Real-Time Computer Vision for Human Interfaces

Yoshio Matsumoto
Nara Institute of Science and Technology, Japan
Introduction to Robotics Lab at NAIST
video
Face Measurement System
and Its Applications

Yoshio Matsumoto
Nara Institute of Science and Technology, Japan
Motivation: Face Measurement

- For Human-Robot Interaction,
  - when a person teaches a task to a robot
  - when a person makes a cooperative task with a robot
  natural ways of communication is required.

- Head motion and gaze direction reflect intention and attention of a person.

Goal of this research:

Head Motion \(\Rightarrow\) Gesture \(\Rightarrow\Rightarrow\Rightarrow\) Intention
Gaze Direction \(\Rightarrow\) Attention
### Gaze Tracking Techniques: Corneal Reflection

- Head should not move, or head pose is measured by magnetic sensor etc.
- Eye rotation is detected using IR reflection on corneal
- Head mounted device prohibits natural behaviors
- Accurate in best condition, however hard to keep it
- Binocular systems and external camera system are also available
Gaze Tracking Techniques: Corneal Reflection (cont’d)

- Purkinje images appear as small white dots in close proximity to the (dark) pupil
- tracker calibration is achieved by measuring user gazing at properly positioned grid points (usually 5 or 9)
- tracker interpolates POR on perpendicular screen in front of user

Figure: Pupil and Purkinje images as seen by eye tracker’s camera
Gaze Tracking Techniques: EOG (electro-oculography)

- most widely used method some 20 years ago (still used today)
- measures eye movements relative to head position
- not generally suitable for POR measurement (unless head is also tracked)

Figure: EOG measurement:
- relies on measurement of skin’s potential differences using electrodes placed around the eye
Gaze Tracking Techniques:
Scleral Contact Lens/Search Coil

Figure: Scleral coil:
- search coil embedded in contact lens and electromagnetic field frames
- possibly most precise
- similar to electromagnetic position/orientation trackers used in motion-capture
Gaze Tracking Techniques: Scleral Contact Lens/Search Coil (cont’d)

Figure: Example of scleral suction ring insertion:
- most intrusive method
- insertion of lens requires care
- wearing of lens causes discomfort

- highly accurate, but limited measurement range (~5°)
- measures eye movements relative to head position
- not generally suitable for POR measurement (unless head is also tracked)
Real-Time Vision for Face Measurement

• One of the important topics in PUI research
• Being actively studied at Microsoft, IBM, MIT, CMU, Univ. of Illinois etc.

• Few of them can measure the quantitative gaze direction
• Most of them use monocular camera systems

Our Approach

- Stereo camera system for 3D measurement
- 3D facial model
- Feature tracking by normalized correlation
- 3D model fitting algorithm based on spring model

• Real-time Face Tracking
• Gaze direction, blinking and lip motion are additionally measured
1. Algorithm of Face Measurement
System Configuration

- PC with Pentium III 450MHz or higher
- OS : Linux 2.2 or 2.4
- Stereo Camera Pair
3D Facial Model

- Corners of eyes and mouth are used for tracking
- 3D Facial Model = Template Images + 3D coordinates

<table>
<thead>
<tr>
<th>Template image for feature tracking</th>
<th>3D coordinates of features</th>
<th>Template image of whole face</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Template Image" /></td>
<td>(-50.7, 13.4, -4.2)</td>
<td><img src="image2.png" alt="Template Image" /></td>
</tr>
<tr>
<td><img src="image3.png" alt="Template Image" /></td>
<td>(-18.2, 12.8, -1.9)</td>
<td></td>
</tr>
<tr>
<td><img src="image4.png" alt="Template Image" /></td>
<td>(16.1, 12.8, -2.1)</td>
<td></td>
</tr>
<tr>
<td><img src="image5.png" alt="Template Image" /></td>
<td>(48.0, 11.0, -10.8)</td>
<td></td>
</tr>
<tr>
<td><img src="image6.png" alt="Template Image" /></td>
<td>(-25.9, -58.7, 10.1)</td>
<td></td>
</tr>
<tr>
<td><img src="image7.png" alt="Template Image" /></td>
<td>(27.6, -58.6, 4.5)</td>
<td></td>
</tr>
</tbody>
</table>
3D Model Fitting for Face Tracking

\[ E = \sum_{i=0}^{n-1} W_i (Rx_i + t - y_i)^T (Rx_i + t - y_i) \]

