The CMU Pose, Illumination, and Expression Database

Terence Sim, Simon Baker, and Maan Bsat

Corresponding Author: Simon Baker

The Robotics Institute
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213

simonb@cs.cmu.edu

Abstract

In the Fall of 2000 we collected a database of over 40,000 facial images of 68 people. Using the CMU 3D Room we imaged each person across 13 different poses, under 43 different illumination conditions, and with 4 different expressions. We call this the CMU Pose, Illumination, and Expression (PIE) database. We describe the imaging hardware, the collection procedure, the organization of the images, several possible uses, and how to obtain the database.

1 Introduction

People look very different depending on a number of factors. Perhaps the 3 most significant factors are: (1) the pose; i.e. the angle at which you look at them, (2) the illumination conditions at the time, and (3) their facial expression; e.g. smiling, frowning, etc. Although several other face databases exist with a large number of subjects [Philips *et al.*, 1997], and with significant pose and illumination variation [Georghiades *et al.*, 2000], we felt that there was still a need for a database consisting of a fairly large number of subjects, each imaged a large number of times, from several different poses, under significant illumination variation, and with a variety of expressions.

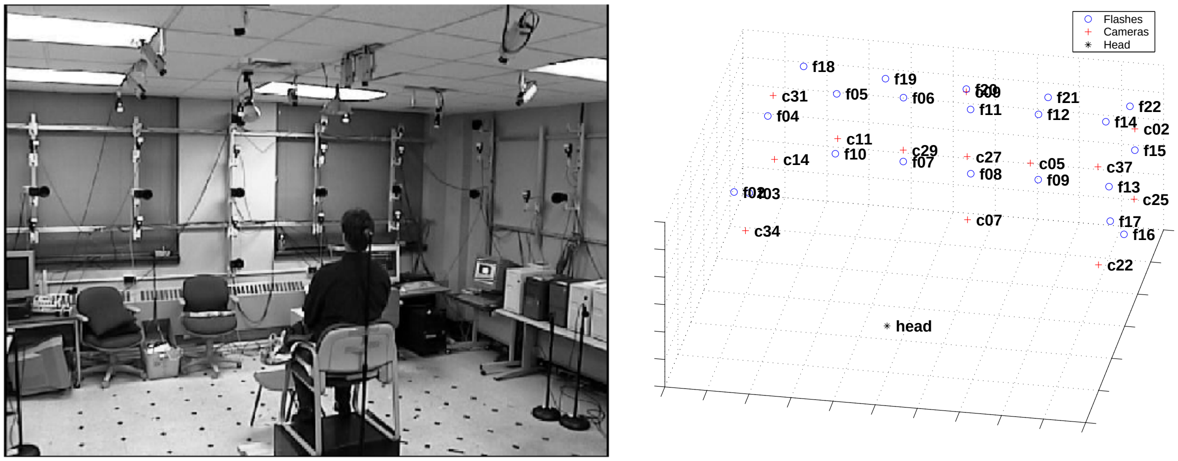
Between October 2000 and December 2000 we collected such a database consisting of over 40,000 images of 68 subjects. (The total size of the database is about 40GB.) We call this the CMU Pose, Illumination, and Expression (PIE) database. To obtain a wide variation across pose, we used 13 cameras in the CMU 3D Room [Kanade *et al.*, 1998]. To obtain significant illumination variation we augmented the 3D Room with a "flash system" similar to the one constructed by Athinodoros Georghiades, Peter Belhumeur, and David Kriegman at Yale University [Georghiades *et al.*, 2000]. We built a similar system with 21 flashes. Since we captured images with, and without, background lighting, we obtained $21 \times 2 + 1 = 43$ different illumination conditions. Furthermore, we asked the subjects to pose with several different expressions.

In the remainder of this paper we describe the capture hardware, the organization of the images, a large number of possible uses of the database, and how to obtain a copy of it.

