Multi-PIE

Ralph Gross[1], Iain Matthews[1], Jeffrey Cohn[2], Takeo Kanade[1], Simon Baker[3]
[1] Robotics Institute, Carnegie Mellon University
[2] Department of Psychology, University of Pittsburgh
[3] Microsoft Research, Microsoft Corporation

Abstract

A close relationship exists between the advancement of face recognition algorithms and the availability of face databases varying factors that affect facial appearance in a controlled manner. The CMU PIE database has been very influential in advancing research in face recognition across pose and illumination. Despite its success the PIE database has several shortcomings: a limited number of subjects, a single recording session and only few expressions captured. To address these issues we collected the CMU Multi-PIE database. It contains 337 subjects, imaged under 15 view points and 19 illumination conditions in up to four recording sessions. In this paper we introduce the database and describe the recording procedure. We furthermore present results from baseline experiments using PCA and LDA classifiers to highlight similarities and differences between PIE and Multi-PIE.

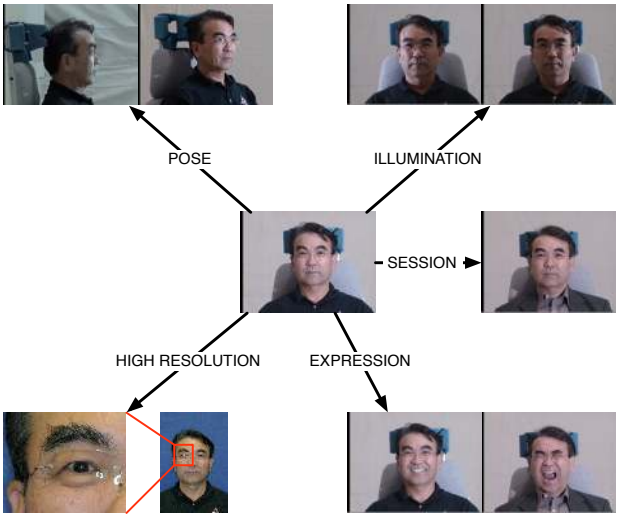


Figure 1. Variation captured in the Multi-PIE face database.

1. Introduction

Facial appearance varies significantly with a number of factors, including identity, illumination, pose, and expression. To support the development and comparative evaluation of face recognition algorithms, the availability of facial image data spanning conditions of interest in a carefully controlled manner is important. Several face databases have been collected over the last decade for this reason, such as the FERET [12], AR [9], XM2VTS [11], Cohn-Kanade [7], and Yale B [5] databases. See [6] for a more comprehensive overview.

To support research for face recognition across pose and illumination the Pose, Illumination, and Expression (PIE) database was collected at CMU in the fall of 2000 [13]. To date more than 450 copies of PIE have been distributed to researchers throughout the world. Despite its success the PIE database has a number of shortcomings; in particular it only contains 68 subjects that were recorded in a single session, displaying a small range of expressions (neutral, smile, blink, and talk).

To address theses issues we collected the Multi-PIE database. The new database improves upon the PIE database in a number of categories as shown in Figure 1 and Table 1, most notably the number of subjects has been substantially increased to 337 with multiple recording sessions (4 vs. only 1 in PIE). In addition the recording environment of the Multi-PIE database has been improved in comparison to the PIE collection through usage of a uniform, static background and live monitors showing subjects during the recording, allowing for constant control of the head position.

This document gives an overview of the Multi-PIE database and provides results of baseline face recognition experiments. Section 2 describes the hardware setup used during the collection. Section 3 explains the recording procedure and shows example images. We detail statistics of the database as well as the subject population in Section 4. Section 5 shows results of evaluations using PCA [14] and LDA [1] in experiments comparing PIE and Multi-PIE as well as in experiments only possible on Multi-PIE.

978-1-4244-2154-1/08/$25.00 ©2008 IEEE

	Multi-PIE	PIE
# Subjects	337	68
# Recording Sessions	4	1
High-Resolution Still Images	Yes	No
Geometrical Calibration Images	Yes	No
# Expressions	6	4
# Cameras	15	13
# Flashes	18	21
Total # Images	750,000+	41,000+
DB Size [GB]	305	40

Table 1. Comparison between the Multi-PIE and PIE databases.