\( n \) : # of features  
\( x_i \) : Coordinate of each feature in the model  
\( y_i \) : Coordinate of each measurement  
\( W_i \) : Reliability for each measurement (0..1)

\[ \text{Minimizing } E \]

\( R(\theta, \phi, \varphi) \) : Rotation matrix  
\( T(x, y, z) \) : Translation vector
3D Model Fitting for Face Tracking
(cont’d)

Assumption: displacement between previous and current frame is small

At time $t$

**Measurement**

$T(x, y, z)$
$R(\theta, \phi, \varphi)$

**Model**

$T(0, 0, 0)$
$R(0, 0, 0)$

At time $t + \Delta t$

**Measurement**

$T(x + \Delta x, y + \Delta y, z + \Delta z)$
$R(\theta + \Delta \theta, \phi + \Delta \phi, \varphi + \Delta \varphi)$

**Model**
3D Model Fitting for Face Tracking
(cont’d)

Gradient 3D model fitting method based on virtual spring model

At time $t + dt$

$R(-\theta, -\phi, -\varphi)$
$T(-x, -y, -z)$

Displacement due to $(\Delta \theta, \Delta \phi, \Delta \varphi)$ and $(\Delta x, \Delta y, \Delta z)$ still remains

Stiffness of spring $k_i = W_i$ (Reliability)

$R$ and $t$ are acquired simultaneously
Result of Face Tracking

System Overview
Accuracy of Face Tracking

Measurable Range

- Roll: $-35 \sim +35$[deg]
- Pitch: $-80 \sim +80$[deg]
- Yaw: $-35 \sim +35$[deg]

Accuracy

- Approx. $2$[deg]
# Estimation of Gaze Direction

1. Get 3D position of both ends of an eye

2. Calculate center of eyeball

3. Calculate size of eyeball

4. Detect iris

5. Calculate 3D gaze direction
Result of Gaze Measurement

- Look at markers from ① to ⑩ with intervals of 10cm
- Intersection of 3D gaze vector and the board are displayed as a fixation point
Result of Gaze Measurement

Gaze Line Detection
Result of Gaze Measurement

Accuracy: Approx. 5[deg]
Result of Face Measurement
Hardware Configuration

Selectable from below combinations depending on requirements and costs

| Desktop PC                  | NTSC Camera × 2            |
|                            | NTSC Camera × 2 (Multiplexed) |
|                            | IEEE1394 Camera × 2        |
| Notebook PC                | IEEE1394 High-Speed Camera (80Hz) |

**IEEE1394 High-Speed Stereo Camera, 80Hz**

- 30Hz, 1.5kg, US$2000
Field Multiplexing Device

- Multiplexed stereo video streams into a single video stream.
- Vertical resolution of each video signal becomes half.
- Can be used for any conventional image processing system.
Attention Recognition

The object that a person is looking at is recognized by calculating $\theta_i$ for all objects.
Attention Recognition
Attention Recognition

Attention Recognition based on Gaze Line Detection
Gesture Recognition

Spotting gesture recognition based on Continuous DP Matching using head motions (velocity, angular velocity)
Gesture Recognition

00:00:13.573 (34ns)
\[
\begin{align*}
\{ & X, Y, Z \\
12 & , -16 , -622 \\
\} \\
\{ & \text{Yaw, Pitch, Roll} \\
12 & , -7 , 1 \\
\}
\end{align*}
\]
R eye: Open
L eye: Open
Gesture: Yes

00:00:30.512 (34ns)
\[
\begin{align*}
\{ & X, Y, Z \\
0 & , 0 , -449 \\
\} \\
\{ & \text{Yaw, Pitch, Roll} \\
3 & , -1 , 3 \\
\}
\end{align*}
\]
R eye: Open
L eye: Open
Gesture: ZoomIn

00:00:45.011 (34ns)
\[
\begin{align*}
\{ & X, Y, Z \\
24 & , 8 , -632 \\
\} \\
\{ & \text{Yaw, Pitch, Roll} \\
0 & , 10 , 0 \\
\}
\end{align*}
\]
R eye: Open
L eye: Open
Gesture: No

