APPENDIX A. DATABASES

1. Databases for Face Recognition

AR FACE DATABASE

1. Type: image and sequences
2. Description: This face database was created by Aleix Martinez and Robert Benavente in the Computer Vision Center (CVC) at the U.A.B. It contains over 4,000 color images corresponding to 126 people’s faces (70 men and 56 women). Images feature frontal view faces with different facial expressions, illumination conditions, and occlusions (sun glasses and scarf). The pictures were taken at the CVC under strictly controlled conditions. No restrictions on wear (clothes, glasses, etc.), make-up, hair style, etc. were imposed to participants. Each person participated in two sessions, separated by two weeks (14 days) time. The same pictures were taken in both sessions.
3. Number of images: over 4,000
4. Conditions:

<table>
<thead>
<tr>
<th>Condition</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>number of subjects</td>
<td>126 (70/56)</td>
</tr>
<tr>
<td>number of poses</td>
<td>1</td>
</tr>
<tr>
<td>number of expressions</td>
<td>4</td>
</tr>
<tr>
<td>number of illuminations</td>
<td>4</td>
</tr>
<tr>
<td>number of sessions</td>
<td>2</td>
</tr>
<tr>
<td>resolution</td>
<td>768 x 576</td>
</tr>
<tr>
<td>simultaneous</td>
<td>X</td>
</tr>
</tbody>
</table>
5. URL: http://cobweb.ecn.purdue.edu/~aleix/aleix_face_DB.html
7. Sample images (see Fig. A.1)

BANCA FACE DATABASE

1. Type: image and video
2. Description: The BANCA database is a new large, realistic and challenging multi-modal database intended for training and testing multi-modal verification systems. The BANCA database was captured in four European languages in two modalities (face and voice). For recording, both high and low quality microphones and cameras were used. The subjects were recorded in three different scenarios, controlled, degraded and adverse over 12 different sessions spanning three months. In total 208 people were captured, half men and half women.
3. Number of images: over 6,240
4. Conditions:
   - number of subjects: 52 (26/26)
   - number of poses: 1
   - number of expressions: 2
   - number of illuminations: -
   - number of sessions: 12
   - resolution: 720 x 576
   - simultaneous: X

Figure A.1. Example of how the images of the AR face database look like
CAS-PEAL DATABASE

1. Type: image
2. Description: The CAS-PEAL face database has been constructed under the sponsors of National Hi-Tech Program and ISVISION by the Face Recognition Group of JDL, ICT, CAS. The goals to create the PEAL face database include: providing the worldwide researchers of FR community a large-scale Chinese face database for training and evaluating their algorithms; facilitating the development of FR by providing large-scale face images with different sources of variations, especially Pose, Expression, Accessories, and Lighting (PEAL); advancing the state-of-the-art face recognition technologies aiming at practical applications especially for the oriental. The CAS-PEAL face database contains 99,594 images of 1040 individuals (595 males and 445 females) with varying Pose, Expression, Accessory, and Lighting (PEAL). For each subject, 9 cameras spaced equally in a horizontal semicircular shelf are setup to simultaneously capture images across different poses in one shot. Each subject is also asked to look up and down to capture 18 images in another two shots. We also considered 5 kinds of expressions, 6 kinds accessories (3 glasses, and 3 caps), and 15 lighting directions, as well as varying backgrounds, distance from cameras, and aging variation.
3. Number of images: 99,594
4. Conditions:
   - number of subjects: 1040 (595/445)
   - number of poses: 27
   - number of expressions: 6
   - number of illuminations: 15
   - number of sessions: 2
   - resolution: 360 x 480
   - simultaneous: partially
5. URL: http://www.jdl.ac.cn/peal/index.html
7. Sample images (see Fig. A.3)

CMU PIE DATABASE

1. Type: image
2. Description: Between October and December 2000, they collected a database of 41,368 images of 68 people. By extending the CMU 3D Room we were able to image each person under 13 different poses, 43 different illumination conditions, and with
4 different expressions. They call this database the CMU Pose, Illumination, and Expression (PIE) database.