Figure 2. Setup for the high resolution image capture. Subjects were seated in front of a blue background and recorded using a Canon EOS 10D camera with a Macro Ring Lite MR-14EX ring flash.

2. Collection Setup

This section describes the physical setup and the hardware used to record the high resolution still images (Section 2.1) and the multi-pose/illumination images (Section 2.2).

2.1. High Resolution Images

We recorded frontal images using a Canon EOS 10D (6.3-megapixel CMOS camera) with a Macro Ring Lite MR-14EX ring flash. As shown in Figure 2, subjects were seated in front of a blue background in close proximity to the camera. The resulting images are 3072×2048 in size with the inner pupil distance of the subjects typically exceeding 400 pixels.

2.2. Pose and Illumination Images

To systematically capture images with varying poses and illuminations during data acquisition we used a system of

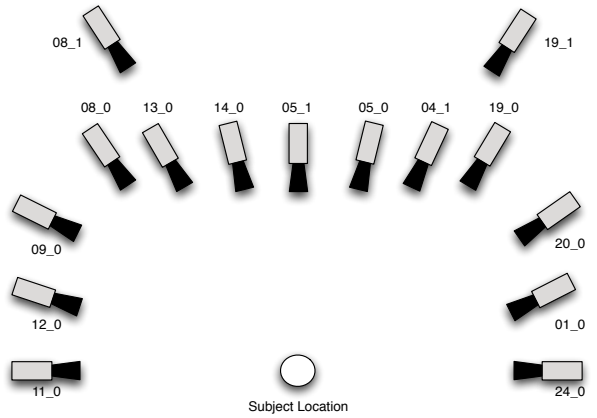


Figure 3. Camera labels and approximate locations inside the collection room. There were 13 cameras located at head height, spaced in 15° intervals. Two additional cameras (08_1 and 19_1) were located above the subject, simulating a typical surveillance camera view. Each camera had one flash attached to it with three additional flashes being placed between cameras 08_1 and 19_1.

15 cameras and 18 flashes connected to a set of Linux PCs. An additional computer was used as master to communicate with the independent recording clients running in parallel on the data capture PCs. This setup is similar to the one used for the CMU PIE database [13]. Figure 3 illustrates the camera positions. Thirteen cameras were located at head height, spaced in 15° intervals, and two additional cameras were located above the subject, simulating a typical surveillance view. The majority of the cameras (11 out of 15) were produced by Sony, model DXC-9000, and the remaining four cameras (positions: 11_0, 08_1, 19_1, and 24_0) Panasonic AW-E600Ps (see Figure 3). Each camera had one flash (model: Minolta Auto 220X) attached to it; above for the 13 cameras mounted at head height and below for the 2 cameras mounted above the subject. In addition, three more flashes were located above the subject between the surveillance-view cameras 08_1 and 19_1. See Figure 4 for a panoramic image of the room with the locations of the cameras and flashes marked with red and blue circles, respectively. All components of the system were hardware synchronized, replicating the system in [8]. All flashes were wired directly to a National Instruments digial I/O card (NI PCI-6503) and triggered in sync with the image capture. This setup was inspired by the system used in the Yale dome [5].

The settings for all cameras were manually adjusted so that the pixel value of the brightest pixel in an image recorded without flash illumination is around 128 to minimize the number of saturated pixels in the flash illuminated images. For the same reason we added diffusers in front of each flash. We also attempted to manually color-balance the cameras so that the resulting images look visually similar.

