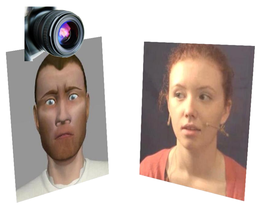
Recognizing Human Facial Emotions in Video: A Psychologically-Inspired Fusion Model

Albert Cruz and Bir Bhanu - {acruz,bhanu}@ee.ucr.edu

Center for Research in Intelligent Systems, University of California, Riverside, CA 92521

Applications and Motivation

- Applications in medicine (treatment of Asperger's), video games (Xbox Kinect), human-computer interaction (intelligent tutoring systems) and affective computing (embodied agents).

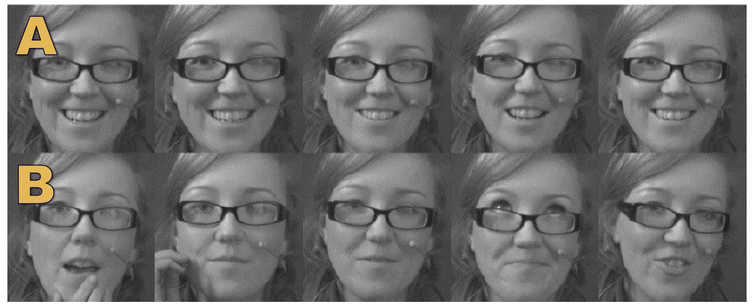


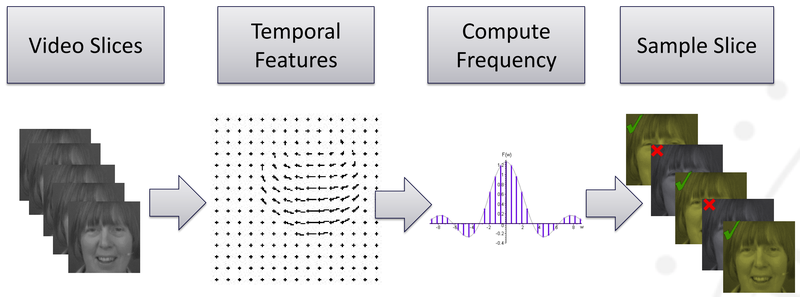
- Affective Computing: The emergent field where computers interpret your emotion and respond in kind with emotion of their own. Example online [3].

- **Technical Challenges to State-of-the-Art:**
  - Can not align faces with extreme pose or occlusion.
  - Systems process the face at each frame. This does not scale to application (not fast enough for real-time interaction).
- **Psych-Inspired Motivation:** Humans perform well at this task so a system needs to emulate behavior of the human visual system.
- **Project Goal:** Design a system that can recognize emotions, properly align faces despite facial dynamics and emulate processes in the human visual system to improve performance.

Vision and Attention Theory Sampling

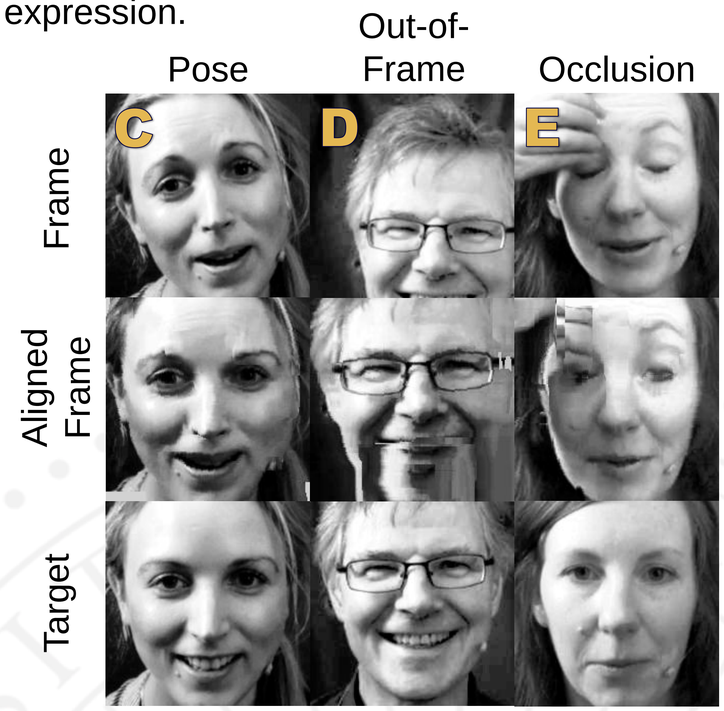
- The HVS selectively processes a scene with a 50ms-1s latency inversely proportional to the change in visual information.
- The pause between processing a new frame should be inversely proportional to the amount of facial motion.
  - An idle subject needs a single frame to describe the expression (A) whereas an actively expressing subject needs many frames (B).





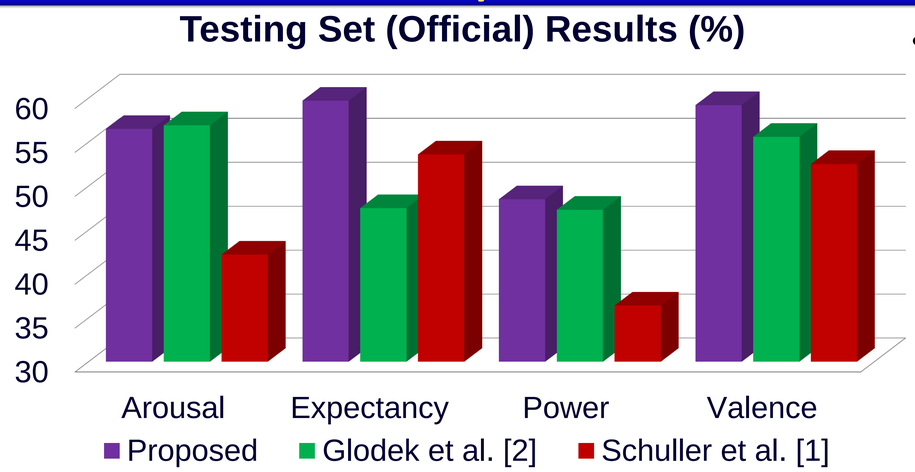
Precision Face Alignment

- Current face alignment algorithms fail for extreme pose, when the subject is out of frame and cases of occlusion.
- We overcome this by warping the image with SIFT-Flow. Facial features are translated to be precisely aligned while maintaining original expression.



- Current algorithms would struggle to give results for the above frames.

Audio/Visual Emotion Challenge 2011 Results



- Proposed approach was submitted to the Audio/Visual Emotion Challenge 2011. Improves results by an average of 10% over Schuller *et al.* [1].

Conclusion

- Concepts in perceptual psychology used to selectively process frames.
- Precision face alignment can align faces while maintaining of facial expressions.

References:

[1] Schuller et al. AVEC 2011 – the first international audio/visual emotion challenge. In Proc. Affective Computing and Intelligent Interaction, 2011.
[2] M. Glodek et al. Multiple classifier systems for the classification of audio-visual emotional states. In Proc. Affective Computing and Intelligent Interaction, 2011.
[3] G. McKeown. Chatting with a virtual agent: The semaine project character spike. Website, February 2011. http://www.youtube.com/watch?v=6KZc6e_EuCg.

Acknowledgement:
This work was supported in part by NSF grant 0727129 and NSF IGERT: Video Bioinformatics Grant DGE 0903667. Contents and information do not reflect the position or policy of the U.S. Government.