Hand Gesture Recognition, Aerobic exercises, Events

Lecture-15

Copyright  Mubarak Shah 2003
Hand Gesture Recognition
Seven Gestures

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(b)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(c)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(d)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(e)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(f)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(g)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Gesture Phases

• Hand fixed in the **start position**.
• Fingers or hand move smoothly to **gesture position**.
• Hand fixed in **gesture position**.
• Fingers or hand return smoothly to **start position**.
Finite State Machine
Main Steps

• Detect fingertips.
• Create fingertip trajectories using motion correspondence of fingertip points.
• Fit vectors and assign motion code to unknown gesture.
• Match
Detecting Fingertips

(a)  (b)  (c)  (d)

Copyright Mubarak Shah 2003
Vector Extraction
Vector Representation of Gestures

Copyright  Mubarak Shah 2003
## Results

<table>
<thead>
<tr>
<th>Run</th>
<th>Frames</th>
<th>L</th>
<th>R</th>
<th>U</th>
<th>D</th>
<th>T</th>
<th>G</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>200</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>2</td>
<td>250</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>3</td>
<td>250</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>X</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>4</td>
<td>250</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>5</td>
<td>300</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>6</td>
<td>300</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>7</td>
<td>300</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>8</td>
<td>300</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>9</td>
<td>300</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>*</td>
</tr>
<tr>
<td>10</td>
<td>300</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

L = Left, R = Right, U = Up, D = Down, T = Rotate, G = Grab, S = Stop, ✓ - Recognized, X - Not Recognized, * - Error in Sequence.

Copyright Mubarak Shah 2003
Publication


Action Recognition Using Temporal Templates

Jim Davis and Aaron Bobick

Copyright  Mubarak Shah 2003
Main Points

• Compute a sequence of difference pictures from a sequence of images.
• Compute Motion Energy Images (MEI) and Motion History Images (MHI) from difference pictures.
• Compute Hu moments of MEI and MHI.
• Perform recognition using Hu moments.
MEI and MHI

Motion-Energy Images (MEI)

\[ E_\tau(x, y, t) = \bigcup_{i=0}^{\tau-1} D(x, y, t-i) \]

Motion History Images (MHI) Change Detected Images

\[ H_\tau(x, y, t) = \begin{cases} \tau & \text{if } D(x, y, t) = 1 \\ \max(0, H_\tau(x, y, t-1) - 1) & \text{otherwise} \end{cases} \]

Copyright Mubarak Shah 2003
MEIs
Color MHI Demo

Copyright  Mubarak Shah 2003
Summary

- Use seven Hu moments of MHI and MEI to recognize different exercises.
- Use seven views (-90 degrees to +90 degrees in increments of 30 degrees).
- For each exercise several samples are recorded using all seven views, and the mean and covariance matrices for the seven moments are computed as a model.
- During recognition, for an unknown exercise all seven moments are computed, and compared with all 18 exercises using Mahalanobis distance.
- The exercise with minimum distance is computed as the match.
- They present recognition results with one and two view sequences, as compared to seven view sequences used for model generation.
Moments

General Moments

\[ m_{pq} = \int \int x^p y^q \rho(x, y) \, dx \, dy \]

Central Moments (Translation Invariant)

\[ \mu_{pq} = \int \int (x - \bar{x})^p (y - \bar{y})^q \rho(x, y) \, d(x - \bar{x}) \, d(y - \bar{y}) \]

\[ \bar{x} = \frac{m_{10}}{m_{00}}, \quad \bar{y} = \frac{m_{01}}{m_{00}} \]