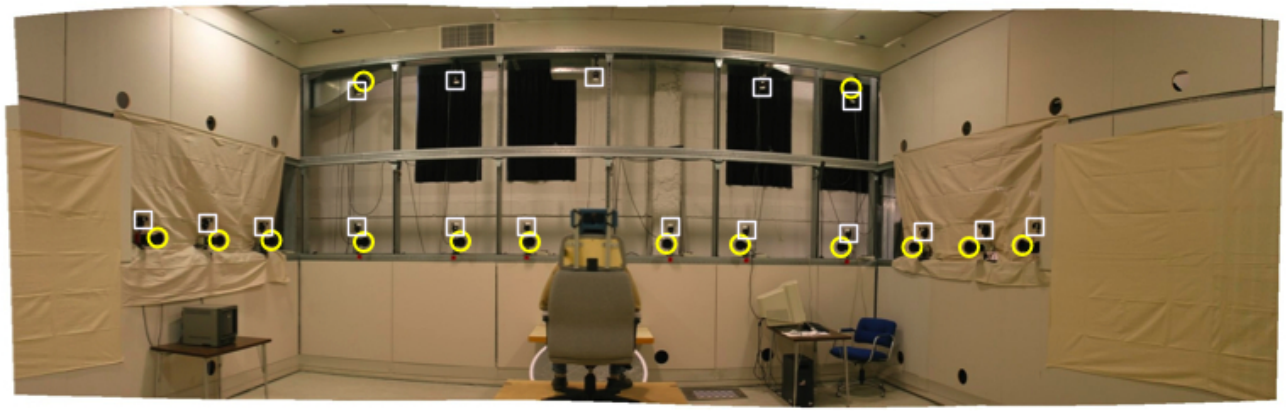


Figure 4. Panoramic image of the collection room. 14 of the 15 cameras used are highlighted with yellow circles, 17 of the 18 flashes are highlighted with white boxes with the occluded camera/flash pair being located right in front of the subject in the chair. The monitor visible to the left was used to ensure accurate positioning of the subject throughout the recording session.

Session 1 Neutral | Session 1 Smile | Session 2 | Session 3 | Session 4

Figure 5. Example high resolution images of one subject across all four recording session. For session 1 we recorded a smile image in addition to the neutral image.

3. Data Collection Procedure

We recorded data during four sessions over the course of six months. During each session we recorded a single neutral high resolution frontal image. In addition, during the first session an additional image showing the subjects smiling was recorded. Figure 5 shows all high resolution images from one subject for sessions 1 through 4.

After the recording of the high resolution images, subjects were taken inside the collection room and seated in a chair. The height of the chair was adjusted so that the head of the subject was between camera `11_0` and camera `24_0`. We used two live monitors attached to cameras `11_0` and `05_1` to ensure correct head location of the subjects throughout the recording procedure. In each session, multiple image sequences were recorded, for which subjects were instructed to display different facial expressions. Subjects were shown example images of the various expressions from the Cohn-Kanade database [7] immediately prior to the recording. Table 2 lists the expressions captured in each session. Figure 6 shows example images for all facial expressions contained in the database.

For each camera 20 images were captured within 0.7 seconds: one image without any flash illumination, 18 images with each flash firing individually, and then another image without any flash illumination. Taken across all cameras a total of 300 images were captured for each sequence. See Figure 7 for a montage of all 15 camera views shown with frontal flash illumination. Unlike in the previous PIE database [13] the room lights were left on for all recordings. Flash-only images can be obtained through simple image differencing of flash and non-flash images as shown in Figure 8. Due to the rapid acquisition of the flash images subject movement between images is neglectible.

4. Database Statistics

In total, the Multi-PIE database contains 755,370 images from 337 different subjects. Individual session attendance varied between a minimum of 203 and a maximum of 249