3. Number of images: 41,368

4. Conditions:
   - number of subjects: 68
   - number of poses: 13
   - number of expressions: 3
   - number of illuminations: 43
   - number of sessions: 1
   - resolution: 640 x 486
   - simultaneous: partially

5. URL: http://www.ri.cmu.edu/projects/project_418.html


FERET DATABASE

1. Type: image

2. Description: The FERET program ran from 1993 through 1997. Sponsored by the Department of Defense's Counterdrug Technology Development Program through the Defense Advanced Research Products Agency (DARPA), its primary mission was to develop automatic face recognition capabilities that could be employed to assist security, intelligence and law enforcement personnel in the performance of their duties. The FERET image corpus was assembled to support government monitored testing

Figure A.3. The 27 images of one subject under pose variation in the CAS-PEAL database. The nine cameras were spaced equally in the horizontal semicircular shelf, each about 22.5° apart. The subject was asked to look upwards, right into the camera C4 (the middle camera) and look downwards. Then, the 27 poses were named after the subject's pose (Up, Frontal, Down) and the number of the corresponding camera (from 0 to 8). The name of each pose was beneath its corresponding image.
Appendix

and evaluation of face recognition algorithms using standardized tests and procedures. The final corpus, presented here, consists of 14051 eight-bit grayscale images of human heads with views ranging from frontal to left and right profiles.

3. Number of images: 14,051
4. Conditions:
   - number of subjects: 1199
   - number of poses: -
   - number of expressions: 2
   - number of illuminations: 2
   - number of sessions: 3
   - resolution: 256 x 384
   - simultaneous: X

5. URL: http://www.itl.nist.gov/iad/humanid/feret/feret_master.html
7. Sample images (see Fig. A.5)

KOREAN FACE DATABASE

1. Type: image
2. Description: They collected a database of 52,000 images of 1,000 people. Two colors and eight directions are considered in the illumination conditions. Five kinds of expressions - neutral, happiness, surprise, anger and blink expressions - are considered under two different illumination colors. Seven poses are considered with or without hair-band or glasses.
3. Number of images: 52,000
4. Conditions:
   - number of subjects: 1000(500/500)
   - number of poses: 7
   - number of expressions: 5
   - number of illuminations: 16
   - number of sessions: 1
   - resolution: 640 x 480
   - simultaneous: partially

5. URL: -
7. Sample images (see Fig. A.6)
ORL DATABASE (AT&T)

1. Type: image
2. Description: The ORL Database of Faces contains a set of face images taken between April 1992 and April 1994 at the lab. The database was used in the context of a face recognition project carried out in collaboration with the Speech, Vision and Robotics Group of the Cambridge University Engineering Department. There are ten different images of each of 40 distinct subjects. For some subjects, the images were taken at different times, varying the lighting, facial expressions (open / closed eyes, smiling / not smiling) and facial details (glasses / no glasses). All the images were taken against a dark homogeneous background with the subjects in an upright, frontal position (with tolerance for some side movement). A preview image of the Database of Faces is available. The files are in PGM format, and can conveniently be viewed on UNIX (TM) systems using the ‘xv’ program. The size of each image is 92x112 pixels, with 256 grey levels per pixel. The images are organized in 40 directories (one for each subject), which have names of the form sX, where X indicates the subject number (between 1 and 40). In each of these directories, there are ten different images of that subject, which have names of the form Y.pgm, where Y is the image number for that subject (between 1 and 10).
3. Number of images: 400
4. **Conditions:**
   - number of subjects: 40
   - number of poses: -
   - number of expressions: -
   - number of illuminations: -
   - number of sessions: 1
   - resolution: 92 x 110
   - simultaneous: X

5. **URL:** http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html


7. **Sample images (see Fig. A.7)**
POSTECH FACE DATABASE (01 version)

1. Type: image
2. Description: We constructed a face database PF01(POSTECH Faces '01). PF01 contains the true-color face images of 103 people, 53 men and 50 women, representing 17 various images (1 normal face, 4 illumination variations, 8 pose variations, 4 expression variations) per person. All of the people in the database are Asians. There are three kinds of systematic variations, such as illumination, pose, and expression variations in the database. The database is expected to be used to evaluate the technology of face recognition for Asian people or for people with systematic variations.

3. Number of images: 1,751
4. Conditions:
   - number of subjects: 103 (53/50)
   - number of poses: 8
   - number of expressions: 4
   - number of illuminations: 5
   - number of sessions: 1
   - resolution: 1280 x 960
   - simultaneous: X

5. URL: http://imlab.postech.ac.kr/special/imdb/imdb.html
6. Paper: -
7. Sample images (see Fig. A.8)

Figure A.7. Ten different images per person. The images were taken at different times, varying the lighting, facial expressions (open / closed eyes, smiling / not smiling) and facial details (glasses / no glasses).
POSTECH FACE DATABASE (07 version)

