Dynamic Training of Hand Gesture Recognition System

Attila Licsár,
Tamás Szirányi

University of Veszprém,
Hungary
Department of Image Processing and Neurocomputing
Motivations

- Effective human-computer interface
  Camera-projector system

- User-interface controlled by hand gestures
  Hand gesture recognition

- User-independence: any user should work with the system with high recognition rate.
  Interactive gesture training
Camera-Projector System

- Desktop image handled by the operating system
- Computer equipped with dual-head video card
- Settings panel of the application
- Projector image
- Monitor

Front-projected system

- Recognition area
- Information area

User interface
System Overview

Image grabbing

Displaying refreshed image by projector

Background & forearm segmentation

background image updating

Gesture analysing

arm mask

Displayed image modification by Finite State Machine

Camera image

Go to the next slide
**Geometrical Distortion of the Camera Image**

Possible geometric distortions: translation, uniform scaling, shearing, keystoning, non-uniform scaling, bending.

Image warping by second order polynomial equations

\[ x' = a_0 + a_1 \cdot x + a_2 \cdot y + a_3 \cdot x^2 + a_4 \cdot xy + a_5 \cdot y^2 \]

\[ y' = b_0 + b_1 \cdot x + b_2 \cdot y + b_3 \cdot x^2 + b_4 \cdot xy + b_5 \cdot y^2 \]
Segmentation Processes

Problem: hand surface reflects the projected background

- Background subtraction
  - reference background image
  - human skin partly absorbs the light

- Forearm segmentation: boundary-based classification

Camera image  ➔  Segmented arm mask  ➔  Segmented contour
Segmentation with Changing Background

- Changed image to be projected
- Warped projector image
- Color corrected projector image
- Camera image
- Background differencing by artificially generated background
- Segmented image
Automatic Wrist Point Detection Based on Wrist Width

Wrist detection condition:

\[ w_i < w_{i+1} < \ldots < w_{i+20} \]

Width values of the hand

Detected wrist point
<table>
<thead>
<tr>
<th>Original Image to Be Projected</th>
<th>Camera Grabbed Image</th>
<th>Segmented Arm Mask</th>
<th>Segmented Hand Contour</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Image 1" /></td>
<td><img src="image2.png" alt="Image 2" /></td>
<td><img src="image3.png" alt="Image 3" /></td>
<td><img src="image4.png" alt="Image 4" /></td>
</tr>
<tr>
<td><img src="image5.png" alt="Image 5" /></td>
<td><img src="image6.png" alt="Image 6" /></td>
<td><img src="image7.png" alt="Image 7" /></td>
<td><img src="image8.png" alt="Image 8" /></td>
</tr>
<tr>
<td><img src="image9.png" alt="Image 9" /></td>
<td><img src="image10.png" alt="Image 10" /></td>
<td><img src="image11.png" alt="Image 11" /></td>
<td><img src="image12.png" alt="Image 12" /></td>
</tr>
</tbody>
</table>
Fourier Descriptors

\[ F(t) = \sum_{n=0}^{N_B-1} f(t) \cdot e^{-j \frac{2\pi nt}{N_B}} \]

F(t) – Fourier descriptors
f(t) – feature vectors

Distance metric (between two contour: \( F^1(t) \) and \( F^2(t) \)):

\[ Dist(F^1, F^2, n) = \sigma \left( \left| \frac{F^1_1}{F^2_1} \right|, \left| \frac{F^1_2}{F^2_2} \right|, \ldots, \left| \frac{F^1_n}{F^2_n} \right| \right) + \sigma \left( \left| \frac{F^2_1}{F^1_1} \right|, \left| \frac{F^2_2}{F^1_2} \right|, \ldots, \left| \frac{F^2_n}{F^1_n} \right| \right) \]

F – Fourier descriptors  \( \sigma \) – standard deviation  \( n \) - cut-off frequency

Metric is invariant to translation, rotation and scaling of feature vectors

Gesture class is estimated by the nearest neighbor rule:

\[ G = \arg \min_i \left( Dist(F_a, F_i) \right) \]

