Challenges in Recognizing Spontaneous and Intentionally Expressed Reactions to Positive and Negative Images

Jennifer Healey
Adobe Research
San Jose, CA, USA
jehealey@adobe.com

Haoliang Wang
Adobe Research
San Jose, CA, USA
hawang@adobe.com

Niyati Chhaya
Adobe Research
Bangalore, India
nchhaya@adobe.com

Abstract

This paper presents a preliminary exploration of the challenges of automatically recognizing positive and negative facial expressions in both spontaneous and intentionally expressed conditions. Instead of recognizing iconic basic emotion states, which we have found to be less common in typical human computer interaction, we instead attempted to recognize only positive versus negative states. Our hypothesis was that this would prove more accurate if participants intentionally expressed their feelings. Our study consisted of analyzing video from seven participants, each participating in two sessions. Participants were asked to view 20 images, 10 positive and 10 negative, selected from the OASIS image data set. In the first session participants were instructed to react normally, while in the second session they were asked to intentionally express the emotion they felt when looking at each image. We extracted facial action coding units (AUs) from the recorded video and found that on average, intentionally expressed emotions generated 33% more AU intensity across action units associated with both negative emotions (AU1, AU2, AU4 and AU5) and 117% more intensity for AUs associated with positive emotions (AU6 and AU12). We also show that wide variation exists both in average participant responses across images and in individual reactions to images and that simply taking a ration of our identified action units is not sufficient to determine if a response is positive or negative even in the intentionally expressed case.

1. Introduction

In this paper we present the challenges we encountered in trying to develop a robust classifier to automatically recognize positive and negative facial expressions in a naturalistic scenario. Although there has been extensive work on recognizing Ekman's six basic emotions, we have found that in typical multi-modal interactions, surprise, disgust and fear are not often expressed and sadness, anger and enjoyment are typically expressed in a moderate way[11]. We wanted to determine whether it was possible to automatically detect if a person was responding positively or negatively to a computer generated suggestion or turn of a conversation, and if possible to determine the degree of the positive or negative response. We were also interested to know if this task would be easier if the person was intentionally expressing their emotion to a computer. Unlike other work that focuses on recognizing the differences between spontaneous ("real") and intentionally expressed ("fake") emotions [10][9][22][5][4] here we studied both spontaneous and intentionally expressed emotions to determine the differences in positive versus negative recognition in each condition.

Affective facial expression recognition technology has the potential to unlock insights about human experiences at scales that would otherwise be impossible. Unfortunately, depictions in popular media have given rise to the belief that this technology can be used with super human accuracy and without restriction in the wild. This popular belief, coupled with the availability of software packages that allow the easy automatic use of facial action coding technology have led to a legitimate concern that this technology could be used irresponsibly[3][7]. An extensive review of these issues was compiled by Barret et al. [3]. This paper challenges the popular idea that affective facial expression in the wild is a solved problem and documents the challenges we faced in even trying to discriminate a positive versus negative reactions.

The limitations of validity of facial expression technology have been well documented in the academic literature for decades. These include: technical issues such as variations in reliability due lighting conditions and when the face is partially occluded; the necessity for understanding context: situational, physiological and cultural and the question of whether or not emotion itself falls unambiguously into any set of neat categories[20][3][20][13][16][14][12]. Ekman and Friesen, who developed the Facial Action Coding

System (FACS) were well aware that multiple contextual factors could complicate the interpretation of facial reactions, among these: differences in stimulating events (and interpretation of stimulating events); co-occurring events such as speech or intentional non-affective facial gestures (e.g. raising an eyebrow in greeting) as well as subsequent interpersonal behavior including learned coping responses and social masking[12][13][16][14].

At another level, emotion recognition algorithms are also limited by the categories of emotion that they have been trained to recognize, which is in turn limited by the labels in the available training data sets. Although there exist multiple frameworks for understanding emotion, including the continuous framework developed by Russell and Mehrabian [19][21], the majority of labeled emotion data is still codified using Ekman's six basic emotion categories. While these categories are important, they were never intended to be an exhaustive set of all human feelings and expressions, however, if these are the only categories an algorithm knows, this is exactly what the algorithm will assume and it will simply do its best to map all expressions into one of these six categories. Inside a sandbox limited to containing only expressions of these six emotions, such an algorithm may perform with high accuracy, but this will not necessarily be representative of its performance "in the wild."

We wanted to leverage the vast amount of work done on recognizing Ekman's six basic emotions, yet keep the algorithm simple in hope that this would generate robust results in the wild. We noted that the only unambiguously positive positive basic emotion, enjoyment, was characterized by action units AU6 and AU12, which differentiated it from the negative basic emotions and we noted that action units AU1, AU2, AU4 and AU5 were commonly occurring in the negative emotions[14]. This became the small and tractable set of features for the positive and negative emotions that we decided to observe in this analysis.

