motion in the x-direction of the image plane

Frame

Global flow vector with labels

Ground truth labels

- FxS
- ML
- FxL
- MR
Results: On-Road

Figure: Data collected using frontal facing camera from on-road experiment is processed first using optical flow to obtain head motions, then annotated using OHMeGA analyzer and finally separated into three types of gestures.

<table>
<thead>
<tr>
<th></th>
<th>FxS</th>
<th>FxR</th>
<th>FxL</th>
<th>FxD</th>
</tr>
</thead>
<tbody>
<tr>
<td>FxS</td>
<td>0.943</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>FxR</td>
<td>0</td>
<td>0.76</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>FxL</td>
<td>0</td>
<td>0</td>
<td>0.833</td>
<td>0</td>
</tr>
<tr>
<td>FxD</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.806</td>
</tr>
<tr>
<td>Samples</td>
<td>926</td>
<td>25</td>
<td>60</td>
<td>31</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>MR</th>
<th>ML</th>
<th>MD</th>
<th>MU</th>
</tr>
</thead>
<tbody>
<tr>
<td>MR</td>
<td>0.8</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ML</td>
<td>0</td>
<td>0.917</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MD</td>
<td>0</td>
<td>0</td>
<td>0.913</td>
<td>0</td>
</tr>
<tr>
<td>MU</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.907</td>
</tr>
<tr>
<td>Samples</td>
<td>145</td>
<td>120</td>
<td>69</td>
<td>86</td>
</tr>
</tbody>
</table>
Conclusion

- OHMeGA is user-independent, simple to implement and set up, and runs in real-time.
- This implementation of OHMeGA relies only on head dynamics.
- OHMeGA can derive higher level cues such as fixation time and relative rate of motion.
Future Works

- Represent 3D head motion in the yaw and pitch rotation angle with both horizontal and vertical motions in the 2D image plane
- Optimize global flow vector calculation for out of plane rotations (currently optimal for in plane movements).
Thank you

- Colleagues in Laboratory for Intelligent and Safe Automobiles, UC San Diego
Thank you!

Any Questions?