Detecting Emotion from People and Scenery

Logan Preston (lpreston2@wisc.edu), Neal Desai (nbdesai2@wisc.edu)
and Nikhil Nigam (nnigam@wisc.edu)

1 Introduction

The motivation behind this proposed project is to investigate approaches for emotion detection and recognition. This builds on existing work of facial recognition and tracking to predict what the user is likely feeling based based on image-specific context. Existing work often looks to face expressions, body language, and scene context. These varied factors combined with individual variation in how people express emotion makes this a challenging but useful problem to solve.

Specifically, there is social value in having a good solution to the emotion recognition problem. The use of AI-infused systems is growing in daily life and generally is expected to continue to do so. Human AI interaction guidelines state that AI infused systems should have "socially appropriate behaviors" [2] or "match relevant social norms" [1], showing that being aware of social expectations is helpful. Detecting user emotions can inform the AI system on appropriate social expectations and also potentially detect emotional anomalies. This could be used to make artificial intelligence applications more socially aware, friendly, or provide better support to humans. For example, identifying if a user is frustrated may prompt the system to check if the user has a question, provide supportive reminders, or reduce the amount of notifications it sends to the user. In general, we expect identifying emotions accurately will enhance interactions between AI infused systems and their users.

2 State of the Art

This problem is especially thought-provoking because we already know computers can identify faces, track objects paths, and other related tasks. However, the work on emotion detection is not as well-established. Classical emotion recognition techniques take a more localized approach by simply focusing on the prominent facial features. These techniques have been now bettered, if not matched, by deep learning algorithms such as CNNs or LSTMs [3].

Thus, prior work reviews emotions by looking at the faces alone rather than considering body language and scene context. Some work adds in context from the surrounding the scene [5] but efforts in this area seem minimal and don't focus on quantifying the potential improvement of considering scene context vs just faces. We are interested in seeing how important the overall scene context is for the accuracy of the model or if the face / body language provides the vast majority of the data needed for accurate classification. Apart from the further exploration by those who initially included the scene for analysis [4], another method adopted by researchers involved masking the face from the image and treating facial features and the scene separately and adopting ensemble learning techniques for their prediction [6]. These findings act as a starting point for a more nuanced technique which can be developed and researched further.

3 Proposed Solution

Our solution will start by identifying people from the image, which will require segmenting the image as a prerequisite. After the humans in each image are identified, we can look to them for potential clues into the emotions using their faces and/or body language. We will further investigate the impact of incorporating the full body and scene context for our emotion recognition as opposed to just isolated faces.

We will evaluate the performance of solutions using only expressions and body language and compare those results to the performance of solutions that also include information on the scene. This will identify the impact of the scene information and determine how valuable of information it is to analyze compared to the individuals in a scene.

To train our model, we plan to use the Emotic dataset. The dataset contains images of several people in real environments. Each person is appropriately annotated with their emotions. The dataset offers both a discrete analysis using 26 distinct emotion categories as well as

a continuous dataset that seeks to quantify each emotion into three dimensions, namely valence, arousal, and dominance [4]. We will incorporate both the discrete and continuous cases in our analysis. We plan on initially using Convolutional Neural Networks to accomplish this classification and will explore other deep learning model architectures that may improve upon the baseline CNN approach.

4 Time Table

Objective	**Due Date**
Identify and Review potential emotion data sets	2/28
Create or modify an existing solution that identifies emotion and can separate people from scenery	3/21
Train and test solution with small portion of the data set to identify impact of scene vs body language	3/28
Gather current results for midterm report and begin preparing report	3/28
Finalize midterm report	4/5
Extend testing with complete data set to identify impact of scene vs body language	4/15
Begin creating final report and presentation	4/15
Finalize final report and presentation	4/25

Table 1: Proposed Time Table

References

[1] S. Amershi, D. Weld, M. Vorvoreanu, A. Fourney, B. Nushi, P. Collisson, J. Suh, S. Iqbal, P. Bennett, K. Inkpen, J. Teevan, R. Kikin-Gil, and E. Horvitz. Guidelines for human-ai interaction. In *CHI 2019*. ACM, May 2019. CHI 2019 Honorable Mention Award.

[2] E. Horvitz. Principles of mixed-initiative user interfaces. In *Proceedings of CHI '99, ACM SIGCHI Conference on Human Factors in Computing Systems, Pittsburgh, PA, ACM Press.*, pages 159–166, May 1999.

[3] B. C. Ko. A brief review of facial emotion recognition based on visual information. volume 18, 2018.

[4] R. Kosti, J. Alvarez, A. Recasens, and A. Lapedriza. Context based emotion recognition using emotic dataset. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.

[5] R. Kosti, J. M. Alvarez, A. Recasens, and A. Lapedriza. Emotion recognition in context. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[6] J. Lee, S. Kim, S. Kim, J. Park, and K. Sohn. Context-aware emotion recognition networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019.