1. Type: image
2. Description: PF07 database includes 100 male and 100 female subjects captured in 5 different poses under 16 illuminations performing 4 different expressions. The pose variation consists of front, left, right, upper, and down, and the angle between the frontal pose and other poses is 22.5°. The illumination variation consists of no light condition and 15 different light conditions, where each light condition means the turn-on of the light on a specific location, and 15 locations are the intersection points of three vertical positions (high, middle, and low) and five horizontal positions (-90°; -45°; 0°; 45°; 90°). The expression variation consists of neutral, happy, surprise, and angry. Therefore, there are 5 x 4 x 16 = 320 images for a subject. Since 200 subjects are exist in this database, the database contains 64000 images.
3. Number of images: 64,000
4. Conditions:
   - number of subjects: 200 (100/100)
   - number of poses: 5
   - number of expressions: 4
   - number of illuminations: 16
   - number of sessions: 1
   - resolution: 640 x 480
   - simultaneous: Y

5. URL: http://imlab.postech.ac.kr/~dkim/new_imlab/faceDB/PF07/PF07.html
6. Paper: -
7. Sample images (see Fig. A.9)

XM2VTS Face Database

1. Type: image and video
2. Description: At the Centre for Vision, Speech and Signal Processing we have captured a large multi-modal database which will enable the research community to test their multi-modal face verification algorithms on a high-quality large dataset. In acquiring the XM2FDB database 295 volunteers from the University of Surrey visited our recording studio four times at approximately one month intervals. On each visit (session) two recordings (shots) were made. The first shot comprised of speech whilst the second
consisted of rotating head movements. Digital video equipment was used to capture the entire database. At the third session a high-precision 3D model of the subjects' head was built using an active stereo system provided by the Turing Institute.

3. Number of images: -

4. Conditions:
   - number of subjects: 295
   - number of poses: -
   - number of expressions: -
   - number of illuminations: 3
   - number of sessions: 4
   - resolution: 720 x 576
   - simultaneous: X

5. URL: http://www.ee.surrey.ac.uk/CVSSP/xm2vtsdb/


7. Sample images (see Fig. A.10)

**YALE FACE DATABASE - A**

1. Type: image
2. Description: The Yale Face Database (size 6.4MB) contains 165 grayscale images in GIF format of 15 individuals. There are 11 images per subject, one per different facial expression.

*Figure A.9. Images of happy facial expression taken under 16 different light conditions*
expression or configuration: center-light, w/glasses, happy, left-light, w/no glasses, normal, right-light, sad, sleepy, surprised, and wink.

3. Number of images: 165
4. Conditions:
   - number of subjects: 15
   - number of poses: 1
   - number of expressions: 6
   - number of illuminations: 3
   - number of sessions: 1
   - resolution: 320 x 240
   - simultaneous: X

5. URL: http://cvc.yale.edu/projects/yalefaces/yalefaces.html/
7. Sample images (see Fig. A.11)

Figure A.11. Images of 6 different facial expression taken under 3 different illumination conditions

Copyright © 2009, IGI Global, distributing in print or electronic forms without written permission of IGI Global is prohibited.
YALE FACE DATABASE - B

1. Type: image
2. Description: The database contains 5760 single light source images of 10 subjects each seen under 576 viewing conditions (9 poses x 64 illumination conditions). For every subject in a particular pose, an image with ambient (background) illumination was also captured. Hence, the total number of images is in fact 5760+90=5850. The total size of the compressed database is about 1GB.
3. Number of images: 5,850
4. Conditions:
   - number of subjects: 10
   - number of poses: 9
   - number of expressions: 1
   - number of illuminations: 65
   - number of sessions: 1
   - resolution: 640 x 480
   - simultaneous: Y
5. URL: http://cvc.yale.edu/projects/yalefacesB/yalefacesB.html
7. Sample images (see Fig. A.12)

BJUT (Beijing University of Technology)-3D FACE DATABASE

1. Type: image
2. Description:
3. Number of images: 500
4. Conditions:

Figure A.12. Images of 10 subjects taken under frontal-view, neutral facial expression
Appendix

age 16~49
male / female 250 / 250
expressions Natural without glasses or other accessories

5. URL: http://bjut.edu.cn/sci/multimedia/mul-lab/3dface/face_database.htm
6. Paper: The BJUT-3D Large-Scale Chinese Face Database, MISKL-TR-05-FMFR-001
7. Sample images (see Fig. A.13)

2. Databases for Facial Expression Recognition

MMI FACE DATABASE Type: Image and Video
1. Description:
2. Number of images: 740 images, 848 video sequences (24fps)
3. Conditions:
   number of peoples 19 (male, female, 3 ethnics)
   views frontal view and dual-view (combine frontal and profile view of the face)
   expressions facial expression of motion

4. URL: http://www.mmifacedb.com
6. Sample images (see Fig. A.14)

CMU-PITTSBURGH AU-CODED FACE EXPRESSION IMAGE DATABASE
1. Type: image
2. Description:
3. Number of images: frontal view 2105 frames, 30-degree view (videotape only)

Figure A.13. Examples of BJUT database. Each column represents 3D scanner, 3D face, shape data, texture image, from left to right.

