Team Smiley

Rene Garcia, Huijun Kim, Clint Ascencio,

Dory Glauberman, Tomas Gonzalez

Final Project – 12/2

Overview

Our project topic is real-time human emotion tracking. Our main goal was to be able to detect basic human emotions such as happiness, sadness, and neutrality based on real-time video of a single human face. Our specific objectives were as follows:

1) Use computer's webcam to implement real-time face tracking.
   Result: Detect human faces when they appear in the webcam's frame.

2) Detect eyes and mouth from a face in a live video
   Result: Track the eyes and mouth as separate nodes (collection of points) on a single face in live video.

3) Recognize basic emotions from a face in a live video
   Result: Recognize happiness, sadness, and neutrality from a single face in a live video.

We were easily able to detect faces, eyes and mouths in a video frame, but choosing the best face and it's actual eyes and mouth were a challenge due to the numerous false positives that the haarcascade classifiers turned up. Once we had these boundaries, we were able to turn the desired regions (eyebrows and mouth) into point clusters and used OpenCV's optical flow implementation to determine the direction of their movement. Next, based on our own research as well as research by Yaser Yacoob and Larry S. Davis that is discussed more in-depth in the

"Background" section, we decided how to associate these point movements with the three basic emotions, and report these emotions to the user in real time.

Background

We used optical flow for real-time emotion recognition based on research by Yacoob and Davis from their 1994 paper "Recognizing Human Facial Expression." While their research was completed over 20 years ago now, it is still considered state of the art in the field. New research in the field still uses the principles behind optical flow, which is tracking the movement of points in a video. This method applies to emotion recognition by tracking the motion of key points on a human face. For example, think of a smile or a frown, points on the lips move upward and outward or upward and inward respectively, if you focus on the direction the points in these regions move, it is a simple conclusion to say that when the majority of the points are moving upward the face is smiling and a human looking at the video would recognize a smile as happiness and therefore say the human in the video is happy, the opposite is true for recognizing sadness. For many emotions it is not as simple as tracking the corners of the mouth, and even for happiness and sadness you can gain more accurate results by considering eyebrow

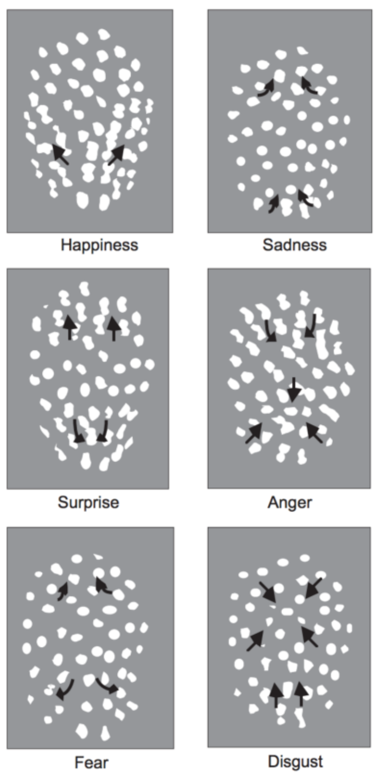


Figure 5: The cues for facial expression as suggested by Bassili.

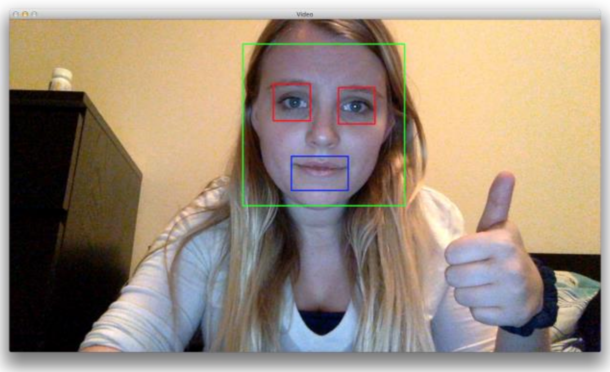
movement. See Figure 5 from Yacoob and Davis' research on the movement of points for six key emotions (above).

Recent, more advanced research on the topic of emotion recognition from video still use the optical flow method, but they use the information gathered from it to train a support vector machine. A 2012 academic journal printed research by Saumil Srivastava in real time emotion recognition. He used optical flow to track the movement of key facial points and then used that information to train a support vector machine to recognize particular point movements as certain emotions (i.e. happiness, sadness, anger, etc.). After researching support vector machines as used in research by Srivastava's and at the Machine Perception Laboratory at The University of California - San Diego, we decided it would be too difficult to implement within the time constraints for this project. We opted instead to set benchmarks based on observations from the optical flow calculations, a more detailed description of our implementation process can be found in the "Methods" section of this report.

