In S.Li and A.Jain, (ed). *Handbook of Face Recognition*. Springer-Verlag, 2005

Face Databases

Ralph Gross
The Robotics Inistitute, Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh, PA 15213
Email: {rgross}@cs.cmu.edu

Because of its nonrigidity and complex three-dimensional (3D) structure, the appearance of a face is affected by a large number of factors including identity, face pose, illumination, facial expression, age, occlusion, and facial hair. The development of algorithms robust to these variations requires databases of sufficient size that include carefully controlled variations of these factors. Furthermore, common databases are necessary to comparatively evaluate algorithms. Collecting a high quality database is a resource-intensive task: but the availability of public face databases is important for the advancement of the field. In this chapter we review 27 publicly available databases for face recognition, face detection, and facial expression analysis.

1 Databases for Face Recognition

Face recognition continues to be one of the most popular research areas of computer vision and machine learning. Along with the development of face recognition algorithms, a comparatively large number of face databases have been collected. However, many of these databases are tailored to the specific needs of the algorithm under development. In this section we review *publicly* available databases that are of demonstrated use to others in the community. At the beginning of each subsection a table summarizing the key features of the database is provided, including (where available) the number of subjects, recording conditions, image resolution, and total number of images. Table 1 gives an overview of the recording conditions for all databases discussed in this section. Owing to space constraints not all databases are discussed at the same level of detail. Abbreviated descriptions of a number of mostly older databases are included in Section 1.13. The scope of this section is limited to databases containing full face imagery. Note, however, that there are databases of subface images available, such as the recently released CASIA Iris database [23].

1.1 AR Database

No. of subjects	Conditions		Image Resolution	No. of Images
116	Facial expressions	4	768×576	3288
	Illumination	4		
	Occlusion	2		
	Time	2		
http://rvl1.ecn.purdue.edu/~aleix/aleix_face_DB.html				

The AR database was collected at the Computer Vision Center in Barcelona, Spain in 1998 [25]. It contains images of 116 individuals (63 men and 53 women). The imaging and recording conditions (camera parameters, illumination setting, camera distance) were carefully controlled and constantly recalibrated to ensure that settings are identical across subjects. The resulting RGB color images are 768×576 pixels in size. The subjects were recorded twice at a 2–week interval. During each session 13 conditions with varying facial expressions, illumination and occlusion were captured. Figure 1 shows an example for each condition. So far, more than 200 research groups have accessed the database.

Database	No. of subjects	Pose	Illumination	Facial Expressions	Time
AR	116	1	4	4	2
BANCA	208	1	++	1	12
CAS-PEAL	66 – 1040	21	9 – 15	6	2
CMU Hyper	54	1	4	1	1 – 5
CMU PIE	68	13	43	3	1
Equinox IR	91	1	3	3	1
FERET	1199	9 – 20	2	2	2
Harvard RL	10	1	77 – 84	1	1
KFDB	1000	7	16	5	1
MIT	15	3	3	1	1
MPI	200	3	3	1	1
ND HID	300+	1	3	2	10/13
NIST MID	1573	2	1	++	1
ORL	10	1	++	++	++
UMIST	20	++	1	++	1
U. Texas	284	++	1	++	1
U. Oulu	125	1	16	1	1
XM2VTS	295	++	1	++	4
Yale	15	1	3	6	1
Yale B	10	9	64	1	1

Table 1: Overview of the recording conditions for all databases discussed in this section. Cases where the exact number of conditions is not determined (either because the underlying measurement is continuous or the condition was not controlled for during recording) are marked with "++."

1.2 BANCA Database

No. of Subjects	Conditions		Image Resolution
208	Image quality Time	3 12	720 × 576
http://www.ee.surrey.ac.uk/Research/VSSP/banca/			

The BANCA multi-modal database was collected as part of the European BANCA project, which aimed at developing and implementing a secure system with enhanced identification, authentication, and access control schemes for applications over the Internet [1]. The database was designed to test multimodal identity verification with various acquisition devices (high and low quality cameras and microphones) and under several scenarios (controlled, degraded, and adverse). Data were collected in four languages (English, French, Italian, Spanish) for 52 subjects each (26 men and 26 women). Each subject was recorded during 12 different sessions over a period of 3 months. Recordings for a true client access and an informed imposter attack were taken during each session. For each recording the subject was instructed to speak a random 12-digit number along with name, address, and date of birth (client or imposter data). Recordings took an average of 20 seconds. Figure 2 shows example images for all three recording conditions. The BANCA evaluation protocol specifies training and testing sets for a number of experimental configurations, so accurate comparisons between algorithms are possible.

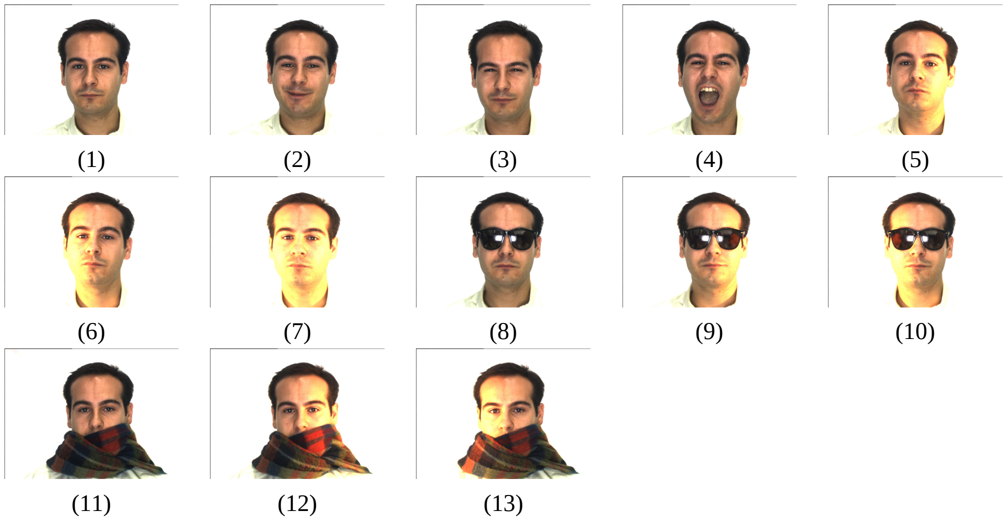


Figure 1: AR database. The conditions are (1) neutral, (2) smile, (3) anger, (4) scream, (5) left light on, (6) right light on, (7) both lights on, (8) sun glasses, (9) sun glasses/left light (10) sun glasses/right light, (11) scarf, (12) scarf/left light, (13) scarf/right light

1.3 CAS-PEAL Database

No. of Subjects	Conditions		Image Resolution	No. of Images
1040	Pose	21	360 × 480	30,900
377	Facial expressions	6		
438	Accessory	6		
233	Illumination	9 – 15		
297	Background	2 – 4		
296	Distance	1 – 2		
66	Time	2		
http://www.jdl.ac.cn/peal/index.html				

The CAS-PEAL (pose, expression, accessory, lighting) Chinese face database was collected at the Chinese Academy of Sciences (CAS) between August 2002 and April 2003.[1] It contains images of 66 to 1040 subjects (595 men, 445 women) in seven categories: pose, expression, accessory, lighting, background, distance, and time [12]. For the pose subset, nine cameras distributed in a semicircle around the subject were used. Images were recorded sequentially within a short time period (2 seconds). In addition, subjects were asked to look up and down (each time by roughly $30°$) for additional recordings resulting in 27 pose images. The current database release includes 21 of the 27 different poses. See Figure 3 for example images.