2 Capture Apparatus

2.1 Setup of the Cameras: Pose

Obtaining images of a person from multiple poses requires either multiple cameras capturing images simultaneously, or multiple "shots" taken consecutively (or a combination of the two.) There are a number of advantages of using multiple cameras: (1) the process takes less time, (2) if the cameras are fixed in space, the (relative) pose is the same for every subject and there is less difficulty in positioning the subject to obtain a particular pose, (3) if the images are taken simultaneously we know that the imaging conditions (i.e. incident illumination, etc) are the same. This



(a) (b)

Figure 1: (a) The setup in the CMU 3D Room [Kanade *et al.*, 1998]. The subject sits in a chair with his head against a pole to fix the head position. We used 13 Sony DXC 9000 (3 CCD, progressive scan) cameras with all gain and gamma correction turned off. We augmented the 3D Room with 21 Minolta 220X flashes controlled by an Advantech PCL-734 digital output board, duplicating the Yale “flash dome” [Georghiades *et al.*, 2000]. (b) The xyz-locations of the head position, the 13 cameras, and the 21 flashes plotted in 3D. These locations were measured with a Leica theodolite and are included in the meta-data.

final advantage can be particularly useful for detailed geometric and photometric modeling of objects. On the other hand, the disadvantages of using multiple cameras are: (1) we actually need to possess multiple cameras, digitizers, and computers to capture the data, (2) the cameras need to be synchronized: the shutters must all open at the same time and we must know the correspondence between the frames, and (3) the cameras will all have different intrinsic parameters.

Setting up a synchronized multi-camera imaging system is quite an engineering feat. Fortunately, such a system already existed at CMU, namely the 3D Room [Kanade *et al.*, 1998]. We reconfigured the 3D Room and used it to capture multiple images simultaneously across pose. Figure 1 shows the capture setup in the 3D Room. There are 49 cameras in the 3D Room, 14 very high quality (3 CCD, progressive scan) Sony DXC 9000’s, and 35 lower quality (single CCD, interlaced) JVC TK-C1380U’s. We decided to use only the Sony cameras so that the image quality is approximately the same across the database. Due to other constraints we were only able to use 13 of the 14 Sony cameras. This still allowed us to capture 13 poses of each person simultaneously.

We positioned 9 of the 13 cameras at roughly head height in an arc from approximately full left profile to full right profile. Each neighboring pair of these 9 cameras are approximately $22.5°$ apart. Of the remaining 4 cameras, 2 were placed above and below the central camera c27, and 2 were placed in the corners of the room, where surveillance cameras are typically located.

The locations of 10 of the cameras can be seen in Figure 1(a). The other 3 are symmetrically opposite the 3 right-most cameras in the figure. We measured the locations of the cameras using a theodolite. The measured locations are shown in Figure 1(b) and are included in the meta-data.

2.2 The Flash System: Illumination

To obtain significant illumination variation we extended the 3D Room with a “flash system” similar to the Yale Dome used to capture the data in [Georghiades *et al.*, 2000]. With help from Athinodoros Georghiades and Peter Belhumeur, we used an Advantech PCL-734, 32-channel digital output board to control 21 Minolta 220X flashes. The Advantech board can be directly wired into the “hot-shoe” of the flashes. Generating a pulse on one of the output channels then causes the corresponding flash to go off. We placed the Advantech board in one of the 17 computers used for image capture in the 3D Room and integrated the flash control code into the image capture routine so that the flash, the duration of which is approximately 1ms, occurs while the shutter (duration approximately 16ms) is open. We then modified the image capture code so that one flash goes off in turn for each image captured. We were then able to capture 21 images, each with different illumination, in $21/30 \approx 0.7$sec. The locations of the flashes, measured with a theodolite, are shown in Figure 1(b) and included in the database meta-data, along with with camera locations.

In the Yale illumination database [Georghiades *et al.*, 2000] the images are captured with the room lights switched off. The images in the database therefore do not look entirely natural. In the real world, illumination usually consists of an ambient light with perhaps one or two point sources. To obtain representative images of such cases (that are more appropriate for determining the robustness of face recognition algorithms to illumination change) we decided to capture images both with the room lights on and with them off. We decided to include images with the lights off to give some partial overlap with the database used in [Georghiades *et al.*, 2000].

