IDIAP COMMUNICATION REPORT

MOBIO DATABASE FOR THE ICPR 2010 FACE AND SPEECH COMPETITION

Chris McCool Sébastien Marcel

Idiap-Com-02-2009

NOVEMBER 2009

Centre du Parc, Rue Marconi 19, P.O. Box 592, CH - 1920 Martigny
T +41 27 721 77 11 F +41 27 721 77 12 info@idiap.ch www.idiap.ch

MOBIO Database for the ICPR 2010 Face and Speech Competition

Christopher McCool and Sébastien Marcel

Idiap Research Institute, PO Box 592, CH-1920 Martigny, Switzerland
christopher.mccool@idiap.ch or sebastien.marcel@idiap.ch

Abstract. This document presents an overview of the mobile biometry (MOBIO) database. This document is written expressly for the face and speech organised for the 2010 International Conference on Pattern Recognition.

1 Introduction

The mobile biometry (MOBIO) database was captured as part of the MOBIO project [1]. This project covers the use of two main forms of biometry for mobile authentication, these being: face and speech. To this end the MOBIO database, a multi-modal face and speech database, was captured to reflect potential real-world scenarios for face and speech authentication on a mobile device.

PhaseI of the database consists of six sessions for 160 participants. Each session consists of 21 recordings which are described in more detail below.

2 Data Collection

The MOBIO database was captured at six separate sites in five different countries. These sites are at the: University of Manchester (UMAN), University of Surrey (UNIS), Idiap Research Institute (IDIAP), Brno University of Technology (BUT), University of Avignon (LIA) and University of Oulu (UOULU).

The database is being acquired primarily on a mobile phone. To address the concerns of both the speech and face researchers the

[1] Funded by the European Commission's (EC's) Seventh Framework Programme (FP7), MOBIO project grant number 214324. For more information please visit http://www.mobioproject.org.

participants were asked to answer a set of 21 questions which varied from (1) set responses, (2) read speech from a paper through to (3) free speech.

1. **Set responses** were given to the user. In total there were five such questions and **fake responses** were supplied to each user. The five questions asked were:
   (a) **What is your name?**
   (b) **What is your address?**
   (c) **What is your birth date?**
   (d) **What is your credit card number?**
   (e) **What is your driver's licence number?**
   and each question took approximately five seconds to answer (although this varies between users).
   Example file: unis/m233/01_mobile/m233_01_p03_i0_0.mp4

2. **Read speech** was obtained from each user by supplying the user with three sentences to read. The sentences were the same for each session and is reproduced below.

   "*I have signed the MOBIO consent form and I understand that my biometric data is being captured for a database that might be made publicly available for research purposes.*

   I understand that I am solely responsible for the content of my states and my behaviour.

   I will ensure that when answering a question I do not provide any personal information in response to any question."

   Example file: uman/m108/01_mobile/m108_01_l11_i0_0.mp4

3. **Free speech** was obtained from each user by prompting the user with a random question. For five of these questions the user was asked to speak for five seconds and for ten questions the user was asked to speak for ten seconds, this gives a total of fifteen such questions. The user was again asked to not provide personal information and it was even suggested to not answer the question used to prompt them provided they could speak for the required time.

Example file: uoulu/m301/01_mobile/m301_01_f14_i0_0.mp4

The collected files are all named according to a particular filename structure. The filename structure is as follows:

PersonID_Recording_ShotNum_Conditions-Channel.mp4

where,
PersonID = Gender + Institute + ID
Recording = Session
ShotNum = Speech Type + Shot
Conditions = Environment + Device
Channel = ChannelID

and
Institute: 0=Idiap, 1=Manchester, 2=Surrey, 3=Oulu, 4=Brno, 5=Avignon
Gender: m=Male, f=Female
ID: from 01 to 99 for each site
Session: ID from 01 to 99
Speech Type: p= set response, l= read speech, r= short free speech or f= long free speech
Shot: ID from 01 to 99
Environment: i=Inside, o=Outside
Device: 0=Mobile, 1=Laptop
ChannelID: ID 0 to 9 (0 - first video/audio channel, 1 - second video/audio channel)

3 Experimental Protocol

The database is split into three distinct sets: one for training, one for development and one for testing. The splitting is such that two sites (in totality) are used for one split, this means that there is no information regarding the individuals or the conditions for a site between sets. For the training set the data can be used in any way deemed appropriate and all of the data is available see Table 1; normally the training set data would be used to derive background models (for instance training a world background model UBM). The

development set can be used to derive fusion parameters, however, it must be used to derive a threshold that is then applied to the test data. To facilitate this the development split and the test split both have the same style of protocol defined for them.

Training Splits		
Session number	Usage	Questions to use
Session 1	Background model training	**All data**
Session 2	Background model training	**All data**
Session 3	Background model training	**All data**
Session 4	Background model training	**All data**
Session 5	Background model training	**All data**
Session 6	Background model training	**All data**

Table 1. Table describing the usage of data for the Training split of the database.

The protocol for the development split and the test split are the same. The first session is used to enrol the user but only the five set response questions can be used for enrollment, see Table 2. Testing is then conducted on each individual file for sessions two to six (there are five sessions used for development/testing) and only the free speech questions are used for testing. This leads to five enrollment videos for each user and 75 test client (positive sample) videos for each user (15 from each session). When producing impostor scores all the other clients are used, for instance if in total there were 50 clients then the other 49 clients would perform an impostor attack. For clarity the enrollment procedure and testing procedure are described again below.

- **Enrollment** data consists of the five **set response** recordings from the first session of the particular user.
- **Testing** data comes from the **free speech** recordings from every other session (the other five sessions) of the users, each video is treated as a separate test observation.

4 Acknowledgements

This work has been performed by the MOBIO project 7th Framework Research Programme of the European Union (EU), grant agreement number: 214324. The authors would like to thank the EU for

Development and Testing Splits		
Session number	Usage	Questions to use
Session 1	Enrollment	**Set questions only**
Session 2	Test Scores	**Free speech only**
Session 3	Test Scores	**Free speech only**
Session 4	Test Scores	**Free speech only**
Session 5	Test Scores	**Free speech only**
Session 6	Test Scores	**Free speech only**

Table 2. Table describing the usage of data for the Testing and Development splits of the database.

the financial support and the partners within the consortium for a fruitful collaboration. For more information about the MOBIO consortium please visit http://www.mobioproject.org. The authors would also like to thank Phil Tresadern (University of Manchester), Bastien Crettol (Idiap Research Institute), Norman Poh (University of Surrey), Christophe Levy (University of Avignion), Driss Matrouf (University of Avignon), Timo Ahonen (University of Oulu), Honza Cernocky (Brno University of Technology) and Kamil Chalupnicek (Brno University of Technology) for their work in capturing this database and development of the protocol.