Copyright © 2009, IGI Global, distributing in print or electronic forms without written permission of IGI Global is prohibited.
4. Conditions:
   - number of peoples: 182 (male, female)
   - age: 18~50
   - number of expressions: 23 (varying skin color and facial conformation)

5. URL: http://vasc.ri.cmu.edu/idb/html/face/facial_expression/index.html


7. Sample images: (see Fig. A.15)

**JAPANESE FEMALE FACIAL EXPRESSION (JAFFE) DATABASE**

1. Type: image
2. Description:
3. Number of images: 213
4. Conditions:
   - Number of peoples: 10 (female)
   - Number of expressions: 7 (6 basic facial expression + 1 neutral)

5. URL: http://www.kasrl.org/jaffe.html

7. Sample images (see Fig. A.16)

BU-3DFE (Binghamton University 3D Facial Expression) DATABASE

1. Type: image
2. Description:
3. Number of images: 100 subjects with 2500 3D facial expression models
4. Conditions:
   age 18–70
   ethnic / racial White, Black / East-Asian, Middle-east Asian, Indian, and Hispanic Latino
   number of expressions 7 (neutral + happiness, disgust, fear, angry, surprise and sadness)

Figure A.16. Images of one subject making 5 different facial expression
5. URL: http://www.cs.binghamton.edu/~lijun/Research/3DFE/3DFE_Analysis.html  
7. Sample images (see Fig. A.17 and Fig. A.18)

3. Databases for Hand Gesture Recognition

MASSEY HAND GESTURE (MHG) DATABASE

1. Type: image  
2. Description: MHG database (2D) contains a variety of hand gesture and hand posture images which is for real-time posture recognition. The data is collected by a digital camera mounted on a tripod from a hand gesture in front of a dark background, and in different lighting environments. Together with the original images, there is a clipped version of each set of images that contains only the hand image. Some of images are consist of black background so that it is quite convenient to detect shape, color of hand.

Figure A.17. Four levels of facial expression from low to high

![Figure A.17](image)

Figure A.18. Seven expressions female with face images and facial models

![Figure A.18](image)
3. Number of images: 15000

4. Conditions

<table>
<thead>
<tr>
<th>Number</th>
<th>Dataset</th>
<th>Lighting Condition</th>
<th>Background</th>
<th>Size</th>
<th>Type</th>
<th>Number of files</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Hand gesture</td>
<td>Normal</td>
<td>Dark background</td>
<td>640x480</td>
<td>Jpeg</td>
<td>169</td>
</tr>
<tr>
<td>2</td>
<td>Hand gesture</td>
<td>Normal</td>
<td>RGB(0,0,0)</td>
<td></td>
<td>Varying/Clipped</td>
<td>169</td>
</tr>
<tr>
<td>3</td>
<td>Hand palm</td>
<td>Normal</td>
<td>Dark background</td>
<td>640x480</td>
<td>Jpeg</td>
<td>145</td>
</tr>
<tr>
<td>4</td>
<td>Hand palm</td>
<td>Normal</td>
<td>RGB(0,0,0)</td>
<td></td>
<td>Varying/Clipped</td>
<td>145</td>
</tr>
<tr>
<td>5</td>
<td>Hand palm</td>
<td>Artificial light/dark room</td>
<td>Dark background</td>
<td>640x480</td>
<td>Jpeg</td>
<td>498</td>
</tr>
<tr>
<td>6</td>
<td>Hand palm</td>
<td>Artificial light/dark room</td>
<td>Dark background</td>
<td>Varying/Clipped</td>
<td>jpeg</td>
<td>498</td>
</tr>
</tbody>
</table>


7. Sample images (see Fig. A.19)

SÉBASTIEN MARCEL – INTERACTPLAY (SMIP) DATABASE

1. Type: Movie
2. Description: The SMIP database made of a 3D hand trajectories so that one can study hand tracking in 3-dimension space. It will be also possible to recognize gestures by tracking the trajectory of the hand. It will be quite suitable for anyone who want to research tracking a hand- trajectory in 3 dimensional space
3. Size of movie: 21 Mb

