The Florence 2D/3D Hybrid Face Dataset

Andrew D. Bagdanov
bagdanov@dsi.unifi.it

Alberto Del Bimbo
delbimbo@dsi.unifi.it

Iacopo Masi
masi@dsi.unifi.it

Media Integration and Communication Center
University of Florence, Italy
http://www.micc.unifi.it/

ABSTRACT

This article describes a new dataset under construction at the Media Integration and Communication Center and the University of Florence. The dataset consists of high-resolution 3D scans of human faces along with several video sequences of varying resolution and zoom level. Each subject is recorded under various scenarios, settings and conditions. This dataset is being constructed specifically to support research on techniques that bridge the gap between 2D, appearance-based recognition techniques, and fully 3D approaches. It is designed to simulate, in a controlled fashion, realistic surveillance conditions and to probe the efficacy of exploiting 3D models in real scenarios.

Categories and Subject Descriptors

H.5.1 [**Multimedia Information Systems**]: Video

General Terms

Experimentation

Keywords

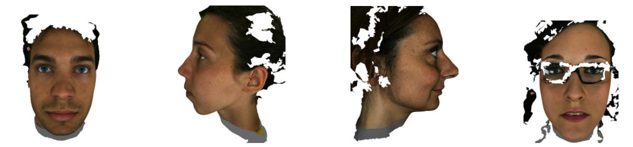
Facial analysis, 3D, face recognition, face retrieval, datasets

1. INTRODUCTION

Human faces are one of the most important biometrics for recognition. Face imagery is easily and non-intrusively collectible, whereas other biometrics such as fingerprints or iris scans are impractical to implement in many scenarios (e.g. a surveillance setting). Because of the universality of faces as a biometric, there has been a proliferation of face recognition approaches proposed in the research literature [4]. Along with the proliferation of algorithms for face recognition there has also been an explosion of datasets designed to support research in face recognition.

Face detection and recognition remains a difficult problem. Much of this difficulty is due to challenging imaging conditions and variations caused by expressions, gender and pose. More recently, 3D scanning technology has matured and the cost of entry is much less. This has led to renewed interest in face recognition using 3D models of human faces. One unexplored avenue of research on facial analysis is the potential of using 3D models to augment the performance of traditional 2D, appearance-based techniques. One of the primary motivations for developing the Florence 2D/3D Face Dataset was the recent advances in tracking with PTZ cameras [2] and also due to our long-term research goal of performing biometric face recognition of tracked, uncooperative targets at a distance.

J-HGBU'11, December 1, 2011, Scottsdale, Arizona, USA.
ACM 978-1-4503-0998-1/11/12.



(a) Frontal (b) Left Side (c) Right Side (d) Glasses

Figure 1: Examples of each type of 3D model for four random subjects.

In this paper we describe the Florence 2D/3D face dataset that is currently under preparation [1]. The dataset has been specifically designed to support research across a range of face analysis and recognition tasks. In section 2 we describe the structure and organization of the Florence 2D/3D Face Dataset and finally we conclude in section 3 with a discussion of the direction our work will take on this new 2D/3D face dataset.

2. DESIGN OF THE DATASET

Our dataset is designed with two main goals in mind. First, we would like to make available accurate and complete 3D models of faces to researchers who are primarily interested in the analysis of 3D meshes and textures of human faces. That is, our dataset is designed to be useful for research on pure 3D analysis techniques.

Second, however, we have designed our dataset as something that goes beyond the scope of 3D analysis techniques, allowing researchers to investigate the possibility of reducing the gap between 2D[2] computer vision algorithms and those methods that work on more precise, 3D models. In particular, our dataset is thought of in the context of evaluating the use of 3D information in computer vision problem like 3D face pose estimation [3] and 3D face recognition [1] *directly from video data or still images.*

[1]Visit `http://www.micc.unifi.it/masi/research/ffd/` for a preview of the dataset. The dataset is scheduled to be completed by Fall 2011.

[2]When we use the term "2D" we refer to video or image sequences of database subjects.

To this end, the pipeline of data acquisition is designed to provide both 3D data and 2D videos consistent with each other. First, a 3D model is captured of the subject using a 3D scanner. Second, we record high definition (HD) video of the subject as he simulates specific head rotations (this corresponds to a cooperative environment). Five levels of zoom are used in order to capture the subject face at multiple image resolutions. After this, the subject is then recorded from two PTZ Cameras, one indoor and the other outdoor. These two scenarios represent a more non-cooperative subject and he is asked to be spontaneous. Three levels of zoom are captured in each video in order to cover a broad range of face resolutions.