Copyright Mubarak Shah 2003
Central Moments

\[
\begin{align*}
\mu_{00} &= m_{00} \equiv \mu \\
\mu_{01} &= 0 \\
\mu_{10} &= 0 \\
\mu_{20} &= m_{20} - \mu \bar{x}^2 \\
\mu_{11} &= m_{11} - \mu \bar{x} \bar{y} \\
\mu_{02} &= m_{02} - \mu \bar{y}^2 \\
\mu_{30} &= m_{30} - 3m_{20} \bar{x} + 2\mu \bar{x}^3 \\
\mu_{21} &= m_{21} - m_{20} \bar{y} - 2m_{11} \bar{x} + 2\mu \bar{x}^2 \bar{y} \\
\mu_{12} &= m_{12} - m_{02} \bar{x} - 2m_{11} \bar{y} + 2\mu \bar{x} \bar{y}^2 \\
\mu_{03} &= m_{03} - 3m_{02} \bar{y} + 2\mu \bar{y}^3
\end{align*}
\]
Moments

Hu Moments: translation, scaling and rotation invariant

\[
\begin{align*}
\nu_1 &= \mu_{20} + \mu_{02} \\
\nu_2 &= (\mu_{20} - \mu_{02})^2 + \mu_{11}^2 \\
\nu_3 &= (\mu_{30} - 3\mu_{12})^2 + (3\mu_{12} - \mu_{03})^2 \\
\nu_4 &= (\mu_{30} + \mu_{12})^2 + (\mu_{21} + \mu_{03})^2 \\
&\vdots
\end{align*}
\]

Copyright  Mubarak Shah 2003
Half size, mirror
Rotated 2, rotated 45
<table>
<thead>
<tr>
<th>Invariant (Log)</th>
<th>Original</th>
<th>Half Size</th>
<th>Mirrored</th>
<th>Rotated 2°</th>
<th>Rotated 45°</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi_1$</td>
<td>6.249</td>
<td>6.226</td>
<td>6.919</td>
<td>6.253</td>
<td>6.318</td>
</tr>
<tr>
<td>$\phi_2$</td>
<td>17.180</td>
<td>16.954</td>
<td>19.955</td>
<td>17.270</td>
<td>16.803</td>
</tr>
<tr>
<td>$\phi_3$</td>
<td>22.655</td>
<td>23.531</td>
<td>26.689</td>
<td>22.836</td>
<td>19.724</td>
</tr>
<tr>
<td>$\phi_4$</td>
<td>22.919</td>
<td>24.236</td>
<td>26.901</td>
<td>23.130</td>
<td>20.437</td>
</tr>
<tr>
<td>$\phi_5$</td>
<td>45.749</td>
<td>48.349</td>
<td>53.724</td>
<td>46.136</td>
<td>40.525</td>
</tr>
<tr>
<td>$\phi_6$</td>
<td>31.830</td>
<td>32.916</td>
<td>37.134</td>
<td>32.068</td>
<td>29.315</td>
</tr>
<tr>
<td>$\phi_7$</td>
<td>45.589</td>
<td>48.343</td>
<td>53.590</td>
<td>46.017</td>
<td>40.470</td>
</tr>
</tbody>
</table>

Hu moments

Copyright Mubarak Shah 2003
PAT (Personal Aerobic Trainer)

Copyright Mubarak Shih 2003
PAT (Personal Aerobic Trainer)


Copyright  Mubarak Shah 2003
PAT (Personal Aerobic Trainer)

Copyright  Mubarak Shah 2003
A Framework for the Design of Visual Event Detectors

Niels Haering

Copyright  Mubarak Shah 2003
Our Framework

- detected events
  - event inference
  - shot summarization
    - motion-blob verification
      - object classification
        - color/texture analysis
        - spatio-temporal analysis
        - motion estimation
        - shot-boundary detection

- video sequences
Object Classification

Fourier transform
Gabor filters
steerable filters
greylevel co-occurrence matrix
fractal dimension
entropy
color

