David Nester
Nikita Haduong, Preethi Vaidyanathan, Emily Prud'hommeaux, Reynold Bailey, Cecilia O. Alm

Project Title:

Capturing, Analyzing, and Aligning Multimodal Responses to Affective Visual Content

Abstract of Project:

Humans routinely extract important information from images and videos. Advances in computer vision techniques are gradually allowing automated systems to approach human-level performance in identifying visual content. However, computational systems still have difficulty annotating important information in visual data in a human-like manner. To address this challenge, we co-captured human gaze, spoken language, and facial expressions in an experiment. Subjects described complex images and videos with abstract positive and negative content and also responded to how they felt about them. Patterns in these sensed modalities were explored to characterize the human reactions. In addition, we expanded a framework to generate automatic alignments between the visual and spoken modalities for visual annotation of complex images, which is a difficult task due to the varying temporal offset of the modalities. We also explored the effect of sparse data and abundant data on the alignment process, in each case performing above baseline comparisons, and provided additional visualizations of the resulting annotations from multimodal alignment. Removing units with sparse data further resulted in a substantial decrease in alignment error rate. This work has implications for areas such as image understanding, media accessibility, and multimodal data fusion.