3 Database Contents

On average the entire capture procedure took about 10 minutes per subject. In that time, we captured (and retained) over 600 images from 13 poses, with 43 different illuminations, and with 4 expressions. The color images have size 640×486. (The first 6 rows of the images contain synchronization information added by the VITC units in the 3D Room [Kanade *et al.*, 1998]. This

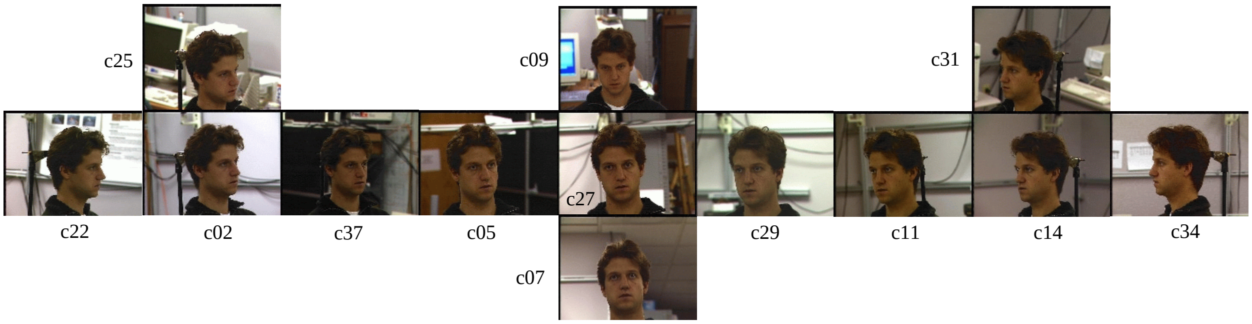


Figure 2: An illustration of the pose variation in the PIE database. The pose varies from full left profile to full frontal and on to full right profile. The 9 cameras in the horizontal sweep are each separated by about 22.5°. The 4 other cameras include 2 above and 2 below the central camera, and 2 in the corners of the room, typical locations for surveillance cameras. See Figure 1 for the camera locations.

information may be discarded.) The storage required per person is approximately 600MB using color "raw PPM" images. Thus, the total storage requirement for 68 people is around 40GB. This can of course be reduced by compression, but we did not do so in order to preserve the original data.

3.1 Pose Variation

An example of the pose variation in the PIE database is shown in Figure 2. This figure contains images of 1 subject from each of the 13 cameras. As can be seen, there is a wide variation in pose from full profile to full frontal. This subset of the data should be useful for evaluating the robustness of face recognition algorithms across pose. Since the camera locations are known, it can also be used for the evaluation of pose estimation algorithms. Finally, it might be useful for the evaluation of algorithms that combine information from multiple widely separated views.

3.2 Pose and Illumination Variation

Examples of the pose and illumination variation are shown in Figure 3. Figure 3(a) contains the variation with the room lights on and Figure 3(b) with the lights off. Comparing the images we see that those in Figure 3(a) appear more natural and representative of images that occur in the real world than those in Figure 3(b). On the other hand, the data with the lights off was captured to reproduce the Yale database used in [Georghiades *et al.*, 2000]. This will allow a direct comparison between the 2 databases. We foresee a number of possible uses for the pose and illumination

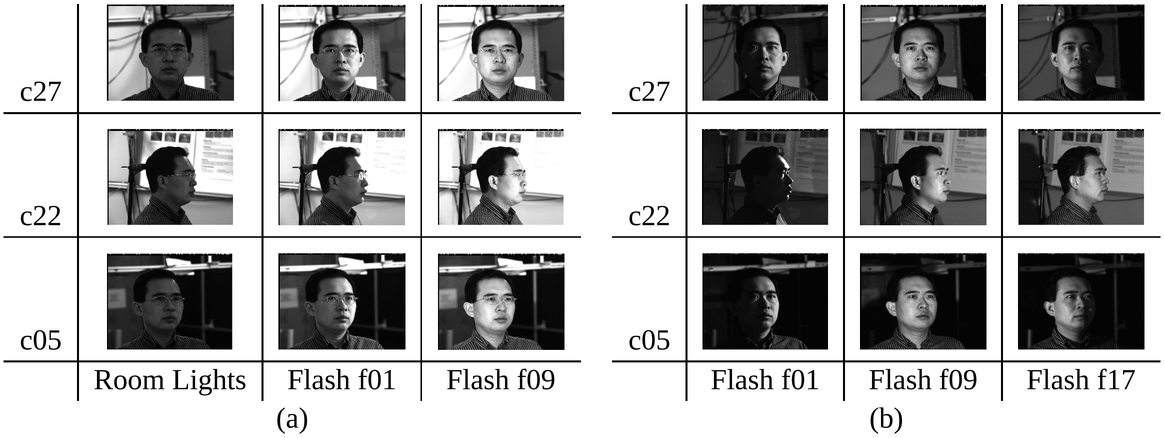


Figure 3: Examples of the pose and illumination variation with (a) the room lights on, and (b) the room lights off. Notice how the combination of room illumination and flashes leads to much more natural looking images than with just the flash alone.