To record faces under varying yet natural looking lighting conditions, constant ambient illumination together with 15 manually operated fluorescent lamps were used. The lamps were placed at (-$90°$, -$45°$, $0°$, $45°$, $90°$) azimuth and (-$45°$, $0°$, $45°$) elevation. Recording of the illumination images typically took around two minutes; therefore small changes between the images might be present. Example images for all illumination conditions are shown in Figure 4. For the expression subset of the database, subjects were asked to smile, to frown, to look surprised, to close their eyes, and to open the mouth. Images were captured using all nine cameras as described above. In the current database release only the frontal facial expression images are included. A smaller number of subjects were recorded wearing three types of glasses and three types of hats. Again, images were captured using all nine cameras, with only the frontal images currently being distributed. To capture the effect of the camera auto white-balance, subjects were also recorded with five uniformly colored backgrounds (blue, white, black, red, and yellow). Furthermore, images were obtained at two distances (1.2 and 1.4 meters). Finally a small number of

[1]The construction of the CAS-PEAL face database has been supported by the China National Hi-Tech Program 2001AA114010.

Controlled Degraded Adverse

Figure 2: Images for the three recording conditions in the BANCA database. A high quality digital camera was used to record the images for the *controlled* and *adverse* conditions. The images of the *degraded* condition were taken with a low quality web cam.

subjects returned 6 months later for additional recordings. Of the 99,594 images in the database, 30,900 images are available in the current release. To facilitate database distribution, the release images are stored as cropped gray-scale images of size 360 $\times$ 480. Figure 5 shows example images of the currently distributed images.

1.4 CMU Hyperspectral Face Database

No. of Subjects	Conditions		Spectral Range	Image Resolution
54	Illumination	4	0.45 – 1.1 μm	640 $\times$ 480
	Time	1 – 5		
http://www.ri.cmu.edu/pubs/pub_4110.html				

The CMU Hyperspectral database, collected under the DARPA HumanID program [28], covers the visible to near-infrared range from 0.45 to 1.1 μm [7]. Using a CMU developed imaging sensor based on an Acousto-Optic-Tunable Filter (AOTF), the wavelength range is sampled in $10nm$ steps, resulting in 65 images. Acquisition of the 65 images took an average of 8 seconds. Because of the relative lack of sensitivity of the system (only 5 – 10% of light is used), comparatively strong illumination from one to three 600 W halogen lamps was used during data collection. The lamps were placed at -$45°$, $0°$, and $+45°$ with respect to the subject. Each of the 54 subjects was then imaged under four illumination conditions (three lamps individually and then combined). Subjects were recorded between one and five times over a 6–week period. Figure 6 shows example images for a selection of wavelengths between 0.5 and 1 μm.

1.5 CMU Pose, Illumination, and Expression (PIE) Database

No. of Subjects	Conditions		Image Resolution	No. of Images
68	Pose	13	640 $\times$ 486	41,368
	Illumination	43		
	Facial expressions	3		
http://www.ri.cmu.edu/projects/project_418.html				

The CMU PIE database was collected between October and December 2000 [38]. It systematically samples a large number of pose and illumination conditions along with a variety of facial expressions. Although only available for 2 years, PIE has

Figure 3: Pose variation in the CAS-PEAL database. The images were recorded using separate cameras triggered in close succession. The cameras are each about $22.5°$ apart. Subjects were asked to look up, to look straight ahead, and to look down. Shown here are seven of the nine poses currently being distributed.

Figure 4: Illumination variation in the CAS-PEAL database. The images were recorded with constant ambient illumination and manually triggered fluorescent lamps.

already made an impact on algorithm development for face recognition across pose [15, 33] and on the evaluation of face recognition algorithms [16]. So far the database has been distributed to more than 150 research groups.

The PIE database contains 41,368 images obtained from 68 individuals. The subjects were imaged in the CMU 3D Room [21] using a set of 13 synchronized high-quality color cameras and 21 flashes. The resulting RGB color images are 640×480 in size. Figure 7 shows example images of a subject in all 13 poses. In addition to the pose sequence, each subject was recorded under four additional conditions.

1. *Illumination 1*: A total of 21 flashes are individually turned on in rapid sequence. The images in the illumination 1 condition were captured with the room lights on, which produces more natural looking images than the second condition. Each camera recorded 24 images, 2 with no flashes, 21 with one flash firing, and then a final image with no flashes. Only the output of three cameras (frontal, three-quarter, and profile view) was kept.

2. *Illumination 2*: The procedure of the illumination 1 condition was repeated with the room lights off. The output of all 13 cameras was retained in the database. Combining the two illumination settings, a total of 43 illumination conditions were recorded.

Figure 5: Example release images of the pose subset of the CAS-PEAL database. Images are gray-scale and 360 $\times$ 480 in size.

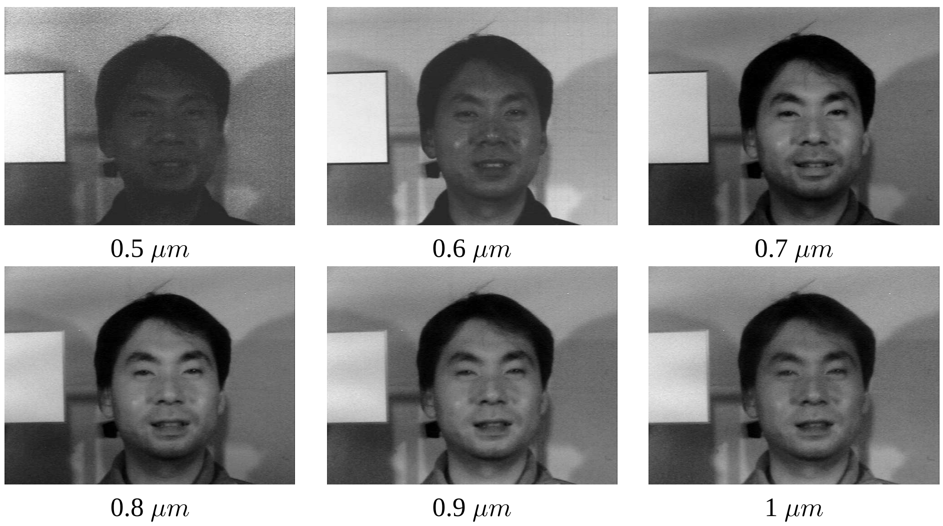


Figure 6: Example images of the CMU Hyperspectral Face Database. Each recording produces images for every $10nm$ step in the range between 0.45 and $1.1\mu m$.

3. *Expression*: The subjects were asked to display a neutral face, to smile, and to close their eyes in order to simulate a blink. The images of all 13 cameras are available in the database.

4. *Talking*: Subjects counted starting at 1 for 2 seconds. 60 frames of them talking were recorded using three cameras (frontal, three-quarter, and profile views).

Examples of the pose and illumination variation are shown in Figure 8. Figure 8a contains variations with the room lights on and Figure 8b with the lights off.

In addition to the raw image data, a variety of miscellaneous “meta-data” were also collected to aid in calibration and other processing.

Head, camera, and flash locations: Using a theodolite, the xyz locations of the head, the 13 cameras, and the 21 flashes were measured. The numerical values of the locations are included in the database and can be used to estimate (relative) head poses and illumination directions.

Background images: At the start of each recording session, a background image was captured from each of the 13 cameras. These images can be used for background subtraction to help localize the face region.

Color calibration images: Although the cameras that were used are all of the same type, there is still a large amount of variation in their photometric responses due to their manufacture and to the fact that the aperture settings on the

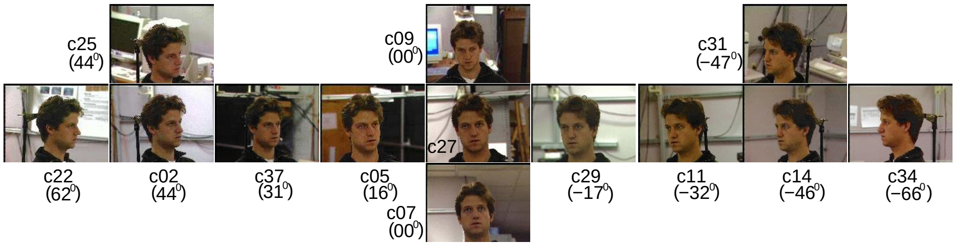


Figure 7: Pose variation in the PIE database. The pose varies from full left profile (c34) over full frontal (c27) to full right profile (c22). Approximate pose angles are shown below the camera numbers.