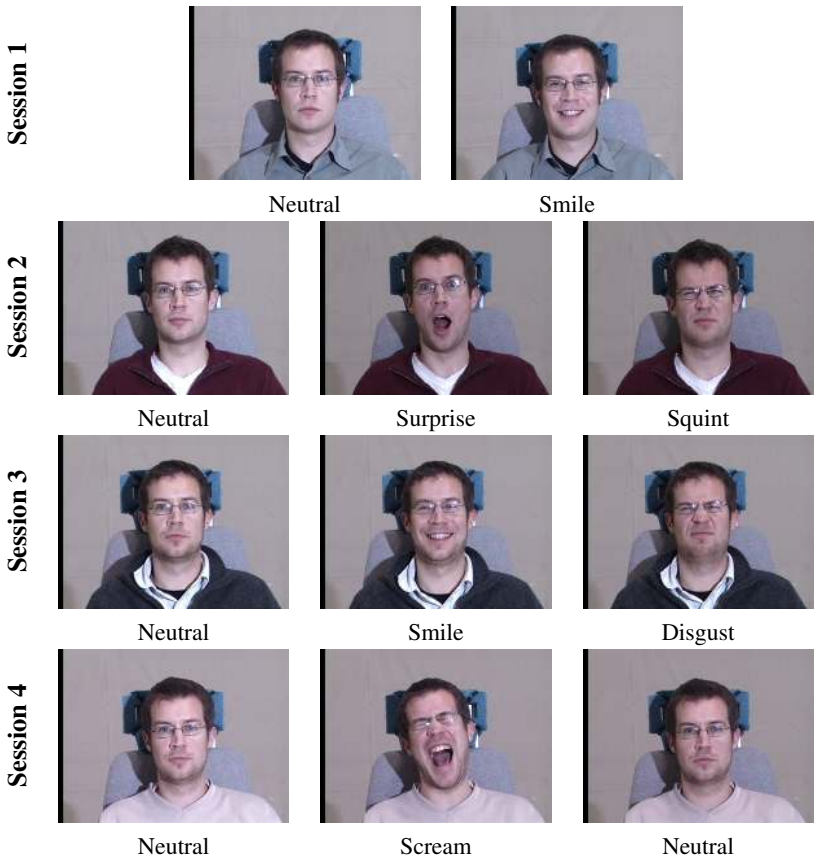


Figure 6. Example images of the facial expressions recorded in the four different sessions. The images shown here were recorded by the camera directly opposite the subject with the flash attached to said camera illuminating the scene.

Figure 7. Montage of all 15 cameras views in the CMU Multi-PIE database, shown with frontal flash illumination. 13 of the 15 cameras were located at head height with two additional cameras mounted higher up to obtain views typically encountered in surveillance applications. The camera labels are shown in each image (see Figure 3).

Expression	S1	S2	S3	S4
Neutral	x	x	x	xx
Smile	x		x	
Surprise		x		
Squint		x		
Disgust			x	
Scream				x

Table 2. Overview of the facial expressions recorded in the different sessions. Note that we recorded two neutral expressions during session four, one before and one after the scream expression.

Figure 8. Computation of flash-only images as difference between flash and non-flash images.

Individual Session Attendance			
Session 1	Session 2	Session 3	Session 4
249	203	230	239
Repeat Recordings			
4 Sessions	$\geq$ 3 Sessions	$\geq$ 2 Sessions	1 Session
129	191	264	73

Table 3. Attendance statistics for the different recording sessions of the Multi-PIE database. 264 of the 337 subjects were recorded at least twice.

subjects. Of the 337 subjects 264 were recorded at least twice and 129 appeared in all four sessions. See Table 3 for details.

The subjects were predominantly men (235 or 69.7% vs. 102 or 30.3% females). 60% of subjects were European-Americans, 35% Asian, 3% African-American and 2% others. The average age of the subjects was 27.9 years. As part of the distribution we make the following demographic information available: gender, year of birth, race and whether the subject wears glasses.

5. Baseline Recognition Results

To illustrate the similarities and differences between the PIE and Multi-PIE databases we report results of baseline experiments with PCA [14] and LDA [1] classifiers, both using a cosine distance measure.[1] We describe the evaluation procedure in Section 5.1 and show results of comparative experiments on PIE and Multi-PIE in Section 5.2. Section 5.3 presents results of new experiments on Multi-PIE that could not be conducted using PIE data.

[1]For face PCA spaces, the whitened cosine distance measure used here has been shown to perform well [2]. For LDA, the optimal distance measure appears to depend on the specific dataset [4].