Figure A.19. Examples of the hand postures of the Massey Hand Gesture Database
4. Conditions
The SMIP database contains 16 hand gestures from 22 persons and provides 5 sessions and 10 recordings per session. The database contains 3D trajectories, including the coordinates of the head and the torso, of segmented gestures. It provides also a calibration sequence for every person and every session, and test sequences of continuous gestures. Each trajectory is stored in one text file. The naming convention is the following \([\text{personid} \_ \text{gestureid} \_ \text{sessionid} \_ \text{recordid}]\). For instance, the file 01_07_5_9.txt is the record 9 of the gesture 7 performed by person 1 during the session 5. For every session and every person, two additional sequences are provided: a calibration sequence called "Vinci" sequence, and a test sequence made of continuous gestures. Test sequences are labelled and will be available soon. The calibration sequence is stored as the gesture 17 and the test sequence as the gesture 18.

5. URL: http://www.idiap.ch/resources/interactplay/
7. Sample images (see Fig. A.20)

**POINTING GESTURE: VIDEO SEQUENCE (PGVS) DATABASE**

1. Type: Movie
2. Description: The PGVS database consists of 24 video sequences of hands pointing onto a desk. It will be quite useful providing that you want to recognize a spot which

*Figure A.20. One of the hand gesture sequences. Database contains 16 hand gestures form 22 persons.*
anyone points out. The sequences are provided as a set of PNG format images with lose-less compression. The .png files are named sequenceType_personID_skinType_sex_frameNumber.png. The sequence Types are described above, the person ID is a number from 01 to 15, skin Type is either (White, Yellow, Black), the sex is either (Male, Female), and frame Number is a four digit number (0000 >> XXXX).

3. Size of movie: 7 GB

4. Conditions
The scene background is a table covered with a black piece of fabric, a Macbeth color-checker, and two small pieces of light gray rubber. The rubber pieces were placed in the scene to give the test persons some objects to interact with. Doing recordings the light in the scene is switched forth and back between four different light sources, which is described in more details below. The scene background is a messy table with a lot of moveable objects, and a Macbeth color-checker. The test persons were asked to interact with as many objects as they felt like, but not to start reading any of the books. The scene background is a messy table with a lot of moveable objects, and a Macbeth color-checker. The test persons were asked to interact with as many objects as they felt like, but not to start reading any of the books. Doing recordings the light in the scene is switched forth and back between four different light sources.

Lighting

<table>
<thead>
<tr>
<th>Light 1</th>
<th>Philips : TLD 58W/965 (6200K)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Light 2</td>
<td>Philips : TLD 58W/950 (4700K)</td>
</tr>
<tr>
<td>Light 3</td>
<td>Philips : TLD 58W/940 (3680K)</td>
</tr>
<tr>
<td>Light 4</td>
<td>Philips : TLD 58W/927 (2600K)</td>
</tr>
</tbody>
</table>

5. URL: http://www.cvmt.dk/~fgnet/Pointing04/

6. Paper:
It was originally recorded as part of the Pointing’04 ICPR Workshop, Cambridge, UK, 22th, Aug, 2004

7. Sample images (see Fig. A.21)

Figure A.21. X: The scene background is a messy table with a lot of moveable objects, and a Macbeth color-checker. Doing recordings the light in the scene was fixed to a specific light type; Y: The scene background is a messy table with a lot of moveable objects, and a Macbeth color-checker. Doing recordings the light in the scene is switched forth and back between four different light sources; Z: The scene background is a table covered with a black piece of fabric, a Macbeth color-checker, and two small pieces of light gray rubber.
4. Databases for Head Gesture Recognition

HEAD POSE (HP) IMAGE DATABASE

1. Type: image
2. Description: HP database proposes a benchmark from head pose estimation system working on static images of known and unknown faces. All images have been taken using the FAME Platform of the PRIMA Team in INRIA Rhone-Alpes. To obtain different poses, we have put markers in the whole room. Each marker corresponds to a pose (h,v). Post-it are used as markers. The whole set of post-it covers a half-sphere in front of the person. In order to obtain the face in the center of the image, the person is asked to adjust the chair to see the device in front of him. After this initialization phase, we ask the person to stare successively at 93 post-it, without moving his eyes. This second phase just takes a few minutes. All images are obtained by using this method (see Fig. A.22 and A.23)
3. Number of images: 2790 images, 6 movies
4. Conditions: It consists in 15 sets of images. Each set contains of 2 series of 93 images of the same person at 93 different poses. There are 15 people in the database, wearing glasses or not and having various skin color. The pose, or head orientation is determined by 2 angles (h,v), which varies from -90 degrees to +90 degrees.
5. URL: http://www-prima.inrialpes.fr/perso/Gourier/Faces/HPDatabase.html
7. Sample image (see Fig. A.24)