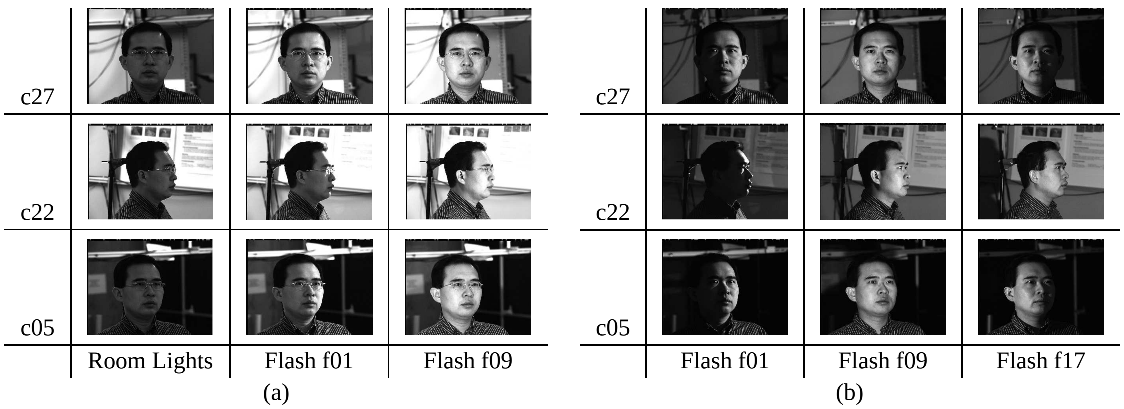


Figure 8: Examples of the pose and illumination variation with the room lights on (a), and the room lights off (b). Note how the combination of room illumination and flashes leads to much more natural looking images than with just the flash alone.

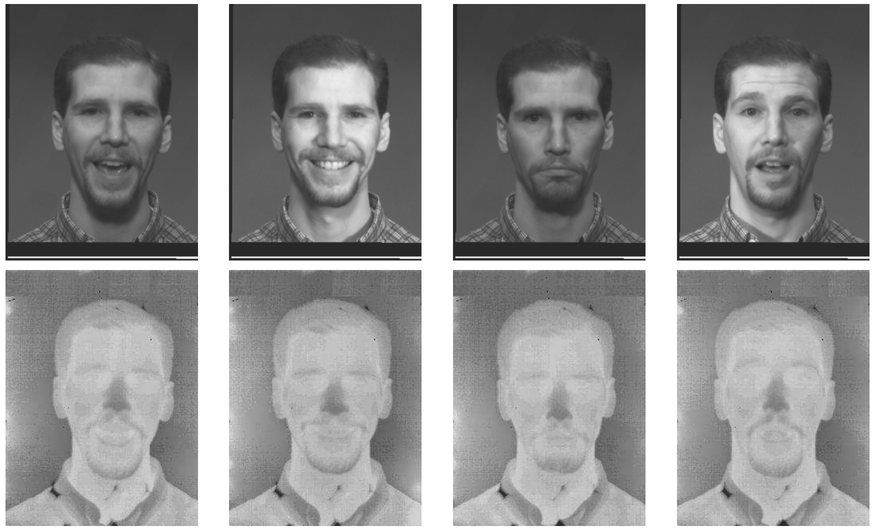
cameras were all set manually. The cameras were "auto white-balanced," but there is still some noticeable variation in their color response. To allow the cameras to be intensity (gain and bias)- and color-calibrated, images of color calibration charts were captured at the start of every session.

1.6 Equinox Infrared Face Database

No. of Subjects	Conditions		Spectral Range	Image Resolution
91	Illumination Facial expressions Speech sequence	3 3 1	8 – 12 μm Visible	240×320
http://www.equinoxsensors.com/products/HID.html				

Various evaluations of academic and commercial face recognition algorithms give empirical evidence that recognizer performance generally decreases if evaluated across illumination conditions [4, 16, 29]. One way to address this problem is to use thermal infrared imagery, which has been shown to be nearly invariant to changes in ambient illumination [42]. As part of the DARPA HumanID program [28], Equinox Corporation collected a database of long-wave infrared (LWIR) imagery in the spectral range of 8 – 12 μm [39]. The database is unique in that the sensor used for the collection simultaneously records video sequences with a visible CCD array and LWIR microbolometer. The resulting image pairs are 240×320 pixels in size and co-registered to within 1/3 pixel. All LWIR images were radiometrically calibrated with a black-body radiator.

The database contains 91 subjects. For each subject, a 4-second (40 frames) video sequence was recorded while the subject pronounced the vowels. Additional still images were obtained in which the subjects display the facial expressions "smile," "frown," and "surprise." All images were recorded under three illumination conditions: frontal, left lateral, and right lateral.



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

Figure 9: Example images of the Equinox IR database. The upper row contains visible images and the lower row long-wave infrared images. The categories are (a) vowel (frontal illumination), (b) "smile" (right illumination), (c) "frown" (frontal illumination), (d) "surprise" (left illumination).

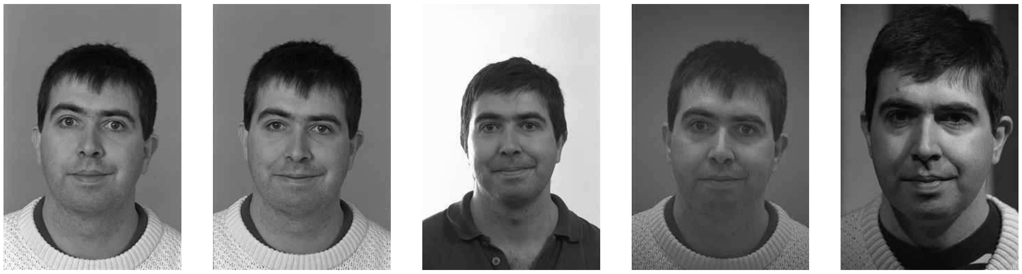
People wearing glasses were imaged twice: with and without glasses. Figure 9 shows example images of both the visual and LWIR imagery across all imaging conditions. For some subjects additional boresighted short-wave infrared (SWIR) (0.9 – 1.7 μm) and middle-wave infrared (MWIR) (3 – 5 μm) data are available.

1.7 FERET

No. of Subjects	Conditions		Image Resolution	No. of Images
1199	Facial expressions Illumination Pose Time	2 2 9 – 20 2	256 $\times$ 384	14,051
http://www.nist.gov/humanid/feret/				

The Facial Recognition Technology (FERET) database was collected at George Mason University and the US Army Research Laboratory facilities as part of the FERET program, sponsored by the US Department of Defense Counterdrug Technology Development Program [30, 32]. The FERET and facial recognition vender test (FRVT) 2000 [4] evaluations as well as independent evaluations [6] used the database extensively, so detailed performance figures are available for a range of research algorithms as well as commercial face recognition systems. The lists of images used in training, gallery, and probe sets are distributed along with the database, so direct comparisons of recognizer performance with previously published results are possible. To date, the database has been distributed to more than 460 research groups.