In our experiment, we exposed people to 20 images, 10 positive and 10 negative, from the OASIS image data set. Images were selected based on the highest and lowest valence scores in the data set and were chosen to inspire strong positive and negative reactions from participants. Given the strength of the stimulus, our results should be seen an upper bound for the kinds of facial reactions people might show in response when viewing more relatively neutral images such as internet based product or service recommendations. While several companies offer services to help evaluate aggregate responses to content over market segments (e.g. Affectiva, Noldus, iMotions and Sitecorp), we instead wanted to look at both differences between spontaneous and intentional reactions to the same content and at individual differences; both in aggregate and in response to individual stimuli. This paper investigates the extent to which that is possible using a simple ratio algorithm on facial action units calculated by open source off the shelf software(OpenFace 2.0 [2][1]). We investigate both the more realistic spontaneous case and the intentionally expressed case, which we believed would be easier to recognize. Figure 2 shows examples of both the spontaneous and the intentionally expressed emotions we wanted to capture as well as the landmarks calculated by OpenFace.

There are currently many different methods for inferring emotion from facial expression. Many of these are based on first calculating facial action units[1] while others use optical flow[15] or deep learning directly from images[18]. There also exist many publicly available data sets of affective facial expression data. While the quantity and diversity of these data sets is increasing there are still limitations on the diversity of the images due to both privacy considerations and the cost of human annotation. As a result, many of these datasets are still limited to more iconic representations of emotion, for example: images or video with professional movies actors, people giving emotionally charged monologues on YouTube or images taken by professional photographers (that may have been filtered for clear emotional expression).

In our experiment, we wanted to focus on how normal people would react to content they liked or disliked on their computer. We wanted see how the same people would react both spontaneously and intentionally to the same set of content. We searched for an image dataset that contained both types of emotional expression and discovered many[8] that contained either spontaneous[25] or posed/acted data [24], but very few with the same participants performing both natural and posed expressions. The closest related data set we were able to identify, that was available to non-academic researchers, was the Natural Visible and Infrared Facial Expression Database for Expression Recognition and Emotion Inference (NVIE) created by Wang et al. [23] which referenced 108 participants contributing both spontaneous and posed facial expressions of the same emotions, however, the original images are not available from this collection and the 72 image feature sets that are available do not contain the per subject spontaneous versus intentionally expressed labels that we were looking to study.

2. Data Collection

Our pilot data collection effort included a convenience sample of seven participants, four female and three male, who were employees of the same company. Participants were told that they were participating in a study of spontaneous and intentionally expressed reactions to a series of images. They were told that they would be filmed and were asked to review and sign a consent form. Participants were then told that they would see a series of images, similar to those shown in Figure 1 and that these images would be separated by a blank slide, also that the slides would advance

Figure 1. Examples of images taken from the OASIS dataset as stimuli. Clockwise from the top left, these are a negative fire image, a positive dog image, a negative garbage dump image and a positive lake image.

automatically. They were told there would be two sessions to the experiment and that they would see identical images in each session. For the first session, participants were told to just react normally to each image, as they would if they were viewing the content online at home. They were told in advance that the experimenter would return after the first session and begin a second session during which they would be asked to intentionally express the emotion each image inspired in them, as they were trying to communicate it to the computer. They were assured that there was no right or wrong response.

2.1. Stimuli

The stimulus consisted of a timed PowerPoint presentation of selected images from the OASIS dataset[17] with blank white slides in between the images. Both the OASIS images and the white slides were timed to be shown for five seconds. To get the highest differentiation in affective reaction, we chose images that had the highest and lowest mean valence ratings in the OASIS dataset, after removing images that we judged might be overly traumatic or sensitive for our participant base (e.g. pictures of an animal carcass, a KKK rally and tumor). We then additionally removed excessively redundant images from the remaining set to create diversity (e.g. fire and explosions on the negative side and lake and beach pictures on the positive side all had similar ratings and were over-represented at the ends of the rating spectrum).

The overall OASIS data set has a valence range of 0.7 (lowest negative) to 6.5 (highest positive). After our rankings and eliminations as previously described, the range of our ten positive images was 6.2 to 6.4 with a mean value

Num.	OASIS ID	N/P	Description	Mean Valence
1	I456	P	Lake	6.39
2	I714	N	Scary face	1.78
3	I496	N	Emaciated Child	1.11
4	I326	N	Fire	1.47
5	I469	P	Lake	6.24
6	I864	N	War	1.76
7	I59	P	Beach	6.37
8	I382	N	Garbage dump	1.64
9	I262	P	Dog	6.29
10	I380	N	Garbage dump	1.60
11	I234	N	Dirt	1.61
12	I616	P	Penguins	6.21
13	I134	P	Cat	6.22
14	I324	N	Fire	1.74
15	I661	P	Rainbow	6.26
16	I60	P	Beach	6.32
17	I209	N	Dead bodies	1.25
18	I466	P	Lake	6.38
19	I871	N	War	1.72
20	I335	P	Fireworks	6.27

