

Chapdelaine, C., Beaulieu, M. and L. Gagnon. "Designing Caption Production Rules Based on Face, Text and Motion Detections" In *Proceedings of IS&T/SPIE's 20th Annual Symposium Electronic Imaging 2008*. San Jose, California, USA, January 27-31, 2008.

Abstract

Producing off-line captions for the deaf and hearing impaired people is a labor-intensive task that can require up to 18 hours of production per hour of film. Captions are placed manually close to the region of interest but it must avoid masking human faces, texts or any moving objects that might be relevant to the story flow. Our goal is to use image processing techniques to reduce the off-line caption production process by automatically placing the captions on the proper consecutive frames. We implemented a computer-assisted captioning software tool which integrates detection of faces, texts and visual motion regions. The near frontal faces are detected using a cascade of weak classifier and tracked through a particle filter. Then, frames are scanned to perform text spotting and build a region map suitable for text recognition. Finally, motion mapping is based on the Lukas-Kanade optical flow algorithm and provides MPEG-7 motion descriptors. The combined detected items are then fed to a rule-based algorithm to determine the best captions localization for the related sequences of frames. This paper focuses on the defined rules to assist the human captioners and the results of a user evaluation for this approach.

Keywords: E-accessibility, Eye-tracking, TV captioning, Image processing, Video object detection.