

# Audio-Visual Features for Biometric Recognition of Identity and Digits

MAYCEL ISAAC FARAJ

Thesis to be defended in public at 14.15 on Friday, December 7,  
2007, in the Wigfors room at Halmstad University, Halmstad,  
Sweden, for the degree of Doctor of Philosophy.

The thesis will be defended in English.

Faculty opponent: Josef Kittler, School of Electronics and  
Physical Sciences, University of Surrey, Guildford, United  
Kingdom.



Department of Signals  
and Systems  
Chalmers University of  
technology  
(+)46 31 772 1000



School of Information Science,  
Computer and Electrical  
Engineering  
Halmstad University  
(+)46 35 167 10

## ABSTRACT

Biometric recognition systems have been established as a powerful security tool to prevent unknown users from entering high-risk systems and areas. They are increasingly being utilized in surveillance and access management (city centers, banks, etc.) by using individuals' physical or biological characteristics. The present study reports on the use of lip motion as a stand-alone biometric modality as well as a modality integrated with audio speech for identity and digit recognition.

First, we estimate motion vectors from a sequence of lip-movement images. The motion is modelled as the distribution of apparent line velocities in the movement of brightness patterns in an image. Then, we construct compact lip-motion features from the regional statistics of the local velocities. These can be used alone or merged with audio features to recognize individuals or speech (digits). Furthermore, we present results when digit recognition is used in a text prompted recognition system to verify the *liveness* of the user. This is useful to reduce audio-based replay attacks on the recognition system.

In this work, we utilized two classifiers for identification and verification of identity as well as with digit recognition. Although the study is focused on processing lip movements in a video sequence, significant speech processing is a prerequisite given that the contribution of video analysis to speech analysis is studied in conjunction with recognition of humans and what they say (digits). Such integration is necessary to understand multimodal biometric systems to the benefit of recognition performance and robustness against noise. We present results using the XM2VTS database representing the video and audio data of 295 people.

Keywords: Biometrics, lip motion, audio-visual signals, speech recognition, speaker recognition, digit recognition.