variation data: First, it can be used to reproduce the results in [Georghiades *et al.*, 2000]. Second, it can be used to evaluate the robustness of face recognition algorithms to pose and illumination.

3.3 Pose and Expression Variation

An example of the pose and expression variation is shown in Figure 4. The subject is asked to provide a neutral expression, to smile, to blink (i.e. to keep their eyes shut), and to talk. For neutral, smiling, and blinking, we kept 13 images, 1 from each camera. For talking, we captured 2 seconds of video (60 frames). Since this occupies a lot more space, we kept this data for only 3 cameras: the frontal camera c27, the three-quarter profile camera c22, and the full profile camera c05. In addition, for subjects who usually wear glasses, we collected an extra set of 13 images without their glasses, asking them to put on a neutral expression.

The pose and expression variation data can be used to test the robustness of face recognition algorithms to expression (and pose.) The reason for including blinking was that many face recognition algorithms use the eye pupils to align a face model. It is therefore possible that these algorithms are particularly sensitive to subjects blinking. We can now test whether this is indeed the case.

3.4 Meta-Data

We collected a variety of miscellaneous “meta-data” to aid in calibration and other processing:

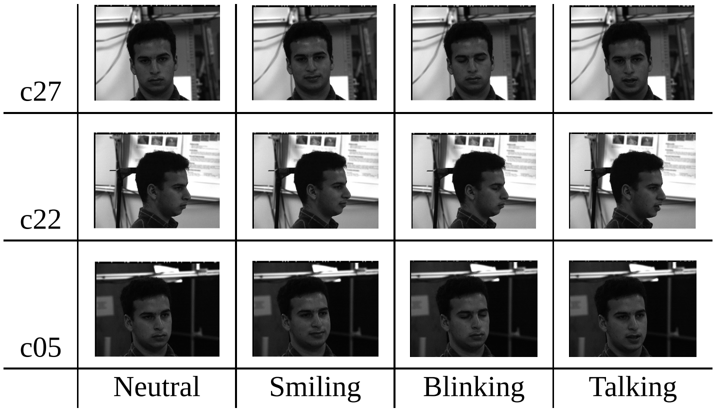


Figure 4: An example of the pose and expression variation in the PIE database. Each subject is asked to give a neutral expression, to smile, to blink, and to talk. We capture this variation across all poses. For the neutral images, the smiling images, and the blinking images, we keep the data from all cameras. For the talking images, we keep 60 frames of video from only 3 cameras (frontal c27, three-quarter profile c05, and full profile c22). For subjects who wear glasses we also capture, from all cameras, one set of neutral-expression images of them without their glasses.

Head, Camera, and Flash Locations: Using a theodolite, we measured the xyz-locations of the head, the 13 cameras, and the 21 flashes. See Figure 1(b). The values 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, we captured a background image from each of the 13 cameras. These images can be used to help localize the face region.

Color Calibration Images: Although the cameras that we used are all of the same type, there is still a large amount of variation in their photometric responses, both due to their manufacture and due to the fact that the aperture settings on the cameras were all set manually. We did perform “auto white-balance” on the cameras, but there is still some noticeable variation in their color response. To allow the cameras to be intensity- (gain and bias) and color-calibrated, we captured images of color calibration charts.

Personal Attributes of the Subjects: Finally, we include some personal information about the 68 subjects. For each subject we record the subject’s sex and age, the presence or absence of eye glasses, mustache, and beard, as well as the date on which the images were captured.