Figure A.22. Top sight
5. Databases for Body Gesture Recognition

CMU MOCAP DATABASE

1. Type: MPEG, c3d, amc
2. Description: MOCAP database is for analyzing the human activities. This DB used a system based on an optical marker-based technology, which yields very clean and detailed motion capture data. Here the actor is equipped with a set of 40-50 retro-reflective markers attached to a suit. There markers are tracked by an array of six to twelve calibrated high-resolution cameras at a frame rate of up to 240 Hz. From the recorded 2D images of the marker positions, the system can then reconstruct the 3D marker positions with high precision. The resulting 3D trajectory data is stored in the C3D mobcap file format.
3. Number of images:

Figure A.23. Side sight

Figure A.24. A sample of a serie, each serie contains 93 images of the same person at 93 different poses
4. Conditions: 144 subjects (various trials per each subject)
   There are 2605 trials in 6 categories and 23 subcategories.

<table>
<thead>
<tr>
<th>Subjects</th>
<th>Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human Interaction</td>
<td>Playground, uneven terrain, path with obstacles</td>
</tr>
<tr>
<td>Interaction with Environment</td>
<td>Running, walking, jumping, varied</td>
</tr>
<tr>
<td>Locomotion</td>
<td>Basketball, dance, gymnastics, acrobatics, martial arts, racquet/paddle sports, soccer, boxing, general exercise and stretching, golf, Frisbee</td>
</tr>
<tr>
<td>Physical Activities &amp; Sports</td>
<td>Common behaviors and expressions, pantomime, communication gestures and signals, cross-category variety</td>
</tr>
<tr>
<td>Situations &amp; Scenarios</td>
<td></td>
</tr>
<tr>
<td>Test Motions</td>
<td></td>
</tr>
</tbody>
</table>

5. URL: http://mocap.cs.cmu.edu
7. Sample image (see Fig. A.25)

**KOREA UNIVERSITY FULL-BODY GESTURE (FBG) DATABASE**

1. Type: Image, HTR, AVI
2. Description: KUG database presents the 2D and 3D Full-Body Gesture database for analyzing 2D and 3D gesture and its related studies. This database contains 14 representative gestures in daily life for 10 male and female subjects of 60~80 ages. This database consists of major three parts (1) 3D motion data (2) 2D stereo-video data (3) 2D silhouette data.
   A. 3D Gesture capture system
      It exploits the Eagle Digital System of Motion Analysis Co. The Eagle Digital System consists of Eagle Digital Cameras, the EagleHub, EVaRT software, which can capture subject’s motion with high accuracy. The motion capture

*Figure A.25. A jumping motion and its animated motion result*
camera, Eagle Digital Camera supports a resolution of 1280x1024 pixels at up to 500 frames per second. It totally positioned the 12 cameras. All subjects wear a black and blue color suit, on which 33 markers reflecting light from LED of 3D cameras are attached. All 3D cameras are synchronized and the 3D position of markers is obtained at 60 frames per second. 3D data motion data is saved in Motion Analysis ‘HTR (Hierarchical Translation Rotation)’ format.

B. 2D Gesture capture system

It captured 2D and 3D gesture data, simultaneously. 2D Video data is captured with stereo camera system (STH-MD-CS2) made by Videre Design. 2D stereo camera systems are 4m away from a subject and placed at +45, -45, 0 degrees for obtaining gestures at 3 different directions. It captures uncompressed video at 320x240 resolution, color and 30 frames per second and saved in uncompressed ‘AVI’ file format.

3. Number of images: 400 MB for 3D data, 90GB for 2D data.
4. Conditions:

14 Gestures

(1) sitting on a chair
(2) standing up from a chair
(3) walking at a place
(4) touching a knee and a waist
(5) raising a right hand
(6) sticking out a hand
(7) bending a waist
(8) sitting on the floor
(9) getting down on the floor
(10) lying down on the floor
(11) waving a hand
(12) running at a place
(13) walking forward
(14) walking circularly