The FERET database documentation lists 24 facial image categories. The images were recorded in 15 sessions between August 1993 and July 1996. Because the recording equipment had to be reassembled for each session, slight variations between recording sessions are present. Images were recorded with a 35 mm camera, subsequently digitized, and then converted to 8-bit gray-scale images. The resulting images are 256 $\times$ 384 pixels in size. Figure 10 shows five frontal image categories of the database. The *fa* and *fb* images were obtained in close succession. The subjects were asked to display a different facial expression for the *fb* image. The resulting changes in facial expression are typically subtle, often switching between "neutral" and "smiling." The images in the *fc* category were recorded with a different camera and under different lighting. A number of subjects returned at a later date to be imaged again. For the images in the duplicate I set, 0 to 1031



fa | fb | duplicate I | fc | duplicate II

Figure 10: Frontal image categories used in the FERET evaluations. For images in the *fb* category a different facial expression was requested. The *fc* images were recorded with a different camera and under different lighting conditions. The duplicate images were recorded in a later session, with 0 and 1031 days (duplicate I) or 540 to 1031 days (duplicate II) between recordings.

bb	bc	bd	be	ba	bf	bg	bh	bi
$+60^\circ$	$+40^\circ$	$+25^\circ$	$+15^\circ$	0°	-15°	-25°	-40°	-60°

Figure 11: Pose variation in the FERET database. The poses vary from $+60^\circ$ (bb) to full frontal (ba) and on to -60° (bi). Images are available for 200 subjects.

days passed between recording sessions (median 72 days, mean 251 days). A subset of these images forms the duplicate II set, where at least 18 months separated the sessions (median 569 days, mean 627 days).

The remaining image categories cover a wide set of pose variations. Figure 11 shows categories *ba* through *bi*. The images were recorded by asking the subject to rotate the head and body. The pose angles range from $+60^\circ$ to -60°. These pose data are available for 200 subjects.

A different set of pose images is shown in Figure 12. Here images were collected at the following head aspects: right and left profile (labeled *pr* and *pl*), right and left quarter profile (*qr*, *ql*) and right and left half profile (*hr*, *hl*). In these categories images were recorded for 508 to 980 subjects. In addition, five irregularly spaced views were collected for 264 to 429 subjects.

Ground-truth information, including the date of the recording and if the subject is wearing glasses, is provided for each image in the data set. In addition, the manually determined locations of left and right eye and the mouth center is available for 3816 images.

In a new release of the FERET database, NIST is making higher resolution (512 $\times$ 768) color images of most of the original gray-scale images available. More information about the color FERET dataset can be found at http://www.nist.gov/humanid/colorferet/.

1.8 Korean Face Database (KFDB)

No. of Subjects	Conditions		Image Resolution	No. of Images
1000	Pose	7	640 $\times$ 480	52,000
	Illumination	16		
	Facial expressions	5		

pr	hr	qr	ql	hl	pl
$+90°$	$+67.5°$	$+22.5°$	$-22.5°$	$-67.5°$	$-90°$

Figure 12: Additional set of pose images from the FERET database. Images were collected at the following head aspects: right and left profile (labeled *pr* and *pl*), right and left quarter profile (*qr*, *ql*), and right and left half profile (*hr*, *hl*).

$+45°$	$+30°$	$+15°$	$+00°$	$-15°$	$-30°$	$-45°$

Figure 13: Pose variation in the Korean face database. The poses vary from $+45°$ to full frontal and on to -$45°$.

The Korean Face Database (KFDB) contains facial imagery of a large number of Korean subjects collected under carefully controlled conditions [19]. Similar to the CMU PIE database described in Section 1.5, images with varying pose, illumination, and facial expressions were recorded. The subjects were imaged in the middle of an octagonal frame carrying seven cameras and eight lights (in two colors: fluorescent and incandescent) against a blue screen background. The cameras were placed between $45°$ off frontal in both directions at $15°$ increments. Figure 13 shows example images for all seven poses. Pose images were collected in three styles: natural (no glasses, no hair band to hold back hair from the forehead), hair band, and glasses. The lights were located in a full circle around the subject at $45°$ intervals. Separate frontal pose images were recorded with each light turned on individually for both the fluorescent and incandescent lights. Figure 14 shows example images for all eight illumination conditions. In addition, five images using the frontal fluorescent lights were obtained with the subjects wearing glasses. The subjects were also asked to display five facial expressions — neutral, happy, surprise, anger, and blink — which were recorded with two different colored lights, resulting in 10 images per subject. In total, 52 images were obtained per subject. The database also contains extensive ground truth information. The location of 26 feature points (if visible) is available for each face image.

1.9 Max Planck Institute for Biological Cybernetics Face Database

No. of Subjects	Conditions		Image Resolution
200	Modality	2	256×256
	Pose	7	
http://faces.kyb.tuebingen.mpg.de/			

The face database from the Max Planck Institute for Biological Cybernetics is unique, as it is based on 3D data collected with a $Cyberware$ laser scanner [5]. The database contains 200 subjects (100 men, 100 women). In addition to the head structure data, which are stored in a cylindrical representation at a resolution of 512 sample points for both horizontal and vertical angles, color values are recorded at the same spatial resolution. Subjects were wearing bathing caps at the time of recording that are later automatically removed. The faces are free of makeup, accessories, or facial hair. A mostly automatic postprocessing procedure normalized the faces into a standard reference frame and removed shoulders and the region behind the ears. The face representation after all processing consists of approximately 70,000 vertices and a similar number of color

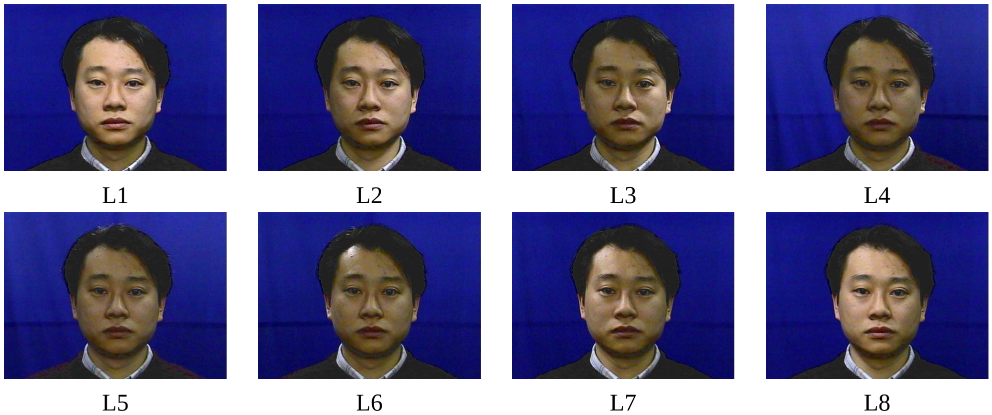


Figure 14: Illumination variation in the Korean face database. Lights from eight different positions (L1 - L8) located in a full circle around the subject were used. For each position images with both fluorescent and incandescent lights were taken.

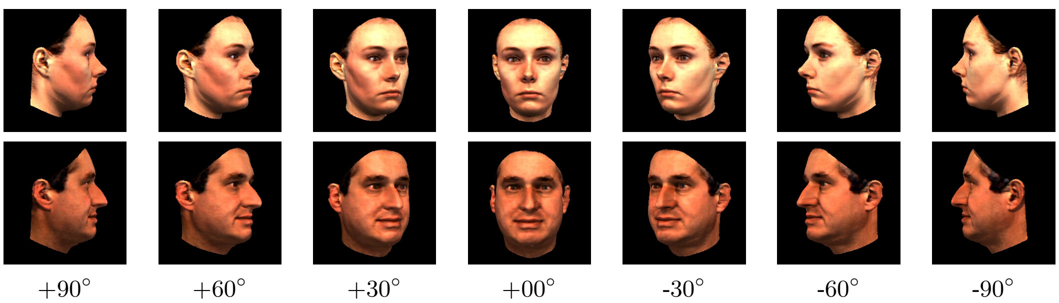


Figure 15: Example images of the MPI face database. The images show seven different poses ranging from full left profile over frontal to full right profile for a female subject (first row) and a male subject (second row).

values. For all subjects, 2D image data at seven face poses are available for download. Figure 15 shows example images for both a female and a male subject. For 7 of the 200 subjects, the full 3D data are available.