\( F_a \) – actual classes
\( F_i \) – stored gesture classes
G – detected gesture class
Interactive Gesture Training

- Traditional training method:
  - 2 phases: preliminary training then recognition

- Interactive training method:
  - there is no separated training phase it is embedded into the recognition phase
  - users could correct only the false recognized gestures by interaction with the training method
  - prototype gesture classes are dynamically corrected by the actual users if it is necessary

System should be used by any users with high gesture recognition rates
Interactive Training Method

Training controlled by user interaction

Gestures detection

No user feedback: correct detection

Displaying recognized gesture

Unsupervised training

Unsupervised training

User feedback: false detection

System selects next probable gesture

Supervised training

Retraining of the selected gesture

Second user feedback: end of the supervised training

Refreshing displayed gesture parameters
Supervised Training

Detection is not correct: user feedback

Camera input

Detected gestures

User feedback

Estimated order:

Training selected gesture parameters

Correct gesture estimation

Selected gesture

Trainer gesture
Interactive Training Method

Gesture detection

No user feedback: correct detection

Displaying recognized gesture

Refreshing displayed gesture parameters

Unsupervised training

User feedback: false detection

System selects next probable gesture

Retraining of the selected gesture

Supervised training

Second user feedback: end of the supervised training

User feedback: correct detection

Displaying recognized gesture

Unsupervised training

No user feedback: correct detection

Refreshing displayed gesture parameters

User feedback: false detection

System selects next probable gesture

Retraining of the selected gesture

Supervised training

Second user feedback: end of the supervised training
Unsupervised Training

Detection is correct: no user feedback

Gestures from camera

Detected gestures class

Trainer gesture (Y)

Trained gesture (x)

Refreshing displayed gesture parameters

\[ |F_n^Z| = \left( \frac{N - 1}{N} \cdot |F_n^X| + \frac{1}{N} \cdot |F_n^Y| \right) \]
Recognition Results with Several Trainer and Tester Users

Results without interactive training method

Columns = trainer users: their gestures are used for the training
Rows = tester users: their gestures are used for the recognition

<table>
<thead>
<tr>
<th></th>
<th>User A</th>
<th>User B</th>
<th>User C</th>
<th>User D</th>
</tr>
</thead>
<tbody>
<tr>
<td>User A</td>
<td>99.8</td>
<td>96.1</td>
<td>86.2</td>
<td>94.6</td>
</tr>
<tr>
<td>User B</td>
<td>90.4</td>
<td>97.6</td>
<td>90.1</td>
<td>93.9</td>
</tr>
<tr>
<td>User C</td>
<td>94.5</td>
<td>92.7</td>
<td>99.6</td>
<td>98.9</td>
</tr>
<tr>
<td>User D</td>
<td>95.5</td>
<td>95</td>
<td>96.7</td>
<td>99.1</td>
</tr>
</tbody>
</table>

Test set: 9 gesture classes and 720 test images per person

If the trainer and the tester users are different the recognition rate could decrease
# Dynamic Gesture Training by User-Interaction

<table>
<thead>
<tr>
<th>User order</th>
<th>Without interactive training</th>
<th>After interactive training</th>
<th>Number of retrained gestures</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. User A</td>
<td>99.5</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2. User B</td>
<td>89.8</td>
<td>96.1</td>
<td>4</td>
</tr>
<tr>
<td>3. User C</td>
<td>94.7</td>
<td>97</td>
<td>2</td>
</tr>
<tr>
<td>4. User D</td>
<td>99.1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5. User B</td>
<td>93</td>
<td>98.8</td>
<td>1</td>
</tr>
<tr>
<td>6. User C</td>
<td>96.8</td>
<td>99.2</td>
<td>2</td>
</tr>
</tbody>
</table>