Table 1. A listing of the image stimuli as presented to the participants. Here we list the OASIS ID of each image, a description and the mean valence associated with each image. Higher valence numbers indicate more positive ratings.

of 6.3, and a standard deviation 0.07 while range of our negative images was 1.1 to 1.8, with a mean of 1.6 and a standard deviation 0.2. We initially used a random draw to order the images (using randperm in Matlab) to set the order of the images but then kept this order fixed for the entire experiment as listed in Table 1. Although our fixed order could introduce a priming bias, we believe this is mediated by the five second recovery time between images. Given our very small participant size, a fixed order also assured that such our bias would at least be identical across across participants and sessions and that we were not accidentally introducing an highly anomalous outlier ordering by chance (e.g. all positive images followed by all negative images).

2.2. Experimental Set-up

Participants viewed the sequential images on a laptop with a 15 inch screen while being recorded by an iPhone 11 mounted on a tripod. Participants sat at a desk approximately one and a half feet from the screen. The camera and tripod were set up behind the laptop such that the camera lens was relatively in line with the participants head, at approximately two feet horizontal distance. During the experiment we did not constrain the participants head motion or body motion but all participants remained seated during the experiment.

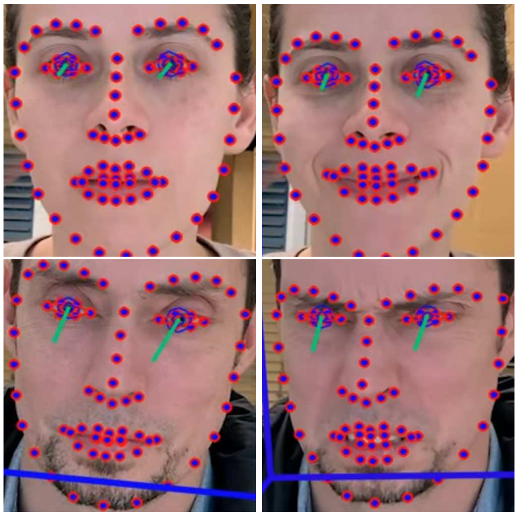

Figure 2. Examples of differences in facial expression in reaction to the same image stimuli in the "natural" condition and intentional expression condition. Clockwise from the top left: a natural reaction to a positive image (Lake), an intentional expressed reaction to the same image, a natural reaction to a negative image (Emaciated Child) and an intentionally expressed reaction to the same image.

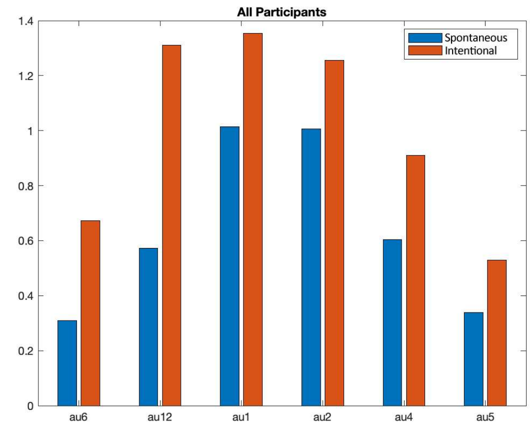

Figure 3. Average AU intensities across all participants for all images. The first two AUs plotted, AU6 and AU12 were associated with a positive emotional response and the remaining four AUs, AU1, AU2, AU4 and AU5 were associated with a negative response. Here the spontaneous average is in the left bar of the pair and the intentional reaction is in the right bar.

3. Analysis Approach

Our hypothesis was that intentionally expressed emotions would be both more intense and more "iconic" than their spontaneous counterparts[6]. To measure the relative strength of the expressions, we chose to look at AU intensity rather than the presence or absence of AUs. To determine if an image was generating an iconic "positive" or "negative" reaction we used action units for happiness, AU6 and AU12 as indicators of positive emotions and action units AU1, AU2, AU4 and AU5 as indicators of negative emotions. We chose these units for the negative emotion indicator as they were found in the sets of AUs that indicate sadness (AU1, AU4 and AU15); fear (AU1, AU2, AU4, AU5, AU7, AU20 and AU26); and anger (AU4, AU5, AU7 and AU23), although in retrospect we believe that a more sophisticated approach is necessary, especially to track emotions such as disgust and contempt. Our main goal was see if we could detect positive and negative reactions in both the spontaneous and expressed conditions. We additionally wanted to compare overall AU intensity between spontaneous and intentionally expressed responses verify our assumption that intentionally expressed emotions would be more intense in this setting.