5.1. Evaluation Procedure

For all experiments, frontal faces were normalized using the location of 68 manually established facial feature points. These points are triangulated and the image warped with a piecewise affine warp onto a coordinate frame in which the canonical points are in fixed locations. This process is similar to the preprocessing used prior to the computation of Active Appearance Models (AAMs) [3, 10]. The resulting images are approximately 90×93 in size (with slight variations for the different data subsets). Throughout we use the data of 14 subjects (20% of the 68 subjects available in PIE) to compute the PCA or LDA subspaces and evaluate performance on the remaining subjects. In all cases we report rank-1 accuracy rates.

Results are reported as averages over 20 independent random assignments of subjects to training and testing sets. In the experiments comparing performance on PIE and Multi-PIE we show results for matched conditions using 68 subjects from each database (labeled as "PIE 68" and "M-PIE 68") as well as results using the full set of subjects available in Multi-PIE (labeled as "M-PIE Full").

5.2. Comparing PIE and Multi-PIE

5.2.1 Recognition across Sessions

The Multi-PIE database contains up to four sessions per subject recorded over a span of six months (see Table 3) whereas subjects were seen only once in the PIE database. As a consequence we can report recognition accuracies as function of time between the acquisition of gallery and probe images (here for neutral expression faces without flash illumination). Figure 9 shows the recognition rates for both PIE and Multi-PIE using a PCA recognizer. For PIE, the probe and gallery images are identical, resulting in perfect recognition. For Multi-PIE, we recorded two neutral expression images in session 4, enabling a within-session test (for time difference 0). Across sessions, recognition accuracies drop with increasing time difference and increasing testing set size (M-PIE 68 vs. M-PIE full).

5.2.2 Recognition across Illumination

In the illumination experiments we use images recorded without flashes as gallery (in the case of PIE from the recording with room lights on) and all flash images in turn as probe. Figure 10 shows recognition accuracies for both PCA and LDA on PIE and Multi-PIE across all illuminations. The physical setup of light sources used in PIE and Multi-PIE is comparable. As a consequence, for matched

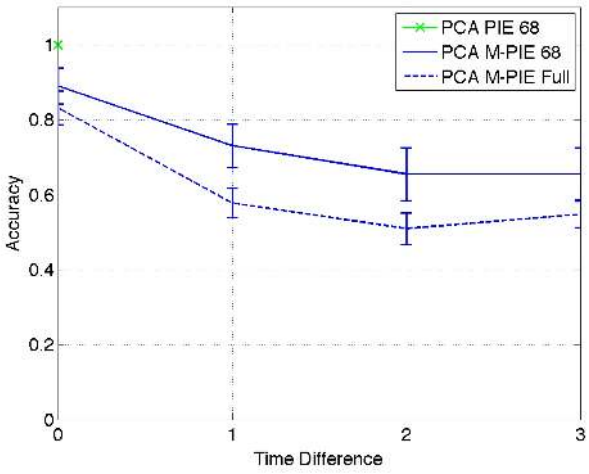


Figure 9. PCA performance for PIE and Multi-PIE across recording sessions. Since PIE only contains images from one session, gallery and probe images are identical, resulting in perfect recognition (PIE 68). For Multi-PIE, accuracies decrease with increasing time difference between the acquistition of gallery and probe images. We show results for a 68 subject subset of Multi-PIE (M-PIE 68) as well as for the full set of available subjects (M-PIE Full).

experimental conditions (PCA PIE 68 in Figure 10(a) and PCA M-PIE 68 in Figure 10(b)), accuracies are nearly identical (36.6% vs. 35.4%). LDA performance saturates over PIE (95%, LDA PIE 68), whereas accuracies on Multi-PIE with the much larger test set of subjects still leaves room for improvement (71.3%, LDA M-PIE Full).

5.3. Beyond PIE

For the most part, PIE supports single factor experiments (e.g. recognition across pose or recognition across illumination). The data in Multi-PIE enables a range of new experiments examining cumulative effects of multiple recording conditions which can not be conducted using PIE data. As examples we show results for recognition across both illumination and sessions in Section 5.3.1 and across expressions and illumination in Section 5.3.2.

5.3.1 Recognition across Illumination and Sessions

The availability of illumination data from multiple sessions enables us to evaluate recognition performance across illumination *and* sessions. Figure 11 shows the performance of PCA and LDA classifiers on the task. Similar to the results in Section 5.2.1 performance decreases with increasing time difference between acquisition of gallery and probe images, at a much lower performance level though than in Figure 9 due to the influence of the illumination differences.