00:00:24.601 (34ns)
\[
\begin{align*}
\{ & X, Y, Z \\
11 & , -1 , -622 \\
\} \\
\{ & \text{Yaw, Pitch, Roll} \\
2 & , -4 , 0 \\
\}
\end{align*}
\]
R eye: Closed
L eye: Open
Gesture: Winking
Gesture Recognition based on Face Tracking
Specs of Developed System

**Accuracy**
- Head pos: 2mm
- Dir: 2deg
- Gaze dir: 5deg

**Processing Speed**
- 30Hz～80Hz
- (depending on camera)
Software Configuration

Flowchart showing the process of face recognition and identification, including modules such as TrueFace Server, Face Identification Module, Face Detection Module, Face Measurement Module, Face Motion Recognition Module, Capture Library, Application Software, and various camera options like NTSC-mix, NTSC, MEGA-D, and IEEE1394 camera.
Software Configuration

Commercial Face Recognition Software

Face Model Database

TrueFace Server

Face Identification Module

Shared Memory

Face Detection Module

Face Measurement Module

Face Motion Recognition Module

Start

Capture Library

NTSC-mix camera

NTSC camera

MEGA-D camera

IEEE1394 camera

Application Software

Success?

Yes

No
Software Configuration

Face Mode Database

TrueFace Server

Face Identification Module

Face Detection Module

Face Measurement Module

Face Motion Recognition Module

Start

Capture Library

Success?

Application Software

Video Capturing Library

NTSC-mix camera

NTSC camera

MEGA-D

IEEE1394 camera
Software Configuration

Start → Face Detection Module → Face Identification Module

Start → Face Measurement Module → Face Motion Recognition Module

Start → Application Software

Capture Library:
- NTSC-mix camera
- NTSC camera
- MEGA-D
- IEEE1394 camera
Potential Application Areas

Since this system is non-contact, passive, and inexpensive, it can be applied to many application areas where conventional systems cannot be used.

- **Human Interfaces**
  - Computer Interface (e.g. Hands-free mouse)
  - Robot Interface (e.g. Eye-contact communication)
  - Safety System (e.g. Driver support)
  - Assistive Products for the disabled
- **Human Modeling**
  - Cognitive Science (Experiment on visual cognition)
  - Ergonomics (e.g. Human-friendly design)
2. Application to Human Interfaces
2.1 Computer Interfaces
Direct Usage of Head Movements: Hands-Free Mouse
How to Use Eye Movements for Computer Interfaces?

- Eye movement [Glenstrup 95]
  - Convergence
  - Rolling
  - Saccades
  - Pursuit motion
  - Nystagmus
  - Drift and micro-saccades
  - Physiological nystagmus

- Need to refine raw data:
  - distinguish Fixations from Saccades
How to Use Eye Movements for Computer Interfaces?

Saccade/Fixation detection

- **Velocity-Threshold**
  - Saccades > 300 deg/sec.
  - Fixations < 100 deg/sec.
  - Usual threshold 200 deg/sec.

- **Dispersion-Threshold**
  \[
  D = \frac{\max(x) - \min(x)}{\max(y) - \min(y)}
  \]
  - Threshold set such that visual angle is between 0.5° and 1°.
How to Use Eye Movements for Computer Interfaces?

• Command based Interface
  – Obvious application: Selection of objects {Menu selection, Window scrolling, …} \(\Rightarrow\) Pointing.
  – Midas Touch problem: Eyes not a control device.
  – Use *dwell time* to trigger a selection.

• Non-Command Interfaces
  – The computer monitors user’s actions instead of waiting for user’s command.
  – Potential Applications: User Support
    \(\Rightarrow\) Detect difficulties and provide translation support of difficult words
Pro-Active Dictionary: How to detect User Difficulties?

Gaze Pattern in Normal Reading

Gaze Pattern when Difficulties Encountered
Pro-Active Dictionary: Implementation

Windows

IE Web Browser

Assistance Module

Data buffering & Feedback

Linux

Gaze Tracking Module

User

Gaze

Action
Pro-Active Dictionary: Demonstration
2.1 Robot Interfaces
Intelligent Wheelchair

CCD cameras

Ultrasonic sensors

PC

Laser Scanner

Notebook PC

CCD cameras
Intelligent Wheelchair

How is facial information used?