4. Feature Extraction

The main features we examined were the intensities of action units (AUs) 1,2,4,5, 6 and 12. To obtain these AU intensities we processed each of the videos using the open source software package OpenFace 2.0[2][1]. Each video was recorded on an iPhone 11, then segmented, extracting the spontaneous and expressed sessions, and converted to MPEG4 format using Adobe Premiere. The segmentation was done manually. An audio click at the beginning of the PowerPoint presentation and the visual action of seeing the start button pressed were used to identify the beginning of each session and synchronize the features generated by OpenFace. We estimate that the synchronization is accurate to within a half second of the actual start time. We believe that this tolerance is acceptable given that in this analysis we present results are based on averages taken over the entire five second window of the image presentation.

As discussed earlier, in this analysis, we were particularly interested in actions units that represented either a positive or negative response. We chose to look at AU6, Cheek Raiser (near eye) and AU12, Lip corner puller as being representative of a positive response (smile) and AU1, inner brow raise; AU2, Outer brow raise; AU4 Brow lowerer; and AU5, Upper lid raiser (eye) as indicators of a negative response as mentioned previously.

We processed our video at a rate of 30 frames per second using OpenFace. A set of features was calculated for each frame. To calculate the response to each stimulus image we took an average over the 600 feature sets we had for each of the participants' five second exposures to each image (30 frames over 5 seconds) in each of the two conditions. We did not analyze the frames where the participant was viewing the white screens between images. There was no weight or kernel applied to the features before averaging, all frames were counted equally.

5. Results

In our analysis we looked at the average intensity of spontaneous versus intentionally expressed emotions and looked a range of individual differences. These included average differences in participant responses across AUs, differences in the AU intensities of the AUs we associated with positive emotions versus the AUs we associated with negative emotions and individual differences in responses by the same participant to the same image. We discuss each in the following subsections.

5.1. Average Spontaneous versus Intentional Responses

We calculated the average intensity of each of the AUs of interest (AU1,AU2,AU4,AU5,AU6 and AU12) across all participants for all images to see the overall difference in expression intensity between the spontaneous and intentionally expressed condition. We found that overall the AUs we associated with a negative responses (AU1, AU2, AU4 and AU5) had 33% more intensity in the expressed condition but that AUs we associated with a positive response (AU6 and AU12) had 117% more intensity in the expressed condition. The results for each AU for each condition can be seen in Figure 3. These average results show the trend we had anticipated, however we had expected to see even more strong differentiation in the negative AUs.

5.2. Variation Across Participant Averages

The overall average profile shown in Figure 3 was not, however, a blueprint for all individual participants. When we look at any individual's average AU intensity profile across all images (both positive and negative) we see a range of variation. Figure 4 shows three examples. The response pattern of Participant 1 (top) exemplifies what we had expected to see from this experiment: spontaneous reactions that were much weaker in AU intensity than the expressed reactions. We note that this follows a very different pattern from the overall average, with higher response in AU6 and AU12 than in the negative emotion AUs. The profile of Participant 6 (middle) most closely mirrors the overall average profile in Figure 3, however the difference where between negative AUs in the two conditions is far less, with the average intensity for AU5 actually being higher in the spontaneous case. Participant 4 (bottom) shows a pattern of response that mirrors the overall pattern in the intensities of the intentionally expressed reactions, however she had nearly zero AU intensity in the positively associated AUs (AU6 and AU12) in the spontaneous condition. We also note that in the intentionally expressed condition her responses showed higher than average intensity (note the y-axis for this participant extends to 3, higher than those for the other two (1.2 and 1.8) or the overall average (1.4)). Assuming that this participant followed our instructions and expressed an emotion she genuinely felt in the expressed condition, we infer that this participant did feel happiness when viewing these images in the spontaneous case but simply did not visibly express it in the facial action units we were tracking.

5.3. Variation Across Individual Reactions

After looking at the differences in individual profiles across all images we wanted to look at differences in participant responses to individual images. In particular we wanted to see if we could use AU intensity in our positive and negative associated AUs to differentiate responses to the positively and negatively valenced images. Our hope had been that many participants would react with an intensity profile similar to the ones shown Figure 5. Here we see Participant 1 reacting with high intensity in our positive associated AUs (AU6 and AU12) in response to the positive valence Image 1, I456 Lake, top, and reacting with higher intensity in our negative associated AUs to the negative valence image I326 Fire (bottom). If this response pattern were typical, which we had anticipated from the prior literature, then we could simply take the ratio of average positive AU activation and compare it to average negative AU activation to determine the nature of the response.

Unfortunately, the responses shown in Figure 5 do not occur either consistently or frequently. As a counterexample, Figure 6 shows Participant 4's the reaction to the positive valence Image 1, I456 Lake (top), compared to the negative valence image I714, Scary Face, (bottom). This activation pattern is actually quite typical for Participant 4 across all images, both positive and negative, as Participant 4 only very rarely showed any activation in either AU6 or AU12 for any image in the spontaneous condition as can be seen in her average profile in Figure 6 (bottom). Using a naive ratio algorithm on Participant 4 would likely show her response as negative to a number of images where she actually expressed that she felt a positive feeling. Such a simplistic algorithm would most often judge Participant 4's response incorrectly, even if were valid, on average, for a general population.

