SmileAtMe – Rating and Recommending Funny Images via Smile Detection

Jonathan Brem

University of Regensburg

93053 Regensburg, DE

jonathan.brem@stud.uni-regensburg.de

Jürgen Hahn

University of Regensburg

93053 Regensburg, DE

juergen.hahn@stud.uni-regensburg.de

Abstract

In this paper, we propose an Android-based assistance system called SmileAtMe which uses smile detection as a means of direct interaction in order to rate meme images. A smartphone's front camera captures the user's face while they are interacting with the application on the device. A system evaluation uncovered that users fully understand how the application uses smile detection to rate images. The users attested that the automatic reaction classification works well. However, several of them were uncomfortable with the idea of being observed by their smartphone.

Author Keywords

Facial Expression-based Interaction; Perceptual User Interface; Smile Detection; Image Rating; Image Recommendation; Memes;

ACM Classification Keywords

H.5.2 User Interfaces: Input devices and strategies, Evaluation/methodology; H.1.2: User/Machine Systems: Human Factors; I.2.10 Vision and Scene Understanding: Perceptual Reasoning; I.4.9 Image Processing and Computer Vision: Applications;

Figure 1: An example for an Internet meme image with the *Doge* meme. Image CC0 Euterpia on Wikipedia.

Introduction

There are several websites and mobile applications with the goal of entertaining users by showing short, funny posts consisting solely of a title and one image or video. Users can react to these posts in different ways. For example, they can be indifferent towards them, grin, smile or laugh.

A more accurate term for these pictures is "meme images". In our context, the term meme is used differently than in cultural studies or memetics: In an analysis of Internet memes, Knobel and Lankshear attest that there are no deep similarities between these concepts [6]. Two example websites containing meme images are 9GAG [1] and HUGELOL [5], the latter of which has a description that suggests how people should use meme image websites: "Scroll. See. Laugh. Repeat." [5] An example is the image in Figure 1.

When people use their smartphone or tablet and look at the device's screen, a built-in front camera can capture their faces, which makes the current facial expression accessible for interaction within mobile applications.

The conventional interaction paradigm for rating posts and funny images requires explicit user input via on-screen controls. In this paper, we propose an Android-based assistance system, SmileAtMe, that shows a user funny images and uses smile detection in order to classify their reaction to the images as positive (smiles) or neutral / negative (other facial expressions). Based on this detection data, the system recommends and displays pictures similar to the ones to which the user reacted positively, i.e. that made them smile or laugh.

We investigate the possibilities of this approach with the following research question:

RQ: Do users understand and accept the use of smile detection as a means of rating funny images?

Related Work

Cowie et al. [2] describe the use of implicit messages such as emotions in communication between humans as well as in human-computer-interaction. This implicit channel provides information that helps interpret the explicit messages. Its use in HCI facilitates communication between humans and machines because it is closer to human-to-human interaction. In this way the other party is easier to understand or, in the case of our work, it enables better assistance to a user. The authors list specific ways in which systems can make use of emotions, such as understanding the user's needs in a tutoring applications or artificial intelligence systems behaving more realistically and life-like. Our work is different from these scenarios insofar as it uses facial expressions and emotions as explicit input instead of gathering additional information about messages conveyed in another channel.

Déniz, Castrillón, Lorenzo-Navarro, Antón-Canalís and Bueno [3] developed a system that uses smile detection to enhance user experience for an Instant Messaging client. When their application detects a smile, it translates it to an emoticon. This is similar to what we propose since smile detection is used as an unobtrusive input method. However, emoticons are understood as a way of conveying tone or implicit information about other messages by the authors analogous to the previously mentioned work, whereas facial expressions are used as a direct means of communication in our system.

System Description

SmileAtMe was developed as a prototypical application targeting Android systems. In order to classify the reactions, we use Google's Mobile Vision API for Android that offers the required functionality to detect

Figure 2: A screenshot of the application. Image CC-BY-NC Randall Munroe on xkcd (https://xkcd.com/643/).

faces, facial features and smiles [4]. For SmileAtMe, the most important ability of the API is to calculate the probability with which the facial expression on every face it recognized in a picture is a smile.

These values are calculated by frame, but there is a temporal aspect that plays a role in classifying reactions. Therefore, the values need to be collected and analyzed to determine user's overall response to the meme image.

In order to investigate if the API can be used to classify the users' reactions, a preliminary study was conducted. 18 participants were asked to look at meme images on a smartphone's screen and rate how much looking at the image made them laugh. Face detection data and smiling probability values were gathered during the study to determine if they could be used to predict the users' ratings. The results are that classification is possible as long as the application detects the users' faces. This is not always the case (e.g. if the user looks away from the device) so it is required to notify the user when the face cannot be found in the current camera frames.

The earlier mentioned description of HUGELOL shows that inducing laughter is a goal of the pictures, so smile detection may be appropriate to see if the images produce that reaction.