1.10 Notre Dame HumanID Database

No. of Subjects	Conditions		Image Resolution	No. of Images
> 300	Time	10/13	1600×1200	$> 15,000$
	Illumination	3		
	Facial Expressions	2		
http://www.nd.edu/~cvrl/HID-data.html				

A number of evaluations of face recognition algorithms have shown that performance degrades if there is a time delay between the acquisition of gallery and probe images [4, 16, 29, 30]. Gross et al. [16] observed differences already for 2–week time laps. The face recognition vendor test 2002 measured a 5% decrease in recognizer performance for each year between recording the gallery and probe images [29]. To further investigate the time dependence in face recognition, a large database is currently being collected at the University of Notre Dame [11] under the DARPA HumanID program [28]. During the spring semester of 2002 approximately 80 subjects were imaged 10 times over an 11–week period. In the fall of 2002 a much larger pool of more than 300 subjects was recorded multiple times over 13 weeks. A minimum of four high-resolution color images

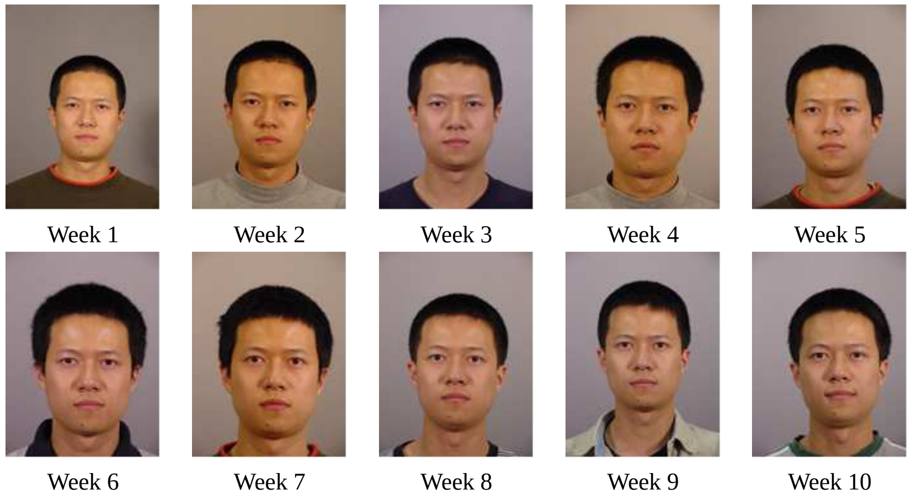


Figure 16: Notre Dame HumanID database. The subjects are imaged on a weekly basis over 11- to 13–week periods. Example images show the subject under controlled studio lighting.

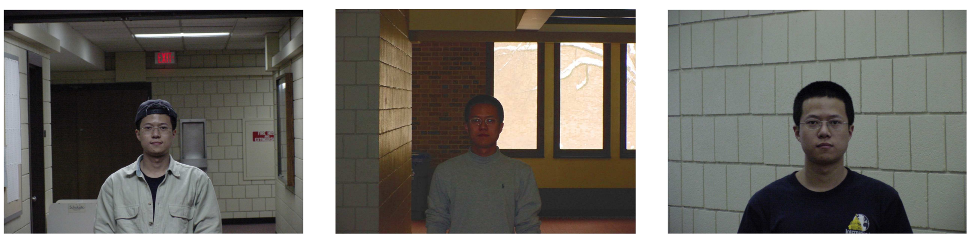


Figure 17: Notre Dame HumanID database. Example images of the “unstructured” lighting condition recorded in the hallway outside of the laboratory.

are obtained during each session under controlled conditions. Subjects are recorded under two lighting configurations. The “FERET style lighting” uses two lights to the right and left of the subject. “Mugshot lighting” uses an additional third light directly opposite the subject’s face. Also, two facial expression images (“neutral” and “smile”) are recorded. Figure 16 shows the 10 images of a subject from weeks 1 through 10. In addition to the studio recordings, two images with “unstructured” lighting are obtained. These images were recorded in a hallway outside the laboratory with a different camera and weekly changing subject positions. Figure 17 shows examples of this condition. As part of the University of Notre Dame collection efforts, a number of additional images are recorded including infrared, range, hand, and ear images.

1.11 University of Texas Video Database

No. of Subjects	Conditions	Image Resolution
284	Pose (still and video) Facial speech (video) Facial expression (video)	720×480
http://www.utdallas.edu/dept/bbs/FACULTY_PAGES/otoole/database.htm		

Wheras the face recognition vendor test 2002 did not measure improvements in face recognition performance from video input [29] it is generally believed that the use of video data has considerable potential for improving face detection and recognition [44]. Recently, the collection of a large database of static digital images and video clips of faces at the University

Figure 18: University of Texas Video Database. Example images for the different recording conditions of the database. First row: Facial speech. Second row: Laughter. Third row: Disgust.

of Texas has been completed [27]. The database contains a total of 284 subjects (208 female, 76 male). Most of the participants were Caucasians between the ages of 18 and 25. The subjects were imaged at close range in front of a neutral background under controlled ambient lighting. To cover up clothing each participant wore a gray smock. Data were collected in four different categories: *still facial mug shots*, *dynamic facial mug shots*, *dynamic facial speech* and *dynamic facial expression*. For the still facial mug shots, nine views of the subject, ranging from left to right profile in equal-degree steps were recorded. The subjects were instructed to use markers suspended from the ceiling at the appropriate angles as guides for correct head positioning. For the dynamic facial mug shot category a moving version of the still facial mug shots was recorded. Aided by a metronome the subjects rotated their heads from one position to the next in 1-second steps. The resulting sequences are 10 seconds in length. The dynamic facial speech sequences were recorded while the subjects responded to a series of mundane questions, eliciting head motions, and changes in facial expression and eye gaze direction. For most faces in the database, a "neutral" and an "animated" sequence are available. The sequence length is cropped to be 10 seconds. For the dynamic facial expression sequences, the subjects were recorded while watching a 10–minute video intended to elicit different emotions. After the recording, short 5-second clips with the subject displaying facial expressions corresponding to happiness, sadness, fear, disgust, anger, puzzlement, laughter, surprise, boredom, or disbelief were hand selected and included in the database along with a 5-second "blank stare" clip, which contains no facial motions but possibly other movements of the head or the eyes. Figure 18 shows example images from three of the four recording conditions.

1.12 Yale Face Database B

No. of Subjects	Conditions		Image Resolution	No. of Images
10	Pose Illumination	9 64	640 × 480	5850
http://cvc.yale.edu/projects/yalefacesB/yalefacesB.html				

The Yale Face Database B [13] was collected to allow systematic testing of face recognition methods under large variations in illumination and pose. The subjects were imaged inside a geodesic dome with 64 computer-controlled xenon strobes. Images of 10 individuals were recorded under 64 lighting conditions in nine poses (one frontal, five poses at $12°$, and three poses at $24°$ from the camera axis). Because all 64 images of a face in a particular pose were acquired within about 2 seconds, only minimal changes in head position and facial expression are visible. The database is divided into four subsets according to the angle between the light source and the camera axis ($12°$, $25°$, $50°$, $77°$). Hand-labeled locations of the eyes and the center of the mouth are distributed along with the database. Example images of the database are shown in Figure 19.

Figure 19: Yale Face Database B: 18 example images from the 64 illumination conditions.

1.13 Older Databases

1.13.1 Harvard Robotics Lab (HRL) Database

# Subjects	Conditions		Image Resolution
10	Illumination	77–84	193 × 254
ftp://cvc.yale.edu/CVC/pub/images/hrlfaces			

The HRL database collected by Peter Hallinan was the first database to systematically sample a wide range of illumination conditions [17, 18]. The subjects were seated on a stool and instructed to maintain a constant head position (although slight movements were unavoidable). The face was then illuminated by a dominant light source whose positions sampled the hemisphere in front of the subject in $15°$ increments (longitude and latitude). In this way at least 75 images of each of 10 subjects were recorded.