4 Potential Uses of the Database

Throughout this paper we have pointed out a number of potential uses of the database. We now summarize some of the possibilities, citing several papers that have already used the database:

- Evaluating pose invariant face detectors [Heisele *et al.*, 2001a, Heisele *et al.*, 2001b].
- Evaluation of head pose estimation algorithms.
- Evaluation of the robustness of face recognition algorithms to the pose of the probe image [Gross *et al.*, 2001, Blanz *et al.*, 2002].
- Evaluation of face recognition algorithms that operate across pose; i.e. algorithms for which the gallery and probe images have different poses [Gross *et al.*, 2002, Blanz *et al.*, 2002].
- Evaluation of face recognizers that use multiple images across pose [Gross *et al.*, 2002].
- Evaluation of the robustness of face recognition algorithms to illumination (and pose) [Sim and Kanade, 2001, Gross *et al.*, 2001, Blanz *et al.*, 2002].
- Evaluation of the robustness of face recognition algorithms to facial expression and pose.
- 3D face model building either using multiple images across pose (stereo) or multiple images across illumination (photometric stereo [Georghiades *et al.*, 2000]).

The importance of the evaluation of algorithms (and the databases used to perform the evaluation) for the development of algorithms should not be underestimated. It is often the failure of existing algorithms on new datasets, or simply the existence of new datasets, that drives research forward.

5 Obtaining the Database

We have been distributing the PIE database in the following manner:

1. The recipient ships an empty (E)IDE hard drive to us.
2. We copy the data onto the drive and ship it back.

To date we have shipped the PIE database to over 50 research groups worldwide. Anyone interested in receiving the database should contact the second author by email at `simonb@cs.cmu.edu` or visit the PIE database web site at `www.ri.cmu.edu/projects/project_418.html`.

Acknowledgements

We would like to thank Athinodoros Georghiades and Peter Belhumeur for giving us details of the Yale “flash dome.” Sundar Vedula and German Cheung gave us great help using the CMU 3D Room. We would also like to thank Henry Schneiderman and Jeff Cohn for discussions on what data to collect. Financial support for the collection of this database was provided by the U.S. Office of Naval Research (ONR) under contract N00014-00-1-0915. A version of this paper appeared in the 5th International Conference on Automatic Face and Gesture Recognition [Sim *et al.*, 2002]. We thank both the PAMI reviewers and the reviewers of [Sim *et al.*, 2002] for their feedback.

References

[Blanz *et al.*, 2002] V. Blanz, S. Romdhani, and T. Vetter. Face identification across different poses and illuminations with a 3D morphable model. In *Proceedings of the 5th IEEE International Conference on Face and Gesture Recognition*, 2002.

[Georghiades *et al.*, 2000] A.S. Georghiades, P.N. Belhumeur, and D.J. Kriegman. From few to many: Generative models for recognition under variable pose and illumination. In *Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition*, 2000.

[Gross *et al.*, 2001] R. Gross, J. Shi, and J. Cohn. Quo vadis face recognition. In *Proceedings of the 3rd Workshop on Empirical Evaluation Methods in Computer Vision*, 2001.

[Gross *et al.*, 2002] R. Gross, I. Matthews, and S. Baker. Eigen light-fields and face recognition across pose. In *Proceedings of the 5th IEEE International Conference on Face and Gesture Recognition*, 2002.

[Heisele *et al.*, 2001a] B. Heisele, T. Serre, M. Pontil, and T. Poggio. Component-based face detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2001.

[Heisele *et al.*, 2001b] B. Heisele, T. Serre, M. Pontil, T. Vetter, and T. Poggio. Categorization by learning and combining object parts. In *Neural Information Processing Systems*, 2001.

[Kanade *et al.*, 1998] 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, Carnegie Mellon University Robotics Institute, 1998.

[Philips *et al.*, 1997] P.J. Philips, H. Moon, P. Rauss, and S.A. Rizvi. The FERET evaluation methodology for face-recognition algorithms. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1997.

[Sim and Kanade, 2001] T. Sim and T. Kanade. Combining models and exemplars for face recognition: An illuminating example. In *Proceedings of the Workshop on Models versus Exemplars in Computer Vision*, 2001.

[Sim *et al.*, 2002] T. Sim, S. Baker, and M. Bsat. The CMU pose, illumination, and expression (PIE) database. In *Proceedings of the 5th IEEE International Conference on Face and Gesture Recognition*, 2002.