5.4. Incongruent Reactions

To get an overall sense of how well a simple ratio would work, assuming that Participant 4 was an outlier, we calculated the average ratio of positive and negative associated AUs, as discussed in the Section 4, for all participants for all images across both conditions. We refer to our AU indicators as being congruent with the stimulus if the average of the positive AU response (AU6 +

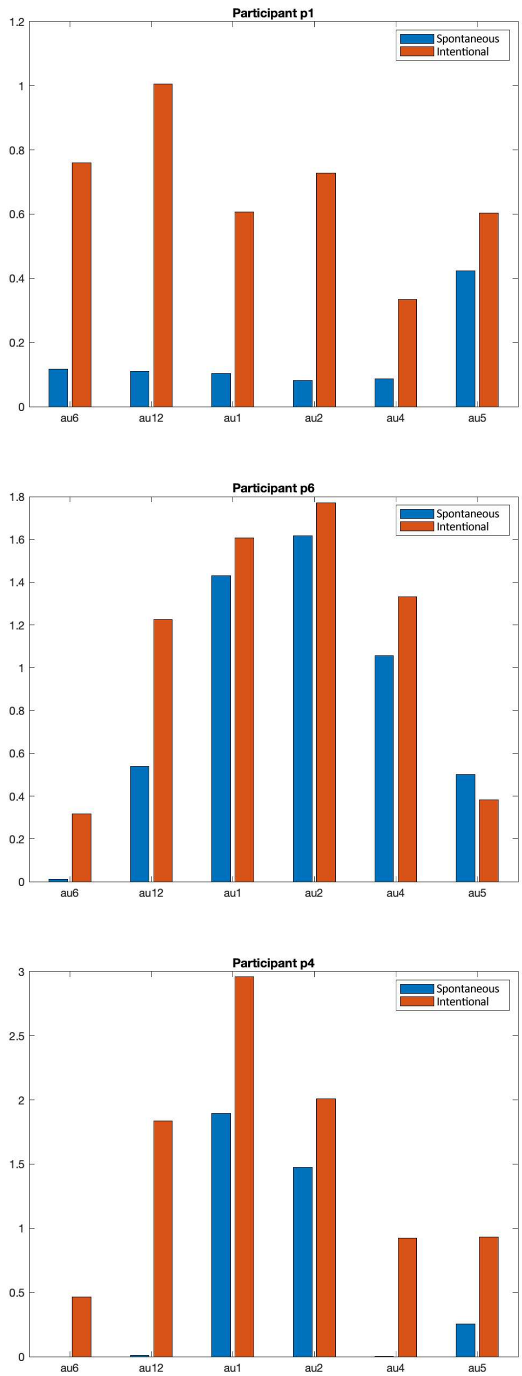

Figure 4. Examples of individual differences in average AU intensity across all images between Participant 1 (top), Participant 6 (middle) Participant 4 (bottom)

AU12) divided by the average of the negative response AUs (AU1+AU2+AU4+AU5) was greater than one for positive images and if it was less than one for the negative images,

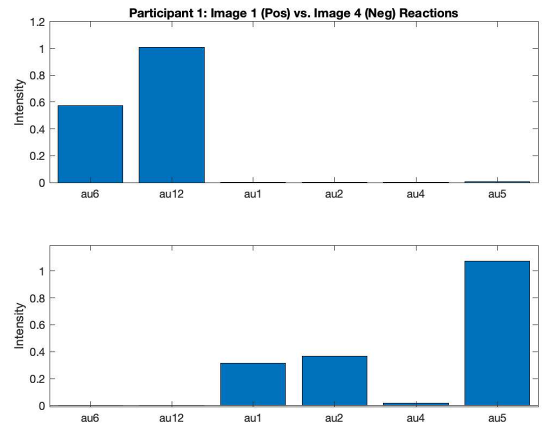

Figure 5. Two individual reactions to a positive (top) and negative (bottom) image from Participant 1. Such AU differentiated reactions, if typical, would allow us to easily detect affective responses from users in real time.