neural network

Copyright Mubarak Shah 2003
Deciduous Trees

Copyright  Mubarak Shah 2003
<table>
<thead>
<tr>
<th>Animals</th>
<th>Sky</th>
<th>Grass</th>
<th>Trees</th>
<th>Rock</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="image4.png" alt="Image" /></td>
<td><img src="image5.png" alt="Image" /></td>
</tr>
<tr>
<td><img src="image6.png" alt="Image" /></td>
<td><img src="image7.png" alt="Image" /></td>
<td><img src="image8.png" alt="Image" /></td>
<td><img src="image9.png" alt="Image" /></td>
<td><img src="image10.png" alt="Image" /></td>
</tr>
<tr>
<td><img src="image11.png" alt="Image" /></td>
<td><img src="image12.png" alt="Image" /></td>
<td><img src="image13.png" alt="Image" /></td>
<td><img src="image14.png" alt="Image" /></td>
<td><img src="image15.png" alt="Image" /></td>
</tr>
<tr>
<td><img src="image16.png" alt="Image" /></td>
<td><img src="image17.png" alt="Image" /></td>
<td><img src="image18.png" alt="Image" /></td>
<td><img src="image19.png" alt="Image" /></td>
<td><img src="image20.png" alt="Image" /></td>
</tr>
</tbody>
</table>

Copyright Mubarak Shah 2003
Motion-blob Verification

- detected events
  - event inference
  - shot summarization
  - motion-blob verification
    - object classification
      - color/texture analysis
        - video sequences
      - spatio-temporal analysis
      - motion estimation
        - shot-boundary detection
Motion Estimation

- three parameter system: x-, y-translation, and zoom,
- 4 motion estimates based on pyramid,
- 4 motion estimates based on previous best match,
- “texture” measure prevents ambiguous matches

Copyright  Mubarak Shah 2003
Motion-blob detection

Motion estimate
\[ \Delta x = -7 \]
\[ \Delta y = 0 \]
zoom = 1.0

Copyright  Mubarak Shah 2003
Shot Summarization

- detected events
- event inference
- shot summarization
- motion-blob verification
- object classification
- color/texture analysis
- spatio-temporal analysis
- motion estimation
- shot-boundary detection

video sequences
Shot Detection

Characteristics of shot boundaries:

- Change of camera/viewpoint
- Change of color characteristics

4 Bins for Value
4 Bins for Saturation
8 bins for hue

\[
\frac{\sum_{n=0}^{15} \min(f_1(n), f_2(n))}{\min(\sum_{n=0}^{15} f_1(n), \sum_{n=0}^{15} f_2(n))} = 0.79
\]
Shot Summaries

A = Ground Truth
B = Result of Algorithm
A ∩ B = Correct detection

recall = \frac{A \cap B}{A}
precision = \frac{A \cap B}{B}
Event Inference

- detected events
- event inference
- shot summarization
- motion-blob verification
- object classification
- color/texture analysis
- spatio-temporal analysis
- motion estimation
- shot-boundary detection

video sequences
Hunt events

Start

Non Hunt Shot

Tracking Fast Animal

Not Tracking Fast Animal

Beginning of Hunt

Tracking Fast Animal

2nd Hunt Shot

Not Tracking Fast Animal

End of Hunt

Valid Hunt

Not Tracking Fast Animal

Valid Hunt

Valid Hunt
Hunts

Hunt

Non-hunt

Copyright  Mubarak Shah 2003
Hunts

Non-hunt

Hunt

Non-hunt

Copyright Mubarak Shah 2003
## Event Detection

<table>
<thead>
<tr>
<th>Sequence Name</th>
<th>Actual Hunt Frames</th>
<th>Detected Hunt Frames</th>
<th>Precision</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hunt1</td>
<td>305 – 1375</td>
<td>305 – 1375</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Hunt2</td>
<td>2472 – 2696</td>
<td>2472 – 2695</td>
<td>100%</td>
<td>99.6%</td>
</tr>
<tr>
<td>Hunt3</td>
<td>3178 – 3893</td>
<td>3178 – 3856</td>
<td>100%</td>
<td>94.8%</td>
</tr>
<tr>
<td>Hunt4</td>
<td>6363 – 7106</td>
<td>6363 – 7082</td>
<td>100%</td>
<td>96.8%</td>
</tr>
<tr>
<td>Hunt5</td>
<td>9694 – 10303</td>
<td>9694 – 10302</td>
<td>100%</td>
<td>99.8%</td>
</tr>
<tr>
<td>Hunt6</td>
<td>12763 – 14178</td>
<td>12463 – 13389</td>
<td>67.7%</td>
<td>44.2%</td>
</tr>
<tr>
<td>Hunt7</td>
<td>16581 – 17293</td>
<td>16816 – 17298</td>
<td>99.0%</td>
<td>67.0%</td>
</tr>
<tr>
<td>average</td>
<td></td>
<td></td>
<td>95.3%</td>
<td>86.0%</td>
</tr>
</tbody>
</table>