• Gesture
  – Nodding -> To start
  – Shaking -> To stop

• Face Direction
  – To determine the direction to move

• Gaze Direction
  – To determine if the user is concentrated
Intelligent Wheelchair
Intelligent Wheelchair

Various lighting conditions
Intelligent Wheelchair
Intelligent Wheelchair

Sensor-based collision avoidance and wall following
Result of Experiment (Trajectory)
Result (Input and Output Values)
Intelligent Wheelchair

How to detect concentration of the user?
Intelligent Wheelchair
Estimation of User’s Attention
Intelligent Wheelchair

Estimation of User’s Attention
## Intelligent Wheelchair

### Estimation of User’s Attention

<table>
<thead>
<tr>
<th>Paying attention to the poster</th>
<th>Not paying attention to the poster</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Diagram" /> u Concentrated on the position of the poster.</td>
<td><img src="image2.png" alt="Diagram" /> u Distributed on various positions.</td>
</tr>
</tbody>
</table>
Intelligent Wheelchair

Estimation of User’s Attention
Receptionist Robot ASKA
Receptionist Robot ASKA

- Speech Recognition and Voice Synthesis
- Human and Face Recognition
- Information Retrieval from Databases
- NLP Dialog Model
- Gesture Generation and Autonomous Mobility

Platform for experiment in real world
ASKA: System Overview

- Camera
- Microphone
- Speaker
- PC for recognition and speech
- PC for vision
- PC for server
- PC for gesture
Hardware Configuration

Base: tmsuk04 (tmsuk Co., LTD.)
Degree of freedom:
  Chest   1 (DOF)
  Arms    14 (DOF)
  Hands   6 (DOF)

telecamera
wide-range camera

Human detection and tracking
Measurement of facial information
Software Configuration

Person Finder  Speech Recognizer  Talker

Server

Data

Body Gesture Controller  Head Motion Controller
Receptionist Robot ASKA
Problem to Solve with Vision

Problem

Wrong responses to:
- Background noises
- Utterances which are not spoken to ASKA

Purpose of This Research:
- To recognize the utterance period
- To detect whether user speaks to ASKA or not

For natural human–robot interaction.
Software Configuration

1. Human detection and tracking
2. Detection of the sign of utterance
Pilot Study

Dialog Experiment

- **Measured information**
  - Head orientation
  - Gaze
  - Lip motion
  - Blink

- **Subject**
  - 10

- **Speech sentence**
  - Fixed sentence: 5
  - Free sentence: 5
Information at Utterance

Subject A

EURON Summer School 2003
Sign of Utterance: Start

**Subject B**

*Head orientation*

*Gaze*

The Face and gaze direct to object before the utterance
Sign of Utterance: Start

Face and gaze vectors are within the threshold area.

A user have an intention of utterance

Allowing area
From the center of ASKA’s face:
- Gaze vector area
  300 [mm]
- Face vector area
  400 [mm]

Detection of Start of Utterance
Experiment:
Detection of Utterance Sign
Demonstration: Interaction Based on Utterance Sign
3. Application to Human Modeling
Measurement of Infants
Measurement of Infants
Measurement of Infants

Measured Results

![Graph showing measured data and missing data markers.](image)
Measurement of TV Game Players
Measurement of TV Game Players

Frequency of Blinking

This result coincides with Psychological Report
Measurement of Drivers

System overview
Measurement of Drivers

Software Configuration
Measurement of Drivers
Measurement of Drivers
Measurement of Drivers

Experiment at night
Measurement of Drivers

Horizontal head movements at curves

Ave. Speed : 20[km/s]  Ave. Speed : 40[km/s]

Not passive head movements due to centrifugal force, but active head movements to cope with centrifugal force.
Measurement of Drivers

Vertical head position in long driving

- Body posture changed after long driving?
- Body lowered due to the softness of the seat?
Measurement of Drivers

Measured fixation point and yaw rate of vehicle

![Graph showing measured fixation point and yaw rate of vehicle over time](graph.png)
Measurement of Drivers

Attention Recognition
Measurement of Drivers

Attention Recognition
Measurement of Drivers

Where are you looking when lane changing?
Measurement of Patients in PET

To compensate head motion during measurement
Summary

Face Measurement System

Accuracy
Head pos: 2mm
dir: 2deg
Gaze dir: 5deg

Processing Speed
30Hz～80Hz
(dependent on camera & CPU)

Human Interfaces

Computer Interface
Robot Interaction

Human Modeling

Cognitive Science
Psychology / Ergonomics
Fin