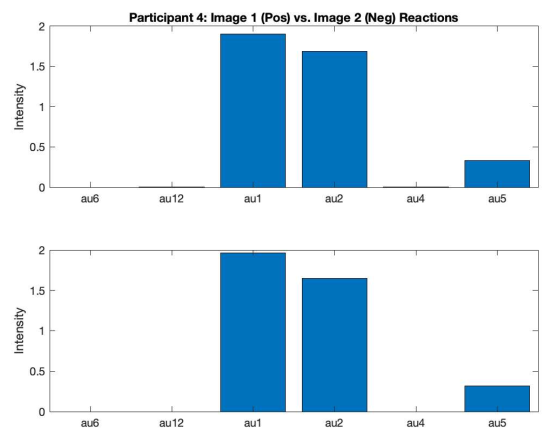

Figure 6. Two individual reactions to a positive (top) and negative (bottom) image from Participant 4. Such AU profiles were typical for all images from Participant 4 who rarely showed natural activation in AU6 or AU12 to any image

where the positivity or negativity of each image is defined by the OASIS valence rating as explained in Section 2.1 and listed in Table 1.

While simplistic, this is the type of model that could be quickly calculated and used in real time with no prior knowledge of the user in interactive situations. The overall results for this method are shown in Figure 7. In this graph, a score of seven indicates that all seven participants responded in a manner "congruent" to the stimulus. An average (but not technically achievable) score of "3.5" would be indicative of random chance. These results show that the ratio method described performs only slightly better than

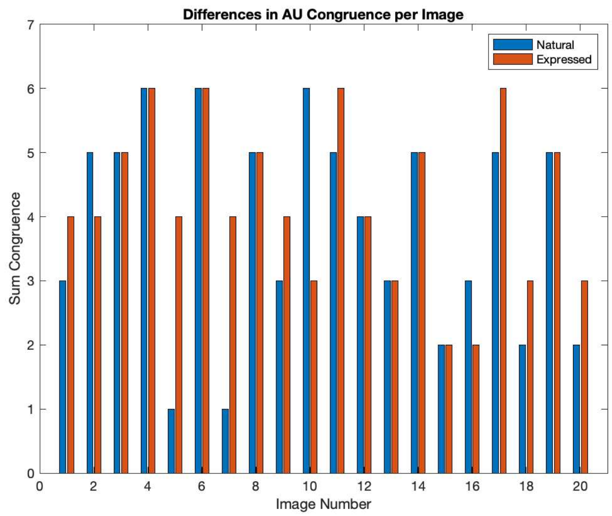

Figure 7. A graph of the sum of congruent responses to each image in for both the spontaneous (left) and intentionally expressive (right) conditions. Here the maximum potential score is seven, indicating all participants expressed a congruent response.

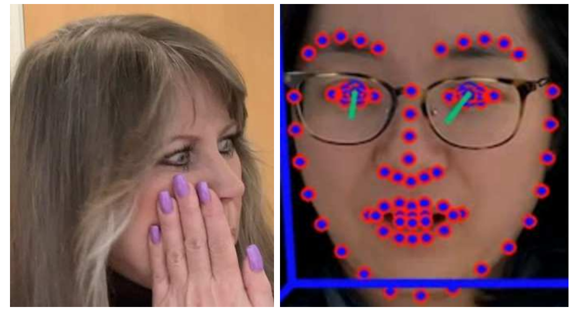

Figure 8. Examples of expressions of emotion that confound AU analysis, on the left the participant covers their face and looks away in horror at the sight of Emaciated Child, on the right participant vocalizes "Wow" in reaction to Fire.

random, with a bias toward detecting negative responses. Specifically, we found that for the spontaneous condition 24 of 70 positive images received a congruent response and 53 of 70 negative responses received a congruent response. For the intentionally expressed condition 33 of 70 positive images received a congruent response whereas 51 of 70 negative images received a negative response. We found that the images with the highest congruent response were all negative: Fire(4), War(6), Dirt(11), Fire(14) and Dead Bodies(17), the images with the lowest congruence were all positive: Lake(5), Beach(7), Rainbow(15), Beach(16), Lake(18) and Fireworks(20).

5.5. Manual Review

We performed a manual review of the video to see better understand why our method performed poorly. We focused our efforts on our detected instances of "incongruence" between stimuli and facial expression where greater positive or negative average AU activation was opposite of the valence of the image being shown. From this review, we found that there were several factors that explained the difference in activation including: variations of the style and intensity of emotional reactions, the occurrence of facial expressions like contempt that could have better been differentiated by action units that we did not consider, by hand motions which occluded that face and by co-occurring events such as vocalizations and head turning (both away in disgust/horror and back and forth in a manner indicating "no" as shown in Figure 8. Although we knew that the guidance for FACS coding is that frames with occlusions such as a hand touching or covering the face cannot be properly evaluated and that these should be removed, however we did not perform this pre-processing in this study as we wished to see how well a simple implementation would perform. Our results indicate that such effects do indeed need to be taken into account.