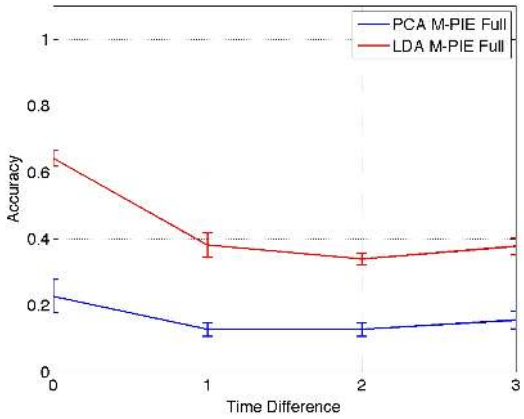


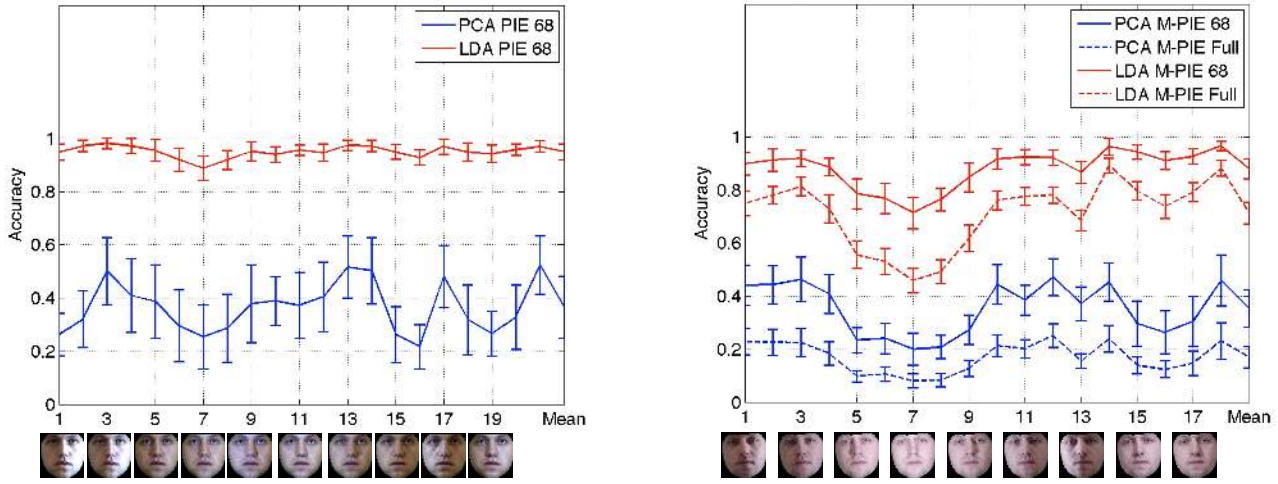
Figure 11. PCA and LDA performance on Multi-PIE across illumination and sessions. Results shown are averages over all illumination conditions. Performance decreases with increasing time difference between the recording of gallery and probe images. Performance overall is lower than in Figure 9 due to the influence of the illumination differences.

5.3.2 Recognition across Expression and Illumination

The range of facial expressions captured in Multi-PIE (*neutral*, *smile*, *surprise*, *squint*, *disgust*, and *scream*) is much larger than the subtle expressions contained in PIE (*neutral*, *smile*, *blink*, and *talk*). Furthermore, Multi-PIE contains images from all illuminations conditions for all facial expressions. We are therefore able to evaluate the cumulative effect of changes in illumination *and* expression on recognition accuracies. Figure 12 shows PCA and LDA accuracies for different probe expressions, averaged over all illumination conditions. In all cases, a neutral expression image recorded in the same session without flash illumination was used as gallery image. As comparison we also show results of PCA recognition with identical illumination conditions for gallery and probe (PCA M-PIE). The combined influence of illumination and expression reduces accuracies drastically, with PCA rates varying between 13.7% (for *scream*) and 21.1% (for *squint*). LDA accuracies are higher on average (41.4% vs. 18.5%), peaking at 50.1% (again for *squint*).