1.13.2 MIT Database

No. of Subjects	Conditions	Image Resolution	No. of Images
16	Head orientation Illumination Scale	120 × 128	433
ftp://whitechapel.media.mit.edu/pub/images/			

The MIT database contains 16 subjects under three lighting conditions (head-on, $45°$, $90°$), three scale conditions, and three head tilt condition (upright, right, left) [41].

1.13.3 NIST Mugshot Identification Database (MID)

No. of Subjects	Conditions	Image Resolution	No. of Images
1573	Frontal and profile view	Varying	3248
http://www.nist.gov/srd/nistsd18.htm			

The NIST Mugshot Identification Database contains frontal and profile views of 1573 (mostly male) subjects.

1.13.4 Olivetti Research Lab (ORL) Database

No. of Subjects	Image Resolution	No. of Images
10	92 × 110	400
http://www.uk.research.att.com/facedatabase.html		

The ORL database was collected between 1992 and 1994 [36]. It contains slight variations in illumination, facial expression (open/closed eyes, smiling/not smiling) and facial details (glasses/no glasses). However, the conditions were not varied systematically. As argued by Phillips and Newton [31] algorithm performance over this database has been saturated.

1.13.5 UMIST Face Database

No. of Subjects	Conditions	Face Resolution	No. of Images
20	Pose	220 × 220	564
http://images.ee.umist.ac.uk/danny/database.html			

The UMIST Face Database [14] contains image sequences of subjects slowly rotating their head from profile to frontal view.

1.13.6 University of Oulu Physics-Based Face Database

No. of Subjects	Conditions		Image Resolution
125	Camera calibration Illumination	16	428 × 569
http://www.ee.oulu.fi/research/imag/color/pbfd.html			

The University of Oulu Physics-based Face database contains color images of faces under different illuminants and camera calibration conditions as well as skin spectral reflectance measurements of each person [24]. Four Macbeth SpectraLight illuminants were used: horizon, incandescent, TL84 fluorescent, and daylight. Images were collected by white-balancing and linearizing the camera for one illumination condition and then recording the subject under all four illumination conditions without changing the camera setting. This results in 16 images per subject. If the subject was wearing glasses, an additional 16 images with the subject wearing them were recorded. Additional information such as camera spectral response and the spectral power distribution of the illuminants are provided.

1.13.7 Yale Face Database

No. of Subjects	Conditions		Image Resolution	No. of Images
15	W/ and w/o glasses Illumination Facial expressions	2 3 6	320 × 243	165
http://cvc.yale.edu/projects/yalefaces/yalefaces.html				

The Yale Face database [2] contains 11 images of 15 subjects in a variety of conditions including with and without glasses, illumination variation, and changes in facial expression.

1.13.8 XM2VTS Database

No. of Subjects	Conditions	Image Resolution
295	Head rotation sequences Speech sequences Time	720 × 576
http://www.ee.surrey.ac.uk/Research/VSSP/xm2vtsdb/		

The XM2VTS database [26] was collected for research and development of identity verification systems using multimodal (face and voice) input data. The database contains 295 subjects, each recorded at four sessions over a period of 4 months. At each session two head rotation shots and six speech shots (subjects reading three sentences twice) were recorded. 3D models of 293 subjects are available as well. The XM2VTS evaluation protocol specifies training, evaluation, and test sets, so detailed comparisons between algorithms are possible. A variety of subsets of the database are available for purchase from the University of Surrey. To date, the XM2VTS database has been distributed to more than 100 institutions.

2 Databases for Face Detection

Face detection algorithms typically have to be trained on face and nonface images to build up an internal representation of the human face. For this purpose the face databases listed in Section 1 are often used. According to a recent survey of face detection algorithms [43], popular choices are the FERET, MIT, ORL, Harvard, and AR databases. Along with these public databases, independently collected, nonpublic databases are often also employed. To comparatively evaluate the performance

Figure 20: Example images from the *Upright Test Set* portion of the MIT/CMU test set.

of face detection algorithms, common testing data sets are necessary. These data sets should be representative of real-world data containing faces in various orientations against a complex background. In recent years two public data sets emerged as quasi-standard evaluation test sets: the combined MIT/CMU test set for frontal face detection [35, 40] and the CMU test set II for frontal and nonfrontal face detection [37]. In the following we describe both databases. They are available for download from http://www.ri.cmu.edu/projects/project_419.html.

2.1 Combined MIT/CMU Test Set

The combined MIT/CMU data set includes 180 images organized in two sets. The first group of 130 images contains 507 upright faces (referred to as the *Upright Test Set* [34]); 23 of the 130 images were collected by Sung and Poggio [40]. The images come from a wide variety of sources including the Internet, newspapers and magazines (scanned at low resolution), analog cameras, and hand drawings. Figure 20 shows example images from this set. A number of images are included that do not contain any faces in order to test tolerance to clutter.

The second set of data was collected to test detection of tilted faces. It contains 50 images with 223 faces, among which 210 are at angles of more than 10° from upright. This set is referred to as the *Tilted Test Set* [34]. Figure 21 shows example images from this set.

2.2 CMU Test Set II

This dataset was collected to test face detection algorithms that are able to handle out-of-plane rotation. It contains 208 images with 441 faces out of which 347 are in profile view. The images were all collected from the Internet. Figure 22 shows example images from this set.

2.3 Other Databases

2.3.1 Nonface Images

Unlike the well defined class of face images, it is much more difficult if not impossible to characterize the class of non-face images. This task is important as face detection algorithms typically operate by discriminating between “images

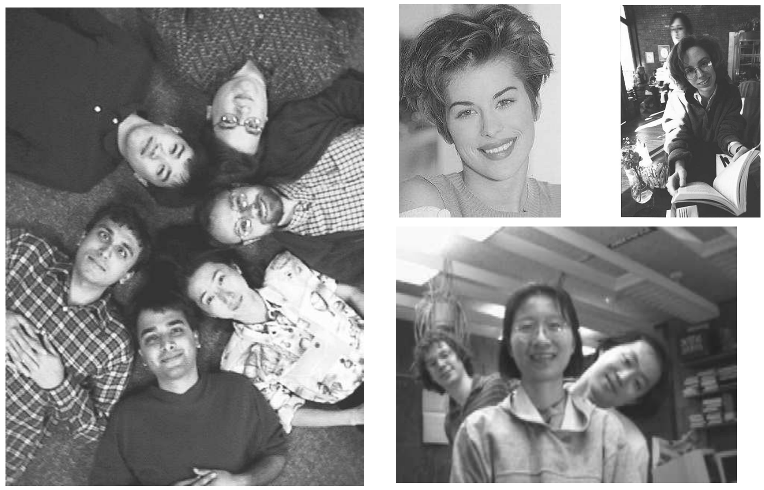


Figure 21: Example images from the *Tilted Test Set* portion of the MIT/CMU test set.

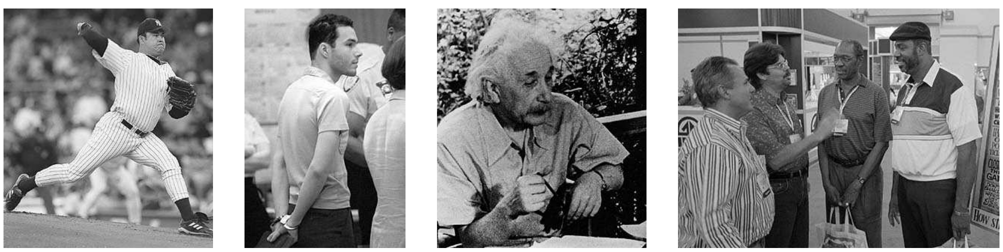


Figure 22: Images from the CMU Test Set II. Most of the faces in this test set are in profile view.

containing faces" and "images not containing faces." A popular source of nonface images other than the World Wide Web is the Washington University archive, which contains a wide variety of nonface images. The archive is accessible at http://wuarchive.wustl.edu/~aminet/pix/.