**Test set:** 9 gesture classes
Gesture Recognition Demo

Camera input

Segmented arm mask

Pictogram of the detected gesture

Segmented palm contour
Conclusion

- Interactive training is user-friendly and user-independent;
- The training is embedded into the recognition phase;
- Supervised training is possible to correct the possible overlap among the different classes;
- We have tested the supervised training system with several users and found that the performance of recognition has increased significantly without terminating the application and retraining all gestures in a preliminary phase.
Thank you very much for your kind attention!
Forearm Segmentation

w_1, w_2, w_3, w_4, w_5

wrist points

main axis
# Gesture Classification Results with Several Features and Users

<table>
<thead>
<tr>
<th>Users</th>
<th>Method I.</th>
<th>Method II.</th>
<th>Method III.</th>
<th>Method IV.</th>
</tr>
</thead>
<tbody>
<tr>
<td>User A.</td>
<td>99.6</td>
<td>96.8</td>
<td>98.1</td>
<td>97</td>
</tr>
<tr>
<td>User B.</td>
<td>98.3</td>
<td>92.6</td>
<td>96.7</td>
<td>98.1</td>
</tr>
<tr>
<td>User C.</td>
<td>97.9</td>
<td>95.2</td>
<td>95.3</td>
<td>91.2</td>
</tr>
<tr>
<td>User D.</td>
<td>98.2</td>
<td>91.2</td>
<td>94.5</td>
<td>92.3</td>
</tr>
</tbody>
</table>

9 gesture classes, 600 gesture samples per user
Recognition efficiency by several cut-off frequencies

- Complex sequence - Method A
- Centroid distance - Method C
- Centroid distance - Method D
- Complex sequence - Method B

Method I.
Method II.
Method III.
Method IV.
. Test results with only preliminary training

<table>
<thead>
<tr>
<th></th>
<th>Recognition results [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Trainer users</td>
</tr>
<tr>
<td></td>
<td>User A</td>
</tr>
<tr>
<td>Tester</td>
<td>Mean</td>
</tr>
<tr>
<td>User A</td>
<td>99.3</td>
</tr>
<tr>
<td></td>
<td>+0.3</td>
</tr>
<tr>
<td>User B</td>
<td>99</td>
</tr>
<tr>
<td></td>
<td>+0.3</td>
</tr>
<tr>
<td>User C</td>
<td>87.7</td>
</tr>
<tr>
<td></td>
<td>+5.3</td>
</tr>
<tr>
<td>User D</td>
<td>94.9</td>
</tr>
<tr>
<td></td>
<td>+1.8</td>
</tr>
</tbody>
</table>
Controlling of Multimedia Presentations

- Go to the next/previous slide
- Go to the last/first slide
- Start presentation
- End presentation
- Draw line to the projected image to select a region

Hand signs
Advantages of Modified Fourier Descriptor

- Robustness to transformation – invariant to translation, rotation and scaling
- Robustness to noise
- Feature extraction efficiency
- Feature matching efficiency – the distance measure requires very low computational cost
- Feature vectors can be calculated efficiently compared to other image moment-based methods
Gestures

Dynamic gestures

Static gestures
Gesture Models

Spatial Gesture Model

3D Hand Model-Based

Parameters:
• joint angles
• palm position
• etc.

Appearance-Based

Posture-Based

Static gestures
• images
• image boundary
• etc.

Motion-Based

Dynamic gestures
• images motion
parameters
• fingertip position
& motion
Hand Signs for the Application

- Tracking
- Select
- Defect
- Good
- Cut
- Reference
- Start play
- Stop
- Back
Example: Selecting ROI in Sample Images

- Tracking: positioning and tracking the cursor
- Select: drawing a continuous line
Application for the old film restoration
Curvature-based Forearm Segmentation

Angles of the contour

Angles

Points of the contour
System Calibration

Geometric calibration

Color calibration

Geometric transformation function

Color transformation function
Forearm segmentation based on curvature analysis

Angles of the contour

Detected wrist points
Measured probability distribution of the false classified gestures

![Graph showing probability distribution](chart.png)
Vision-Based Gesture Recognition System

Model parameter space and classes