The Florence 2D/3D Face Dataset is currently a continuous work in progress and consists of about 50 adult subjects as of October 1st, 2011. The aim is to complete the dataset by Fall 2011 with a total of 100+ subjects.

2.1 3D Face Models

In terms of 3D data, our dataset contains a minimum of three high-resolution 3D models (both meshes and textures) for each subject. Some example models are shown in figure 1. The models are summarized below:

- Two frontal models with the head in approximately the same pose. This is designed so that one may be used for training and the other for testing. See figure 1(a) for an example
- One left lateral model, which covers approximately the whole side of the face: from mid-nose to the back of the head. See figure 1(b) for examples.
- One right lateral model, with the same properties as above. See figure 1(c) for examples.
- If the subject wears glasses, we also capture a 3D model with glasses. See figure 1(d).

All 3D models are provided in three different formats: OBJ, PLY and VRML. They contain the mesh and the high resolution texture.

The geometry of the mesh is extremely accurate: on average about 0.2mm of RMS error in the reconstruction or better, depending on the exact pre-calibration and configuration[3]. Considering this, our 3D models are very accurate and precise in the number, yielding on average a mesh of about 40,000 vertices about 80,000 facets. The texture acquired is a stereo image with a resolution of 3341×2027 pixels.

2.2 2D Face Imagery

Once 3D models are acquired, the face of the subject is recorded in several 2D videos in order to obtain the same face as seen from a camera. In particular, we record three different videos in sequence. The video sequences are designed to provide video across a range of standard camera qualities and resolutions. They are also recorded at multiple levels of zoom in order to investigate the 2D face resolution necessary to support accurate facial analysis. Finally, the imaging scenarios used are designed to simulate and indoor environment with a cooperative subject, and indoor environment with an uncooperative subject, and an outdoor environment with an uncooperative subject.

The three type of videos recorded for each subject are defined as follows:

- One indoor HD video (1280×720 pixel resolution) of a cooperative subject recorded at 4 levels of zoom. To generate a uniform source of light without highlights, the subject is illuminated frontally by two lamps. The subject is asked to generate some out-of-plane head rotations, viewing six points: top-right, top-left, middle-right, middle-left, bottom-right, bottom-left. The frame rate for this video is about 20 fps and is acquired with an AXIS Q1755 HD camera.
- One indoor video (704×576 - 4CIF resolution) of a an uncooperative subject recorded with a PTZ camera using 3 levels of zoom. Here the subject is asked to be spontaneous. The frame rate for this video is about 20 fps and is acquired with an AXIS PTZ Q6032-E.
- One outdoor video (736×544 pixel resolution) of uncooperative subject from a PTZ camera using 3 levels of zoom. Here the subject is again asked to be spontaneous, but this time the recorded video is very challenging because of the uncontrolled lighting conditions and the presence of shadows and highlights. The frame rate for this video is about 5-7 fps and it is acquired with a SONY RZ30-P camera.

3. DISCUSSION AND CONCLUSIONS

By far the most common type of dataset available to support research on facial analysis algorithms are standard, 2D datasets consisting of multiple, static images of each subject. Pure 3D datasets for facial analysis are starting to appear, though an unexplored area of research remains the potential intersection of 2D and 3D facial analysis algorithms. This is the niche in which we have positioned the Florence 2D/3D Face Dataset. The Florence dataset is, at the time of writing, still a work in progress. It is scheduled to be a complete resource of 100+ subjects by fall of 2011. Our hypothesis is that 3D information in the form of high-resolution models of subjects can be useful for a range of facial analysis techniques that are classically very challenging using only traditional 2D imagery.

4. REFERENCES

[1] S. Berretti, A. Del Bimbo, and P. Pala. 3D face recognition using isogeodesic stripes. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 32(12):2162–2177, December 2010.

[2] A. Del Bimbo, F. Dini, G. Lisanti, and F. Pernici. Exploiting distinctive visual landmark maps in pan-tilt-zoom camera networks. *Computer Vision and Image Understanding (CVIU)*, 114(6):611 – 623, 2010.

[3] G. Fanelli, J. Gall, and L. van Gool. Real time head pose estimation with random regression forests. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'11)*, June 2011.

[4] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Computing Surveys*, 35:399–458, December 2003.

[3]To capture each 3D model we used a state-of-the-art scanner "3dMDface System".