Copyright  Mubarak Shah 2003
Landing Events
Landing Events

1. **Approach**
   - Tracking descending non-sky motion blob
   - Fast horizontal motion

2. **Non-landing**
   - Tracking descending non-sky motion blob
   - Sky below motion blob
   - Violation of context constraints

3. **Deceleration**
   - Tracking non-sky motion blob
   - No/slow horizontal motion
   - No sky below motion blob

4. **Touch-down**
   - Tracking non-sky motion blob
   - Fast horizontal motion

**Start**

---

Copyright Mubarak Shah 2003
Landing Events

Non-landing  Approach

Touch-down  Deceleration  Non-landing

Copyright Mubarak Shah 2003
Landing Events

Non-landing Approach

Touch-down Deceleration Non-landing

Copyright Mubarak Shah 2003
Conclusions

• Many natural objects are easily recognized by their color and texture signatures (shape is often not needed)

• Many events are easily detected and recognized by the classes of the comprising objects and their approximate motions

• The proposed visual event detection is robust to changes in scale, color, shape, occlusion, lighting conditions, view points and distances, and image compression

Copyright  Mubarak Shah 2003
Publications


Copyright  Mubarak Shah 2003
Monitoring Human Behavior

Lecture-16

Copyright  Mubarak Shah 2003
Monitoring Human Behavior


Copyright  Mubarak Shah 2003
Goals of the System

• Recognize human actions in a room for which *prior knowledge* is available.
• Handle multiple people
• Provide a textual description of each action
• Extract “key frames” for each action
Possible Actions

- Enter
- Leave
- Sitting or Standing
- Picking Up Object
- Put Down Object
- .....
Prior Knowledge

• Spatial layout of the scene:
  – Location of entrances and exits
  – Location of objects and some information about how they are use

• Context can then be used to improve recognition and save computation
Layout of Scene 1
Layout of Scene 2
Layout of Scene 4
Major Components

• Skin Detection
• Tracking
• Scene Change Detection
• Action Recognition

Copyright  Mubarak Shah 2003
State Model For Action Recognition

Start

End

Standing

Near Cabinet

Near Terminal

Near Phone

Using Terminal

Opening / Closing Cabinet

Open / Close Cabinet

Sit / 0

Stand / 0

Pick Up Phone

Put Down Phone

Talking on Phone

Hanging Up Phone

Copyright Mubarak Shah 2003
Flow of the System

Skin Detection

Track people and Objects for this Frame

Determine Possible Interactions Between People and Objects

Scene Change Detection

Update States, Output Text, Output Key Frames
Key Frames

• Why get key frames?
  – Key frames take less space to store
  – Key frames take less time to transmit
  – Key frames can be viewed more quickly

• We use heuristics to determine when key frames are taken
  – Some are taken before the action occurs
  – Some are taken after the action occurs
Key Frames

- “Enter” key frames: as the person leaves the entrance/exit area
- “Leave” key frames: as the person enters the entrance/exit area
- “Standing/Sitting” key frames: after the tracking box has stopped moving up or down respectively
- “Open/Close” key frames: when the % of changed pixels stabilizes
Key Frames Sequence 1 (350 frames), Part 1
Key Frames Sequence 1 (350 frames), Part 2
Key Frames Sequence 2 (200 frames)
Key Frames Sequence 3 (200 frames)
Key Frames Sequence 4 (399 frames), Part 1

Copyright  Mubarak Shah 2003
Key Frames Sequence 4 (399 frames), Part 2

Copyright Mubarak Shah 2003