6. Availability

Multi-PIE is available to all interested researchers for the cost of media (a 400GB hard drive) and shipping. Details of the distribution procedure along with a reference set of experiments will be published on the database website at `http://multipie.org`.



(a) PIE (b) Multi-PIE

Figure 10. Comparison of PCA and LDA recognition across illumination conditions in PIE and Multi-PIE. For matched experimental conditions (PCA PIE 68 in (a) and PCA M-PIE 68 in (b)), performance is comparable, experimentally veryifying the similarity in the physical setup of the two collections. Whereas LDA performance over PIE nearly saturates at 95%, the average accuracy over Multi-PIE using the largest test set (LDA M-PIE Full) indicates further room for improvement.

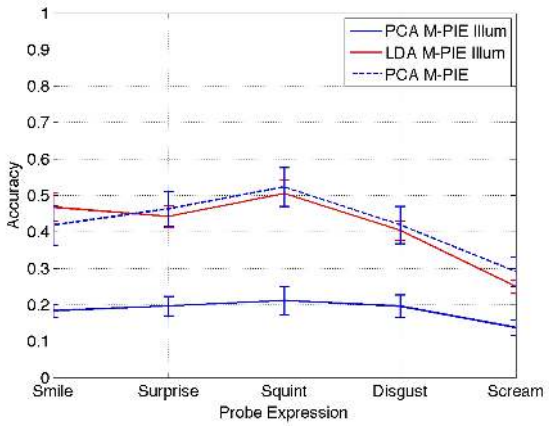


Figure 12. PCA performance on Multi-PIE across expressions and illuminations. We use the neutral images (without flash illumination) recorded in the same session as gallery and the expression images under all illumination conditions as probe. The combined influence of illumination and expression reduces accuracies drastically, with PCA rates varying between 13.7% (for *scream*) and 21.1% (for *squint*). LDA accuracies are higher on average (41.4% vs. 18.5%), peaking at 50.1% (again for *squint*). As comparison we also show PCA recognition rates for identical gallery and probe illumination conditions (labeled "PCA M-PIE").

7. Conclusion

In this paper we introduced the CMU Multi-PIE face database. Multi-PIE improves upon the highly successful PIE database in a number of aspects: a larger set of subjects, more recording sessions, more facial expressions, and the inclusion of high resolution images. We reported results of baseline experiments using PCA and LDA classifiers discussing both the similarities and as well as the differences between the two databases. All experiments shown here only used frontal face images. In future work we plan on expanding the evaluations across pose as well.

Acknowledgment

We would like to thank Jonathon Phillips, James Wayman, and David Harrington for discussions and support. We furthermore would like to thank Athinodoros Georghiades and Peter Belhumeur for providing us with the setup details of the Yale "flash system." Collection of the Multi-PIE database was supported by United States Government's Technology Support Working Group (TSWG) through award N41756-03-C-402 and by Sandia National Laboratories.

References

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

[2] R. Beveridge, D. Bolme, B. Draper, and M. Teixeira. The CSU face identification evaluation system. *Machine Vision and Applications*, 16:128–138, 2005.

[3] T. Cootes, G. Edwards, and C. Taylor. Active appearance models. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 23(6), 2001.

[4] K. Delac, M. Grgic, and S. Grgic. Independent comparative study of PCA, ICA, and LDA on the FERET data set. *International Journal of Imaging Systems and Technology*, 15(5):252–260, 2005.

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

[6] R. Gross. Face databases. In S. Li and A. Jain, editors, *Handbook of Face Recognition*. Springer Verlag, 2005.

[7] T. Kanade, J. Cohn, and Y.-L. Tian. Comprehensive database for facial expression analysis. In *Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition*, pages 46–53, 2000.

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

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

[10] I. Matthews and S. Baker. Active appearance models revisited. *International Journal of Computer Vision*, 60(2):135–164, 2004.

[11] 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 (AVBPA)*, 1999.

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

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

[14] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.