Sensor Fusion for Context Understanding

Huadong Wu, Mel Siegel
The Robotics Institute, Carnegie Mellon University
Sevim Ablay
Applications Research Lab, Motorola Labs
sensor fusion

- how to combine outputs of multiple sensor perspectives on an observable?
- modalities may be “complementary”, “competitive”, or “cooperative”
- technologies may demand registration
- variety of historical approaches, e.g.:
  - statistical (error and confidence measures)
  - voting schemes (need at least three)
  - Bayesian (probability inference)
  - neural network, fuzzy logic, etc
context understanding

- best algorithm for human-computer interaction tasks depends on context
- context can be difficult to discern
- multiple sensors give complementary (and sometime contradictory) clues
- sensor fusion techniques needed
- (but best algorithm for sensor fusion tasks may depend on context!)
agenda

• a *generalizable* sensor fusion architecture for "*context-aware computing*
  - or (my preference, but not the standard term) "*context-aware human-computer interaction*"

• a realistic test to demonstrate usability and performance enhancement

• improved sensor fusion approach (to be detailed in next paper)
background

- current context-sensing architectures (e.g., Georgia Tech Context Toolkit) tightly couple sensors and contexts
- difficult to substitute or add sensors, thus difficult to extend scope of contexts
- we describe a modular hierarchical architecture to overcome these limitations
toward context understanding

Identification, representation, and understanding of context
Adapt behavior to context

Sensing hardware: cameras, microphones, etc.
Environment situation: people in the meeting room, objects around a moving car, etc.

Humans understand context naturally & effortlessly

Traditional system

Information Separation + Sensor Fusion
methodology

- top-down
- adapt/extend Georgia Tech Context Toolkit (Motorola helps support both groups)
- create realistic context and sensor prototypes
- implement a practical context architecture for a plausible test application scenario
- implement sensor fusion as a mapping of sensor data into the context database
- place heavy emphasis on real sensor device characterization and (where needed) simulation
context-sensing methodology: sensor data-to-context mapping

\[
\begin{bmatrix}
?_1 \\
?_2 \\
\vdots \\
?_n
\end{bmatrix} = \begin{bmatrix}
f_{11}(\cdot) & f_{12}(\cdot) & \cdots & f_{1m}(\cdot) \\
f_{21}(\cdot) & f_{22}(\cdot) & \cdots & f_{2m}(\cdot) \\
\vdots & \vdots & \ddots & \vdots \\
f_{n1}(\cdot) & f_{n2}(\cdot) & \cdots & f_{nm}(\cdot)
\end{bmatrix} \begin{bmatrix}
sensor_1 \\
sensor_2 \\
\vdots \\
sensor_m
\end{bmatrix}
\]

context

observations & hypotheses

sensory output
dynamic database

- example: user identification and posture for discerning focus-of-attention in a meeting
- tables (next) list basic information about environment (room) and parameters, e.g.,
  - temperature, noise, lighting, available devices, number of people, segmentation of area, etc
  - initially many details are entered manually
  - eventually a fully “tagged” and instrumented environment can reasonably be anticipated
- weakest link: maintaining currency
## Context Classification and Modeling

### Inside (Personal Information, Feeling & Thinking, Emotional)
- **Mood**: Merry, sad, sad, satisfied...
- **Agitation/Alertness**: Nervousness
- **Stress**: Focus of attention
- **Concentration**: Preferences
- **Physical Information**: Personal state: Heart rate, respiration rate, blood pressure, Galvanic Resistance, body temperature, sweat

### Outside (Environment)
- **Sound Processing**: Speaker recognition, speaking understanding
- **Image Processing**: Face recognition, object recognition, 3-D object measuring

### Information: Sensors
- **GPS, DGPS, serverIP, RFID, gyro, accelerometers, dead-reckoning**
- **Network Resource**: Thermometer, humidity sensor, photodiode sensors, accelerometers, gas sensor

### Activity
- **Work Content**: Work, entertainment, living, etc.
- **Body Motion**: Drive, walk, sit, etc.
- **Aural (Listen/Talk)**: Content: work, entertainment, living, etc.
- **Hands**: Type, write, use mouse, etc.

### Location and Orientation
- **City, Altitude, Temperature, Humidity, Barometer Pressure, Forecast**
- **Location and Vicinity**
- **Room, Car, Devices**: Function, state, etc.
- **Vicinity Temperature, Humidity**

### Time and Change
- **Day, Date, Time of the Day**
- **Change**: Travelling, Speed, Heading

### Outside (Social)
- **People**: Individuals or group (audience of show, attendees in cocktail party, people interaction, casual chat, formal meeting, eye contact, attention)}
context information architecture: dynamic context information database

