IDIAP RESEARCH REPORT

MOBIO: MOBILE BIOMETRIC FACE AND SPEAKER AUTHENTICATION

Sébastien Marcel Chris McCool
Cosmin Atanasoaei Flavio Tarsetti Jan Pesan
Pavel Matejka Jan Cernocky Mika Helistekangas
Markus Turtinen

Idiap-RR-31-2010

AUGUST 2010

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: Mobile Biometric Face and Speaker Authentication

Sébastien Marcel, Chris McCool, Cosmin Atanasoaei, Flavio Tarsetti
Idiap Research Institute
1920 Martigny, Switzerland
marcel@idiap.ch, christopher.mccool@idiap.ch, cosmin.atanasoaei@idiap.ch

Jan Pešán, Pavel Matějka, Jan Černocký
Brno University of Technology
61266 Brno, Czech Republic
xpesan00@stud.fit.vutbr.cz,matejkap@fit.vutbr.cz,cernocky@fit.vutbr.cz

Mika Helistekangas, Markus Turtinen
Visidon Ltd
90570 Oulu, Finland
mika.helistekangas@visidon.fi,markus.turtinen@visidon.fi

Abstract

This paper presents a mobile biometric person authentication demonstration system. It consists of verifying a user's claimed identity by biometric means and more particularly using their face and their voice simultaneously on a Nokia N900 mobile device with its built-in sensors (frontal video camera and microphone).

1. Introduction

Face and speaker recognition are both mature fields of research. For instance, face recognition has been explored since the mid 1960's. Speaker recognition by humans has been done since the invention of the first recording devices, but automatic speaker recognition has only been a topic of research since the 1970's. Most of these research activities have been carried out in environments with "controlled" setups and a priori conditions to better study their influence on the recognition process.

The mobile environment offers a natural combination of challenging conditions to operate including adverse illumination, noisy background and noisy audio data. The face and speaker recognition system shortly described in this paper provides a unique opportunity to analyze two mature biometrics side by side in a mobile environment. This is one of the direct outcomes of the Mobile Biometry (MOBIO) European Project (http://www.mobioproject.org).

This paper presents a mobile biometric person authentication demonstration system. It consists of verifying/authenticating a user's claimed identity by biometric means and more particularly using their face and their voice simultaneously on a mobile device with its built-in sensors (frontal video camera and microphone).

Figure 1. A capture of the MOBIO face and speaker authentication demonstration system running on the Nokia N900.

The mobile device chosen for this demonstration is a Nokia N900 as it has several requirements needed for performing this task such as: a frontal VGA camera, a fast CPU (600Mhz ARM Cortex-A8), a floating-point unit, 256MB of memory, and an open operating system (Maemo – a Linux environment) with development tools. Although this device has been selected, the source code is written to be cross-platform (C++) and can be used on any operating system such Symbian OS, Android or iPhone/iPad OS.

2. MOBIO Face and Speaker Authentication

The MOBIO demonstration system is based on two independent face and speech sub-systems to perform face authentication (FA) and speaker authentication (SA).

2.1. Face Processing Sub-System

The MOBIO face sub-system is decomposed in two steps: Face Localization (FL) and Face Authentication (FA). Face localization is achieved by a cascade of classifiers based on a variant of Local Binary Patterns (LBP) [1] as implemented in [5]. Many different ways to address the problem of face recognition have been proposed [6] but it still remains a very challenging task due to changes in the appearance caused by facial expressions, poses, lighting and aging of the subjects.

We have chosen a simple yet efficient technique for FA based on histograms of Local Binary Patterns (LBP) and a parts based topology [1]. First, the gray-scale face image is normalized to a size of 64×80 pixels (width $\times$ height). Then the normalized image is divided into 8×8 non-overlapped blocks and for each block a histogram of Local Binary Patterns (LBP) is formed. During enrollment, these histograms are stored as a face template, then during test, these histograms are compared to the stored template using histogram intersection.

It should be noted that both the FL and FA modules have been implemented in fixed-point arithmetic.

2.2. Speech Processing Sub-System

The MOBIO speech sub-system consists of two parts: Voice Activity Detection (VAD) and Speaker Authentication (SA).

The role of VAD is to select only speech for the following processing and to discard silence and noises. Our VAD currently employs a simple energy measure, the next version will be based on a neural network phone recognizer. As features, classical Mel-Frequency Cepstral Coefficients (MFCC) with deltas and double deltas (approximating the 1st and 2nd derivation) and short-time Gaussianisation are used.

Although many advanced techniques for SA have recently been proposed [2, 3], the MOBIO SA sub-system is based on one of the most prevalent techniques for the task: the Gaussian Mixture Model (GMM) paradigm that uses a Universal Background Model (UBM) [4]. In this paradigm, a UBM is trained on a set of independent speakers. Then a client is enrolled by adapting from this UBM using the speaker specific data. When testing, two likelihoods are produced, one for the UBM and one for the client specific model: these two scores are combined using the log-likelihood ratio and compared to a threshold to produce a "client/impostor" decision.

2.3. The N900 Mobile Device

Based on extensive review, the Nokia N900 Linux device has been selected as the most suitable device for the MOBIO demonstration system.

The ARM Cortex A8, used in the N900, has 2 Floating Point Units (FPU), a non-pipelined VFP-lite and a pipelined SIMD NEON coprocessor. The VFP-lite can handle both single and double precision arithmetic, as well as properly handling exceptions and subnormal numbers. However, probably due to the full specification compliance and presence of the NEON, it is a relatively slow implementation in the A8, usually taking between 18-21 cycles to perform a single precision multiply accumulate. The NEON unit on the other hand is designed for very fast single precision vector math, it can sustain multiply accumulates at a rate of two per cycle. Although the FPUs are present, the best performance on the mobile device is achieved by converting floating-point arithmetics into the fixed-point integer calculations.

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 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`.

References

[1] T. Ahonen, A. Hadid, and M. Pietikainen. Face Recognition with Local Binary Patterns. *Lecture Notes in Computer Science*, pages 469–481, 2004.

[2] W. Campbell, D. Sturim, D. Reynolds, and A. Solomonoff. Svm based speaker verification using a gmm supervector kernel and nap variability compensation. *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings*, 1:I–I, 2006.

[3] P. Kenny, P. Ouellet, N. Dehak, V. Gupta, and P. Dumouchel. A study of inter-speaker variability in speaker verification. In *IEEE Transactions on Audio, Speech and Language Processing*, July 2008.

[4] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn. Speaker verification using adapted gaussian mixture models. *Digital Signal Processing*, 10(1-3):19 – 41, 2000.

[5] Y. Rodriguez. *Face Detection and Verification using Local Binary Patterns*. PhD thesis, EPFL, 2006.

[6] X. Tan, S. Chen, Z.-H. Zhou, and F. Zhang. Face recognition from a single image per person: A survey. *Pattern Recognition*, 39(9):1725–1745, 2006.