2.3.2 BioID Face Database

The BioID database contains 1521 frontal view images of 23 subjects. Images are 384×288 in size. Because the images were recorded at different locations, significant variations in illumination, background, and face size are present. Manually marked eye locations are available for all images. The database can be downloaded from http://www.humanscan.de/support/downloads/facedb.php.

2.3.3 MIT CBCL Face Database #1

The MIT Center for Biological and Computation Learning distributes a database of 2901 faces and 28,121 nonfaces. The images are 19×19 in size and gray-scale. The database is available for download from http://www.ai.mit.edu/projects/cbcl/software-datasets/F

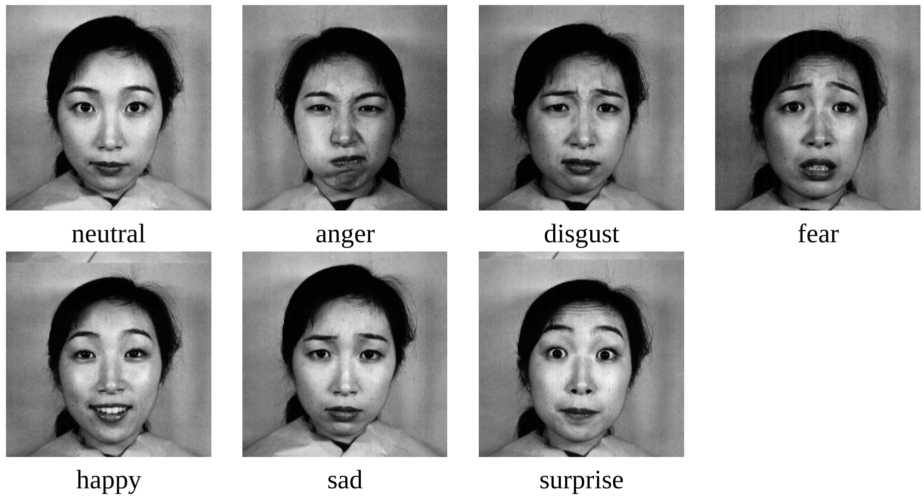


Figure 23: Example images from the JAFFE database. The images in the database have been rated by 60 Japanese female subjects on a 5-point scale for each of the six adjectives. The majority vote is shown underneath each image (with neutral being defined through the absence of a clear majority) .

3 Databases for Facial Expression Analysis

The human face is able to display an astonishing variety of expressions. Collecting a database that samples this space in a meaningful way is a difficult task. Following the work on automated facial expression analysis, available databases cluster into two categories. In one group we find work that has concentrated on recognizing what Ekman and Friesen called the six *basic emotions* [8]—happiness, sadness, fear, disgust, surprise and anger—from either single images or image sequences. In the other group research has focused on extracting a more fine-grained description of facial expressions. Here the underlying data are typically coded using the facial action coding system (FACS) [9], which was designed to describe subtle changes in facial features in terms of 44 distinct action units. Of the 44 action units, 30 are anatomically related to a specific set of facial muscles. For the remaining 14 units the anatomic basis is unspecified. The following section describes well known databases in both groups. A question unique to collecting data for facial expression analysis is how facial actions are elicited during data collection. To facilitate the collection process, subjects are usually asked to perform the desired actions. However, appearance and timing of these directed facial actions may differ from spontaneously occurring behavior [10]. Most of the databases collected primarily for face recognition also recorded subjects under changing facial expressions (see Section 1). The video database collected at the University of Texas [27] deserves special mention here, because it also contains a wide range of spontaneous facial expressions from a large number of subjects (see Section 1.11).

3.1 Japanese Female Facial Expression (JAFFE) Database

No. of Subjects	Expressions	Image Resolution
10	7	256×256
http://www.mis.atr.co.jp/˜mlyons/jaffe.html		

The JAFFE database contains 213 images of 10 Japanese female models obtained in front of a semireflective mirror [22]. Each subject was recorded three or four times while displaying the six basic emotions and a neutral face. The camera trigger was controlled by the subjects. The resulting images have been rated by 60 Japanese women on a 5-point scale for each of the six adjectives. The rating results are distributed along with the images. Figure 23 shows example images for one subject along with the majority rating. The images were originally printed in monochrome and then digitized using a flatbed scanner.

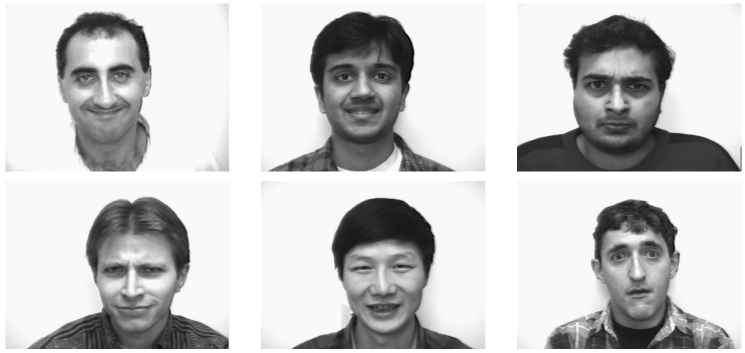


Figure 24: Images from the University of Maryland database. The images show peak frames taken from an image sequence in which the subjects display a set of facial expressions of their choice.

Figure 25: Cohn-Kanade AU-Coded Facial Expression database. Examples of emotion-specified expressions from image sequences.

3.2 University of Maryland Database

No. of Subjects	Expressions	Image Resolution
40	6	560×240
http://www.umiacs.umd.edu/users/yaser/DATA/index.html		

The University of Maryland database contains image sequences of 40 subjects of diverse racial and cultural backgrounds [3]. The subjects were recorded at full frame rate while continuously displaying their own choice of expressions. In contrast to other databases the subjects were instructed to move their heads but avoid profile views. The resulting sequences were later ground-truthed manually. The database contains 70 sequences with a total of 145 expressions. Each sequence is about 9 seconds long and contains one to three expressions. Occurrences of the six basic emotions were not balanced, with “happiness,” “surprise,” “disgust,” and “anger” being more frequent than “sadness” and “fear.” Figure 24 shows peak frame examples from the database.

3.3 Cohn-Kanade AU-Coded Facial Expression Database

No. of Subjects	Expressions	Image Resolution
100	23	640×480
http://vasc.ri.cmu.edu/idb/html/face/facial_expression/index.html		

The Cohn-Kanade AU-Coded Facial Expression Database is publicly available from Carnegie Mellon University [20]. It contains image sequences of facial expressions from men and women of varying ethnic backgrounds. The subjects perform a series of 23 facial displays that include single action units and combinations of action units. A total of 504 sequences are available for distribution. The camera orientation is frontal. Small head motion is present. There are three variations in lighting: ambient lighting, single-high-intensity lamp, and dual high-intensity lamps with reflective umbrellas. Facial expressions are coded using the facial action coding system [9] and assigned emotion-specified labels. Emotion expressions included happy, surprise, anger, disgust, fear, and sadness. Examples of the expressions are shown in Figure 25.

Acknowledgments

We thank the following individuals for their help in compiling this chapter: Simon Baker, Kevin Bowyer, Jeffrey Cohn, Patrick Flynn, Athos Georghiades, Patrick Grother, Josh Harms, Bon-Woo Hwang, Stan Janet, Stan Li, Michael Lyons, Iain Matthews, Aleix Martinez, Jonathon Phillips, Michael Renner, Shiguang Shan, Henry Schneiderman, Diego Socolinsky, Tieniu Tan, Ying-Li Tian, Alice O'Toole, Christian Wallraven, and Yaser Yacoob. This work was supported by U.S. Office of Naval Research contract N00014-00-1-0915 and by U.S. Department of Defense contract N41756-03-C4024.