We examined several examples where the ratio model failed to detect a congruent response. One example is shown in Figure 9, where the participant is displaying an emotion that would be most likely contempt. We infer this from both the side lip pull and the subsequent slow side to side head shaking action that the participant displayed. We could possibly have detected this as a negative response if we had been considering AU14 (one sided dimpling) or if we had tracked that the intensity in AU12, lip pull was one sided in contrast to the two sided action in AU12 associated with smiling. We would also have to be more careful in how we looked at the sequence as both the one side lip pull and one side dimpling is very brief and might be overwhelmed in an straight average across the entire reaction. Specifically, during this five second response period, the participant first seems slightly confused, then smiles a little, then begins the one sided lip pull and then initiates lateral head shaking with a small frown. This last action is not well captured due to the head movement. This response is different from the same participants spontaneous response, which is detected as negative as shown in Figure 10. We consider that some social display rule may be responsible if, in the intentional condition, the participant was expressing how she thought she *should* respond to this image.

6. Discussion

In this analysis, we presented some examples of challenges in recognizing emotion "in the wild" using currently available facial expression recognition technology. In our pilot study, we show that on average, spontaneous facial expressions have less intensity than intentional expressions of

Figure 9. An example of a segment where a participant's expressed reaction to a negative image was considered positive due to higher AU activation in AU6 and AU12 which are normally associated with the emotion "happiness"

Figure 10. A comparison of the natural (left) vs the expressed (right) reaction of a participant to the trash image from Figure 9

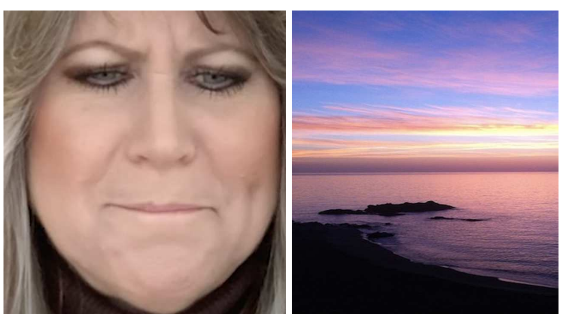

Figure 11. An examples of where a participant's expressed reaction to a positive image, I60 Beach, was considered negative due to higher activation in AU1, AU2, AU4 and AU5. Although the overall reaction to the image does actually seem positive, the participant does not smile

emotion to the same stimuli. We found that the difference in intensity between these two conditions for our positive associated AUs: AU6 and AU12 than for our negative associated AUs: AU1, AU2, AU4 and AU5. We believe that this might be partially due this to social display rules: in an office environment it is less appropriate to show negative emotions than positive ones. We also show that there is wide variation in individual response profiles, where some

Figure 12. A side by side comparison of the natural (left) and expressed (right) reactions of the same participant to I60 Beach. In this case the natural reaction was judged to be negative but the expressed reaction was judged to be positive.

participants do not spontaneously smile at images that inspire happiness. We also noted that our ratio algorithm misclassified contempt as a positive response and failed on occlusion, motion and vocalizations. We found that a naïve ratio classifier that simply compared the average intensity AUs normally associated with positive emotions versus a subset of AUs normally associated with negative emotions was highly unreliable. For both spontaneous and expressed positive emotions, accuracies were 24% and 27% respectively. For negative emotions we achieved accuracy rates of 65% and 72% respectively, but this could simply a bias towards negative recognition.

In this paper we present the result of a naïve approach to detecting positive and negative reactions using facial action units associated with smiling and brow actions associated with multiple negative emotions. We found that our approach inaccurate in many cases due to: very low intensity spontaneous expression, confounds of happiness and contempt, facial occlusions and vocalizations. This study and evaluation are limited, but we had hoped that the task was so straightforward that discrimination would be trivial. We found this not to be the case. Both individual differences and the specific recognition of different types of categorical emotions (including "non-basic" emotions such as contempt) seem necessary for discriminating positive and negative responses. We are aware that this study has a very small number of participants, and we hope to greatly increase this number in future work. We are also aware that taking an average over the entire five minute window is a very coarse method and that it is very likely that our discriminator would be improved by looking at these features as they evolve over time[5]. We also believe that detecting and differently processing frames with occlusion and vocalizations will provide valuable signals to improve results. We hope that this initial study will help shed light on some of the challenges we found in emotion recognition even in a simple task.

References

[1] Tadas Baltrušaitis, Marwa Mahmoud, and Peter Robinson. Cross-dataset learning and person-specific normalisation for automatic action unit detection. In *2015 15th IEEE International Conference on Automatic Face Gesture Recognition*, 2015.

[2] Tadas Baltrušaitis, Amir Zadeh, Yao Chong Lim, and Louis-Philippe Morency. Openface 2.0: Facial behavior analysis toolkit. In *2018 18th IEEE International Conference on Automatic Face Gesture Recognition*, 2018.

[3] Lisa Feldman Barrett, Ralph Adochs, and Stacy Marsella. Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological Science in the Public Interest*, 20(1):1–68, 2019.

