Content-Based Representation of Sign Language Video Sequences

Nariman Habili

B.Sc., B.Eng. (Flinders)

A thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

in the

School of Electrical and Electronic Engineering

The University of Adelaide

Australia

September, 2002

Contents

Abstract

Sign language is a visual language used by deaf or hearing-impaired people to communicate. For distant communication, deaf people commonly use the text telephone, which is at least 10 times slower than sign language. Moreover, sign language is the first language of many pre-lingually deaf individuals, and its speed is comparable to that of normal speech. Video communication would allow deaf individuals to communicate remotely via sign language, providing them the equivalent of the telephone for individuals of normal hearing. Therefore, video communication would be a boon to the deaf community.

Block-based video coding strategies, the cornerstone of the H.261 and H.263 coding standards for video conferencing, are unsuitable for the transmission of sign language video over affordable low bit-rate channels. This is mainly due to the presence of rapid hand and arm motion in sign language video, as well as the necessity of smooth motion perception. Accordingly, sign language video will require content-based coding strategies to achieve the image quality and frame rate necessary for accurate perception. Using content-based coding, video sequences are typically segmented into different objects which may be independently coded and transmitted. More resources are allocated to the perceptually important objects, which in the case of sign language, are the face and hands.

In this thesis, a methodology is devised for the segmentation of the face and hands in sign language video sequences. As well as an improved coding performance, the content-based representation of video data would allow other functionalities, such as improved error-robustness and scalability. The proposed algorithm employs color and motion cues to segment the face and hands. First, a color segmentation algorithm is devised to locate skin-color regions in each frame. Second, we note that sign language is characterized by the motion of the hands and the face. Based on this observation, the proposed face and hand segmentation

methodology employs motion information to locate the moving skin-color regions in each frame. To this end, a statistical change detection method is proposed based on the F test and block-based motion estimation. In addition to the face and hand segmentation methodology, a face detection and temporal tracking method is also presented. This has applications in lip-reading, where more coding resources are allocated to the face.

The performance of the skin-color segmentation algorithm is demonstrated by simulations carried out on both still images and video sequences. The proposed change detection method is tested on four video sequences. The simulation results demonstrate the effectiveness of the proposed face and hand segmentation methodology, and the face detection and tracking method.