References

[1] E. Bailly-Bailliere, S. Bengio, F. Bimbot, M. Hamouz, J. Kittler, J. Mariethoz, J. Matas, K. Messer, V. Popovici, F. Poree, B. Ruiz, and J.-P. Thiran. The BANCA database and evaluation protocol. In *Audio- and Video-Based Biometric Person Authentication (AVBPA)*, pages 625–638, 2003.

[2] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. fisherfaces: recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720, July 1997.

[3] M. J. Black and Y. Yacoob. Recognizing facial expressions in image sequences using local parameterized models of image motion. *International Journal of Computer Vision*, 25(1):23–48, 1997.

[4] D. Blackburn, M. Bone, and P. J. Phillips. Facial recognition vendor test 2000: evaluation report, 2000.

[5] V. Blanz and T. Vetter. A morphable model for the synthesis of 3d faces. In *Computer Graphics Proceedings, Annual Conference Series (SIGGRAPH)*, 1999.

[6] D. Bolme, R. Beveridge, M. Teixeira, and B. Draper. The CSU face identification evaluation system: its purpose, features and structure. In *International Conference on Vision Systems*, pages 304–311, 2003.

[7] L. Denes, P. Metes, and Y. Liu. Hyperspectral face database. Technical report, Robotics Institute, Carnegie Mellon University, 2002.

[8] P. Ekman and W. Friesen. Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, 17(2):124–129, 1971.

[9] P. Ekman and W. Friesen. *Facial Action Coding System*. Consulting Psychologist Press, Palo Alto, CA, 1978.

[10] P. Ekman and E. Rosenberg, editors. *What the face reveals*. Oxford University Press, New York, 1997.

[11] P. Flynn, K. Bowyer, and P. J. Phillips. Assesment of time dependency in face recognition: an initial study. In *Audio- and Video-Based Biometric Person Authentication (AVBPA)*, pages 44–51, 2003.

[12] W. Gao, B. Cao, S. Shan, D. Zhou, X. Zhang, and D. Zhao. CAS-PEAL large-scale Chinese face database and evaluation protocols. Technical Report JDL-TR-04-FR-001, Joint Research & Development Laboratory, 2004.

[13] A. Georghiades, D. Kriegman, and P. Belhumeur. From few to many: generative models for recognition under variable pose and illumination. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):643–660, 2001.

[14] D. Graham and N. Allison. Characterizing virtual eigensignatures for general purpose face recognition. In H. Wechsler, P. J. Phillips, V. Bruce, F. Fogelman-Soulie, and T. Huang, editors, *Face Recognition: From Theory to Applications*, pages 446–456. NATO ASI Series F, 1998.

[15] R. Gross, I. Matthews, and S. Baker. Eigen light-fields and face recognition across pose. In *Proceedings of the Fifth International Conference on Face and Gesture Recognition*, pages 1–7, 2002.

[16] R. Gross, J. Shi, and J. Cohn. Quo vadis face recognition? In *Third Workshop on Empirical Evaluation Methods in Computer Vision*, 2001.

[17] P. Hallinan. *A Deformable Model for Face Recognition under Arbitrary Lighting Conditions*. PhD thesis, Harvard University, 1995.

[18] P. Hallinan, G. Gordon, A. Yuille, P. Giblin, and D. Mumford. *Two- and Three-dimensional Patterns of the face*. A.K. Peters, Wellesley, MA, 1999.

[19] B.-W. Hwang, H. Byun, M.-C. Roh, and S.-W. Lee. Performance evaluation of face recognition algorithms on the asian face database, KFDB. In *Audio- and Video-Based Biometric Person Authentication (AVBPA)*, pages 557–565, 2003.

[20] T. Kanade, J. Cohn, and Y. Tian. Comprehensive database for facial expression analysis. In *Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, pages 46–53, 2000.

[21] T. Kanade, H. Saito, and S. Vedula. The 3D room: digitizing time-varying 3D events by synchronized multiple video streams. Technical Report CMU-RI-TR-98-34, CMU Robotics Institute, 1998.

[22] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba. Coding facial expressions with Gabor wavelets. In *3rd International Conference on Automatic Face and Gesture Recognition*, pages 200–205, 1998.

[23] L. Ma, T. Tan, Y. Wang, and D. Zhang. Personal identification based on iris texture analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(12):1519–1533, 2003.

[24] E. Marszalec, B. Martinkauppi, M. Soriano, and M. Pietkainen. A physics-based face database for color research. *Journal of Electronic Imaging*, 9(1):32–38, 2000.

[25] A. R. Martinez and R. Benavente. The AR face database. Technical Report 24, Computer Vision Center(CVC) Technical Report, Barcelona, 1998.

[26] K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre. XM2VTSDB: the extended M2VTS database. In *Second International Conference on Audio and Video-based Biometric Person Authentication*, 1999.

[27] A. O'Toole, J. Harms, S. Snow, D. R. Hurst, M. Pappas, and H. Abdi. A video database of moving faces and people. submitted, 2003.

[28] P. J. Phillips. Human identification technical challenges. In *IEEE International Conference on Image Processing*, volume 1, pages 22–25, 2002.

[29] P. J. Phillips, P. Grother, J. M. Ross, D. Blackburn, E. Tabassi, and M. Bone. Face recognition vendor test 2002: evaluation report, March 2003.

[30] P. J. Phillips, H. Moon, S. Rizvi, and P. J. Rauss. The FERET evaluation methodology for face-recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1090–1104, 2000.

[31] P. J. Phillips and E. M. Newton. Meta-analysis of face recognition algorithms. In *5th IEEE Conf. on Automatic Face and Gesture Recognition*, Washington, DC, 2002.

[32] P. J. Phillips, H. Wechsler, and P. Rauss. The FERET database and evaluation procedure for face-recognition algorithms. *Image and Vision Computing*, 16(5):295–306, 1998.

[33] S. Romdhani, V. Blanz, and T. Vetter. Face identification by matching a 3D morphable model using linear shape and texture error functions. In *Proceedings of the European Conference on Computer Vision*, pages 3–19, 2002.

[34] H. Rowley. *Neural Network-Based Face Detection*. PhD thesis, Carnegie Mellon University, 1999.

[35] H. Rowley, S. Baluja, and T. Kanade. Neural network-based face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(1):23–38, 1998.

[36] F. Samaria and A. Harter. Parameterisation of a stochastic model for human face identification. In *2nd IEEE Workshop on Applications of Computer Vision*, Sarasota, FL, 1994.

[37] H. Schneiderman and T. Kanade. A statistical method for 3D object detection applied to faces and cars. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 746–751, 2000.

[38] T. Sim, S. Baker, and M. Bsat. The CMU pose, illumination, and expression database. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(12):1615–1618, 2003.

[39] D. Socolinsky, L. Wolff, J. Neuheisel, and C. Eveland. Illumination invariant face recognition using thermal infrared imagery. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2001.

[40] K.-K. Sung and T. Poggio. Example-based learning for view-based human face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(1):39–51, 1999.

[41] M. Turk and A. Pentland. Face recognition using eigenfaces. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1991.

[42] L. Wolff, D. Socolinsky, and C. Eveland. Quantitative measurement of illumination invariance for face recognition using thermal infrared imagery. In *IEEE Workshop on Computer Vision Beyond The Visible Spectrum: Methods and Applications*, 2001.

[43] M.-H. Yang, D. Kriegman, and N. Ahuja. Detecting faces in images: a survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(1):34–58, 2002.

[44] W. Zhao, R. Chellappa, A. Rosenfeld, and P. J. Phillips. Face recognition: a literature survey. Technical Report CS-TR4167, University of Maryland, 2000.