5. URL: http://gesturedb.korea.ac.kr
7. Sample image (see Fig. A.26)

APPENDIX B. DEMONSTRATION SYSTEMS

1. Face Detection and Recognition Systems

VISION ACCESS SDK

1. Author: Bioscrypt
2. Sales: Commercial
3. Specification: The SDK tools can be used for development 3D face recognition system. Two versions SDK components are available, BioAPI®-based components and ActiveX® components. From door to desk top computer, many applications are useful.
4. WWW: http://www.bioscrypt.com/
5. Environment:
   A. Intel Pentium® 4-based PC 2.0+ GHz RAM 512 MB, 60GB HDD
   B. Microsoft® Windows 2000™ and XP Professional™,
   C. DirectX®9.0
   D. Off-the-shelf video capture cards
   E. Microsoft Visual Studio™.NET 2003, C++, C#

6. System performance: Perform facial recognition calculations at processing rates of 10 - 12 full capturing- matching cycles per second

7. Available anywhere from 3-6 feet away

8. Demo image (see Fig. B.1)

**VISION 3D/2D ENROLLMENT APPLICATION**

1. Author: A4vision
2. Sales: Commercial
3. Specification: To perform enrollment of the 3D (optionally 2D) face image. That system is used for a verity of surveillance system such as ePassport programs or

---

*Figure A.26. 3D motion examples of 14 gestures*
traveler initiatives. The enrollment process is simple. And the system is robust to lighting flexibility

4. WWW: http://www.a4vision.com/
5. Environment:
   A. Intel Pentium® 4 PC 3 GHz or faster, dual processor is supported
   B. RAM 512 MB 1GB preferred, 120MB HDD
   C. Microsoft Windows 2000 Professional / XP Professional
   D. Compatible with ODBC compliant database
6. System performance
   A. Verification time: less then 1second
   B. Identification time: less then 1second
   C. Enrollment time: 5~10 seconds
7. Demo image (see Fig. B.2)

PC-BASED FACE RECOGNITION TECHNOLOGY

1. Author: Neuro technologija
2. Sales: Commercial

Figure B.1. A sample image

Figure B.2. An execution window
3. Specification: Like fingerprint biometrics, facial recognition technology is widely used in various systems, including physical access control and computer user accounts security. Usually these systems extract certain features from face images and then perform face matching using these features. A face does not have as many uniquely measurable features as fingerprints and eye irises, so facial recognition reliability is slightly lower than these other biometrical recognition methods. However, it is still suitable for many applications, especially when taking into account its convenience for user. Facial recognition can also be used together with fingerprint recognition or another biometrical method for developing more security-critical applications. The multi-biometrical approach is especially important for identification (1:N) systems. In general, identification systems are very convenient to use because they do not require any additional security information (smart cards, passwords etc.). However, using 1:N-matching routines with only one biometrical method, can result in a higher false acceptance probability, which may become unacceptable for applications with large databases. Using face identification as an additional biometric method can dramatically decrease this effect. This multi-biometrical approach also helps in situations where a certain biometric feature is not optimal for certain groups of users. For example, people who do heavy labor with their hands may have rough fingerprints, which can increase the false rejection rate if fingerprint identification was used alone.

4. WWW: http://www.neurotechnologija.com/

5. Environment:
   Intel Pentium® PC 1 GHz or faster
   Microsoft Windows, Linux, Mac OS X

6. System performance:
   Multiple faces detection time (640x480) : 0.07 second
   Single face processing time(after detecting all faces) : 0.13 second
   Matching speed: 100,000 face/ second
   Size of one record in the database: 2.3 Kbytes

7. Demo image (see Fig. B.3)
IDFEND SECURITY SHIELD

1. Author: Neuro technologija
2. Sales: Commercial
3. Specification: This system provides biometric authentication to complement standard card entry systems that use magnetic stripe, proximity or smart card technologies. It operates as an intermediary, routing and filtering access card information to and from legacy systems (e.g., Chubb/ADT MIS). It integrates easily with multi-portal security systems and accommodates up to 200 access points and 20,000 users per installation. It can also be integrated with third-party time & attendance systems
4. WWW: http://www.tcc.us.com/
5. Environment:
   A. Client
   B. Intel Pentium® 1.3GHz
   C. RAM 256 MB 512GB preferred, 40MB HDD
   D. Microsoft Windows XP
   E. Server
   F. Intel Pentium® 1.3GHz
   G. RAM 512 MB 1 GB preferred, 40MB HDD
6. System performance:
   A. Speed: less then 2 second
   B. Scalability: Number of access points is not restricted
   C. Reliability: Reduces fraud compared to card or taken based system
7. Highly Configurable: Multiple configurations of facial biometric authentication and Wiegand access control devices may be applied within the same installation. Unobtrusive Operation-Uses idfend Biometrics Facial Recognition for fast/accurate biometric authentication
8. Demo images (see Fig. B.4)