[4] Jeffrey Cohn, Lawrence Reed, Tsuyoshi Moriyama, Jing Xiao, Karen Schmidt, and Zara Ambadar. Multimodal coordination of facial action, head rotation, and eye motion during spontaneous smiles. pages 129–138, 01 2004.

[5] Jeffrey Cohn and Karen Schmidt. The timing of facial motion in posed and spontaneous smiles. *International Journal of Wavelets, Multiresolution and Information Processing*, 02, 11 2011.

[6] Jeffrey F Cohn, Zara Ambadar, and Paul Ekman. Observer-based measurement of facial expression with the facial action coding system. *The handbook of emotion elicitation and assessment*, 1(3):203–221, 2007.

[7] Kate Crawford, Roel Dobbe, Theodora Dryer, Genevieve Fried, Ben Green, Elizabeth Kaziunas, Amba Kak, Varoon Mathur, Erin McElroy, Andrea Nill Sánchez, Deborah Raji, Joy Lisi Rankin, Rashida Richardson, Jason Schultz, Sarah Myers West, , and Meredith Whittaker. Ai now 2019 report. Technical report, AI Now Institute, 2019.

[8] Abhinav Dhall, Roland Goecke, Simon Lucey, and Tom Gedeon. Acted facial expressions in the wild database. 10 2011.

[9] Hamdi Dibeklioglu, Albert Salah, and T. Gevers. Recognition of genuine smiles. *IEEE Transactions on Multimedia*, 17:279–294, 03 2015.

[10] Hamdi Dibeklioğlu, Albert Ali Salah, and Theo Gevers. Are you really smiling at me? spontaneous versus posed enjoyment smiles. In Andrew Fitzgibbon, Svetlana Lazebnik, Pietro Perona, Yoichi Sato, and Cordelia Schmid, editors, *Computer Vision – ECCV 2012*, pages 525–538, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[11] Paul Ekman. An argument for basic emotions. *Cognition and Emotion*, 6(3-4):169–200, 1992.

[12] Paul Ekman. Emotional and conversational nonverbal signals. In J.M.; Miranda. L.A.P. Larrazabal, editor, *Language, Knowledge, and Representation*. Springer, Dordrecht, 2004.

[13] Paul Ekman and Wallace V. Friesen. *Facial Action Coding System: A technique for the measurement of facial movement*. Consulting Psychologists Press, Palo Alto, CA.

[14] Paul Ekman, Wallace V. Friesen, and Joseph C. Hager. *Facial Action Coding System: Manual and Investigator's Guide*. Research Nexus, Salt Lake City, UT.

[15] Irfan A. Essa and Alex P. Pentland. Coding, analysis, interpretation, and recognition of facial expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):757–763, 1997.

[16] Wallace V. Friesen and Paul Ekman. Emfacs-7: Emotional facial action coding system. 1983.

[17] Benedek Kurdi, Shayn Lozano, and Mahzarin R Banaji. Introducing the open affective standardized image set (OASIS). *Psychology Review*, 97:315–331, 1990.

[18] S. Li and W. Deng. Deep facial expression recognition: A survey. *IEEE Transactions on Affective Computing*, 2020.

[19] A. Mehrabian. *Basic Dimensions for a General Psychological Theory: Implications for Personality, Social, Environmental, and Developmental Studies*. Social Environmental and Developmental Studies. Oelgeschlager, Gunn & Hain, 1980.

[20] Andrew Ortony and Terrance J. Turner. What's basic about emotions? *Psychology Review*, 97:315–331, 1990.

[21] James A. Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6):1161–1178, 1980.

[22] Michel F. Valstar, Maja Pantic, Zara Ambadar, and Jeffrey F. Cohn. Spontaneous vs. posed facial behavior: Automatic analysis of brow actions. In *Proceedings of the 8th International Conference on Multimodal Interfaces*, ICMI Äô06, page 162Äì170, New York, NY, USA, 2006. Association for Computing Machinery.

[23] Shangfei Wang, Zhilei Liu, Siliang Lv, Yanpeng Lv, Guobing Wu, Peng Peng, Fei Chen, and Xufa Wang. A natural visible and infrared facial expression database for expression recognition and emotion inference. *IEEE Transactions on Multimedia*, 12(7):682–691, Nov 2010.

[24] Lijun Yin, Xiaochen Chen, Yi Sun, Tony Worm, and Michael Reale. A high-resolution 3d dynamic facial expression database. In *2008 8th IEEE International Conference on Automatic Face Gesture Recognition*, pages 1–6, 2008.

[25] Zheng Zhang, Jeffrey M. Girard, Yue Wu, Xing Zhang, Peng Liu, Umur Ciftci, Shaun Canavan, Michael Reale, Andrew Horowitz, Huiyuan Yang, Jeffrey F. Cohn, Qiang Ji, and Lijun Yin. Multimodal spontaneous emotion corpus for human behavior analysis. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3438–3446, June 2016.