Room-table: NSH A417

<table>
<thead>
<tr>
<th>Area-table</th>
<th>Inside Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature</td>
<td>72°F (σ = 3°F)</td>
</tr>
<tr>
<td>Light condition</td>
<td>Brightness grade</td>
</tr>
<tr>
<td>Noise level</td>
<td>60 db (σ = 6 db)</td>
</tr>
<tr>
<td>Devices</td>
<td>...</td>
</tr>
<tr>
<td>Detected people #</td>
<td>6 (κ &gt; 0.5)</td>
</tr>
<tr>
<td>Detected user</td>
<td>...</td>
</tr>
</tbody>
</table>

Preference-table-user[Hd]

<table>
<thead>
<tr>
<th>Name</th>
<th>Huadong Wu (κ = 1.0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight</td>
<td>144 lb (σ = 4 lb)</td>
</tr>
<tr>
<td>Height</td>
<td>5'6&quot; (σ = 0.5&quot;)</td>
</tr>
<tr>
<td>Preference</td>
<td>Preference-table-user[Hd]</td>
</tr>
</tbody>
</table>

History-table-user[Hd]

<table>
<thead>
<tr>
<th>Name</th>
<th>Huadong Wu</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>Place</td>
</tr>
</tbody>
</table>

User-table

<table>
<thead>
<tr>
<th>Name</th>
<th>Background</th>
<th>Place</th>
<th>Confidence</th>
<th>Activity</th>
<th>First detected</th>
<th>History</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alan</td>
<td>Background-table-user[Alan]</td>
<td>Entrance Area</td>
<td>In meeting</td>
<td>Speaking</td>
<td>10:32 AM, 06/06/2001</td>
<td>History-table-user[Alan]</td>
</tr>
<tr>
<td>Mel</td>
<td>Background-table-user[Mel]</td>
<td>Entrance Area</td>
<td>Speaking</td>
<td>Head pose</td>
<td>11:48 AM, 06/06/2001</td>
<td>History-table-user[Mel]</td>
</tr>
<tr>
<td>Hd</td>
<td>Background-table-user[Hd]</td>
<td>Entrance Area</td>
<td>Speaking</td>
<td>Focus of attention</td>
<td>2:48 PM, 06/06/2001</td>
<td>History-table-user[Hd]</td>
</tr>
</tbody>
</table>

User activity

<table>
<thead>
<tr>
<th>user</th>
<th>user ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>contact info.</td>
<td>PK user ID</td>
</tr>
<tr>
<td>preference</td>
<td></td>
</tr>
<tr>
<td>conf. room usage activity</td>
<td></td>
</tr>
</tbody>
</table>

User activity

<table>
<thead>
<tr>
<th>user</th>
<th>user ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>in meeting</td>
<td>PK user ID</td>
</tr>
<tr>
<td>speaking</td>
<td></td>
</tr>
<tr>
<td>head pose</td>
<td></td>
</tr>
<tr>
<td>focus of attention</td>
<td></td>
</tr>
</tbody>
</table>
implementation

- low-level sensor fusion done
- sensing and knowledge integration via LAN
- context maintained in a dynamic database
- each significant entity (user, room, house, ...) has its own dynamic context repository
- dynamic repository maintains and serves context and data to all applications
- synchronization and cooperation among disparate sensor modalities achieved by “sensor fusion mediators” (agents)
system architecture to support sensor fusion for context-aware computing
practical details ...

- context type => sensor fusion mediator
- mediator integrates corresponding sensors, e.g., by designating some primary others secondary based on observed or specified performance
- Dempster-Shafer “theory of evidence” implementation in accompanying paper
- (white-icon components in following cartoon inherited from Georgia Tech CT)
details inherited from GT TC

- JAVA implementation
- BaseObject class provides communication functionality: sending, receiving, initiating, and responding via multithread server
- context widgets, interpreters, and discovers subclass from the BaseObject, and inherit its functionality
- service is part of the widget object
- aggregators subclass from widgets, inheriting their functionality
object hierarchy and subclass relationship in context toolkit
focus-of-attention application

- neural network estimates head poses
- focus of attention estimate based on head pose probability distribution analysis
- audio reports speaker, assumed to be focus of other participants’ attention
- situation is not easy to analyze due to, e.g., dependence of behavior on discussion topic
- initial results suggests we need more general fusion approach than provided by Bayesian
next paper ...

- Dempster-Shafer approach ...
- provides mechanism for handling “belief” and “plausibility”
- cautiously stated, generalizes Bayes’ Law a priori probabilities to distributions
- (difficulty, of course, is that usually neither the requisite probabilities nor the distributions are actually known)
focus-of-attention estimation from video and audio sensors
expectations

Expected Performance Boost

1. Uncertainty & ambiguity representation to user applications
2. Information consolidation & conflict resolving for users
3. Adaptive sensor fusion support switch to suitable algorithms
4. Robust to configuration change — and for some to die gracefully
5. Situational description support — using more & complex context
conclusions

• preliminary experiments demonstrate feasibility of context-sensing architecture and methodology
• expect our further improvements via
  - better uncertainty and ambiguity handling
  - fusion of overlapping sensors
  - context-adaptive widget capability
  - sensor fusion mediator coordinates resources
  - context information server supports applications