2. Hand Gesture Recognition Systems

SMART CAR INTERFACE SYSTEM

1. Authors: GM collaborative research lab. And dept. of ECE, Carnegie Mellon Univ.
2. Sales: Commercial

Figure B.4. Sample images
3. Specification
   A. A **companion** that recognizes your settings and keeps you alert.
   B. “**Context aware**” -- responsive to your needs and preferences, road and weather conditions, and information from Internet, on demand.
   C. Equipped with a **gesture interface** that lets you control the car’s electronics with a wave of your hand.
   D. Built with a **speech recognition system** tuned to your voice that connects your car to your handheld computers and cell phone.
   E. Outfitted with the latest **wireless networks** and **Global Positioning Satellite technology** to keep you safe and on time.
   F. Assembled with a **heads-up display** for operating the radio, navigating, checking email, and your schedule.
   G. Able to automatically **modify its own behavior**, make “**graceful upgrades**” to new versions, monitor mechanical and electrical problems, and repair itself until you can get to the shop.

4. WWW: http://gm.web.cmu.edu/research/index.php
5. Demo images (see Fig. B.5 and Fig. B.6)

HANDVU

1. Author: The moves institute
2. Sales: Non-commercial
3. Specification: The HandVu is a vision-based hand gesture interface. It detects the hand in a standard posture, then tracks it and recognizes key postures — all in real-time and without the need for camera or user calibration. It is very useful for interfacing between human and machine with only hand gesture combined with virtual reality and wearable computer.
4. WWW: www.movesinstitute.org/~kolsch/HandVu/HandVu.html#overview
5. Demo images (see Figures B.7, B.8, B.9, and B.10)

**Figure B.5. Smart car system using hand gesture**

**Figure B.6. Training for a certain hand gesture**
3. Head Gesture Recognition Systems

USE YOUR HEAD SYSTEM

1. Author: Cybernet Systems Corp.
2. Sales: Commercial
3. Specification: The system makes it possible to assign just about any keyboard command you want to have easily accessible to your head motions. The UseYourHead converts these head movements into user definable keystrokes. In effect, it acts like a second keyboard. This simple implementation allows it to work well with almost all existing games (virtually any game that accepts keyboard input). Additionally, game developers will be able to directly integrate support for UseYourHead (through use of a DirectInput wrapper) to allow a more continuous motion.
5. Demo image: Nothing

WATSON: Real-Time Head Tracking And Gesture Recognition

1. Author: MIT CSAIL, Vision Interfaces Lab
2. Sales: Non-Commercial
3. Specification: The Waton can track rigid objects in real-time with 6 degrees of free-
dom using a tracking framework called Adaptive View-Based Appearance Model. The tracking library can estimate the pose of the object for a long period of time with bounded drift. Our main application is head pose estimation and gesture recognition using a USB camera or a stereo camera. Our approach combines an Adaptive View-based Appearance Model (AVAM) with a robust 3D view registration algorithm. AVAM is a compact and flexible representation of the object that can be used during the tracking to reduce the drift in the pose estimates. The model is acquired online during the tracking and can be adjusted according to the new pose estimates. Relative poses between frames are computed using a hybrid registration technique which combine the robustness of ICP (Iterative Closest Point) for large movement and the precision of the normal flow constraint. The complete system runs at 25Hz on a Pentium 4 3.2GHz.


5. Demo images (see Fig. B.11)

4. Body Gesture Recognition Systems

GESTURE + PLAY

1. Author: MIT CSAIL, Vision Interfaces Lab
2. Sales: Non-Commercial
3. Specification: We are developing a perceptual interface toolkit based on stereo vision sensing. Stereo vision allows accurate estimation of 3-D position and orientation cues, and also allows robust segmentation of the image of a user from other objects or people in a scene. We have developed software that can track users’ movement and gesture in real-time and that is robust to crowds, lighting changes, motion, and clothing variation. Our toolkit returns the 3D position and articulated body posture of multiple users as they move in an arbitrary workspace. Gestures, actions, and motions in this space cause motion and action in a virtual workspace, e.g. in a game or avatar-world. For this study we are focusing primarily on game environments, for the playful and evocative nature of their visual content and the variety of possible perceptual interaction styles. The virtual world is created using the HALF-LIFE

Figure B.11. Sample images
game engine from Sierra/Valve, which enabled us to create rich environments with texture mapping and animations that also contain active objects with behavior. Our installation is physically realized as a simple video projection wall, a stereo-camera and an ample open space in front.

4. WWW: http://groups.csail.mit.edu/vision/vip/index.htm
5. Demo image (see Fig. B.12)

Figure B.12. Sample images