Methods

Face Tracking

We relied heavily on haarcascade classifiers for face and facial feature tracking. Two of the cascade classifiers we used were provided by OpenCV, "haarcascade_frontalface_alt.xml" and "haarcascade_eye.xml", the third "haarcascade_mcs_mouth.xml" was created by Modesto Castrillón Santana. The frontal face classifiers is very good at finding faces in the frame, so when we present it with a frame that has a single face against a plain background in good lighting conditions, we were able to find that face without any modifications. The eyes and mouth were less accurate. For the eyes, we found that most of the time, the first two eyes it detects are the correct eyes, instead of the false positives it finds such as the nostrils. The mouth was the hardest thing to detect because the haarcascade classifier was not very robust and thought a lot of things were mouths, even



forehead wrinkles. We used common sense facts about the layout of a face to pick out the right mouth from the list of mouths the classifier returned. We decided that the top of the mouth has to be a certain distance below the bottom of the eyes and that the bottom of the mouth should be above the bottom of the face (results pictured above). Eyebrows are used in Yacoob and Davis' research, but we found we were able to get accurate results by only using the mouth.

Optical Flow

The definition of optical flow according to Dr. Klingner's slides is "the apparent motion of brightness patterns in the image." Which in ideal cases, such as constant lighting can detect the motion of points in a video. We are using OpenCV's implementation to calculate the optical flow frame by frame. We are isolating which points we care about so that we are only displaying the optical flow lines for the eyes and the mouth. We are converting each frame of the video to gray scale to do the optical flow calculations as it allows for the difference in brightness patterns to represent the motion field as accurately as possible. The direction of the flow lines in the mouth is what we use to determine emotion.

Emotion Recognition

We experimented with several methods of using the optical flow calculation to determine emotion. We did so based on observational experiments of determining which direction the majority of the points move in while a test subject (one of our group members) is in the frame making a smile, neutral face, and sad face. We initially thought that the important points for the mouth were the corners and assumed they moved upward/downward based on a smile or frown, but this proved to be ineffective and unsupported by our observational tests. Yacoob and Davis had a better idea, they say that for happiness the majority of the points in the mouth region are moving upward and outward, and for sadness the points in the mouth region are still moving upward but inward. We use the endpoints of the lines given from the optical flow calculation to determine how many lines meet those criteria and whichever is greater determines the emotion. Neutrality is easier and more difficult, it's easy because it can be detected by very few points of motion. However, your mouth isn't moving when you are holding a smile or frown, only when you are entering one so lack of movement is not enough to determine neutrality. This led us to realize that the key to why determining emotion is so much easier with video than static images is that what we are really detecting is a change in emotion. So we decided to keep track of the current emotional state, always beginning with neutral, and use that to determine what the relative optical flow calculations mean for emotion. If the face is in a neutral state and the lines move upward and inward then it is transitioning into a state of happiness, when the face stops smiling, the lines move similarly to how they would for a frown, but since we know that the current state is happiness, we know that this transition is back to neutrality instead of sadness, the opposite is true for transition from sadness back to neutrality. Due to time constraints, our implementation requires that happiness and sadness can only be recognized when transitioning from a neutral state.

Results

We were able to successful determine whether a single face in a video is displaying three basic human emotions: happiness, sadness, or neutrality. We did so by using optical flow to track the movement of the mouth, which is a key facial region for determining these basic emotions (Note, more complex emotions requiring taking into account the movement of more facial regions such as the eyebrows and nose). We report these results in real time feedback displayed on the video frame. Several factors determine the quality of our results. We get the most accurate facial feature detection (and therefore emotion recognition) from a video containing a single face on a plain background (such as a white wall) in a well-lit area with unchanging light. Our other limitation is only being able to recognize emotion transitions from a neutral state. Overall, we achieved our main objectives: we can detect a face and key its key facial feature regions, calculate the optical flow of the points in these key regions, and use that data to determine the human emotion being displayed in the video.

Bibliography

Bartlett, Marian S., Gwen Littlewort, Ian Fasel, and Javier R. Movellan. Real Time Face Detection and Facial Expression Recognition: Development and Applications to Human Computer Interaction. Machine Perception Laboratory, Institute for Neural Computation University of California, San Diego, n.d. Web.

Srivastava, Saumil. "Real Time Facial Expression Recognition Using a Novel Method." The International Journal of Multimedia & Its Applications (IJMA) April 4.2 (2012): 49-57. Web.

Yacoob, Yaser, and Larry S. Davis. Recognizing Human Facial Expression. Center For Automation Research University of Maryland, n.d. Web.