When the user has finished looking at an image, they can request to see the next image by clicking a button. SmileAtMe then displays the result of the classification before showing the next meme. Additionally, there are warnings on the screen when the front camera does not capture a face.

Upon the system's first start, a tutorial is displayed which explains the application's concept. It addresses potential security concerns by clarifying that no video is created and no face data is permanently stored. After the tutorial or in subsequent starts, the application displays a meme image on most of the screen (see Figure 2).

Tags such as "cat picture" are assigned to the images and used to categorize them. That information is used to calculate the percentage of images of a category a user has laughed at and to recommend more images from categories the user likes best.

System Evaluation

We conducted a user study with the prototype to find answers to our research question, whether users understand the input method and would actually use the smile detection application.

16 students, aged 20 to 26, participated in the study. They were asked to open the application and explore it for the amount of time they saw fit. The only instructions users got are from the tutorial and error messages. External factors such as adequate lighting are not artificially set up to support the face detection. This free exploration allowed participants to view as many meme pictures as they wanted and needed in order to form an opinion of the ratings method.

Afterwards, they had to fill out a questionnaire that consisted of five questions. The first two questions were how well they had understood the application in general and the input method via smile detection specifically. Then the users were asked how well the input method performed, if they prefer this method or a more conventional input method, and how uncomfortable they were with being permanently observed by their smartphone. All questions except for their preferred way of input were answered on a five-point Likert scale, 5 being the most favorable answer for the system.
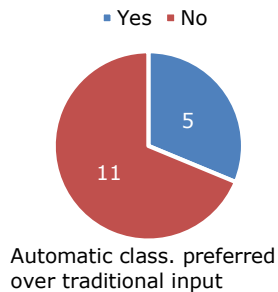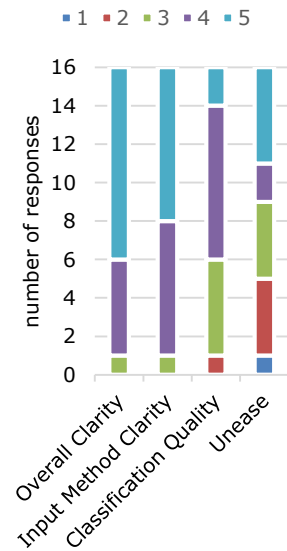
Figure 3: Evaluation results, based on 16 participants' responses

The median response to the question about the system's technical performance was a 4 on a scale of 1 to 5, i.e. the classification worked well.

Figure 3 shows the answers to the questions in the questionnaire that concern the RQ except for the question about the preferred way of input. 11 participants (69%) prefer traditional input methods over the automatic classification.

The idea of using smile detection was very clear to users as shown by a median value of 4.5 and the median response of 5 for the overall clarity gives proof that the general concept of the application, including the recommender component, is not confusing to users and therefore works.

With regards to the acceptance of the input method, 9 of the 16 users were at least somewhat uncomfortable with the idea of the application monitoring their facial expression, 5 being very uneasy with it.

Four of the participants stated that it would help if the detection also notices when they finished looking at a meme image and automatically displays the next one. Seven users expressed the concern that the system might make many errors or function poorly in some conditions whereas conventional interaction with buttons is not prone to error.

The answer to the RQ is that users fully understand how smile detection is used to rate funny images. However, some reservations exist relating to privacy issues. The acceptance could be improved if the classification results were even more reliable.

Conclusion and Future Work

Based on the evaluation results, the rating method works and users understand how to use it. However, the biggest problem with the system is that 9 of 16 participants felt at least some unease when operating the system due to the constant observation by the device's front camera. This makes investigating this unease the most urgent future work, i.e. if it disqualifies the use of smile detection completely or if there are ways to make users feel more comfortable while using the system.

If that is possible, other future work includes implementing and testing a more detailed classification of users' reaction (perhaps with a custom facial expression detection).

References

1. 9GAG. 2016. 9GAG – Go Fun Yourself. Retrieved March 26, 2016 from http://www.9gag.com

2. Roddy Cowie, Ellen Douglas-Cowie, Nicolas Tsapatsoulis, George Votsis, Stefanos Kollias, Winfried Fellenz and John G. Taylor. 2001. Emotion recognition in human-computer interaction. *IEEE Signal Processing Magazine* 18, 1: 32-80.

3. Oscar Déniz, Modesto Castrillón Santana, Javier Lorenzo-Navarro, Luis Antón-Canalís and Gloria Bueno. 2008. Smile Detection for User Interfaces. In *Advances in Visual Computing*, George Bebis et al. (eds.). Springer Berlin Heidelberg, Berlin / Heidelberg, 602-611.

4. Google. 2015. Mobile Vision | Google. Retrieved March 26, 2016 from https://developers.google.com/vision/

5. HUGELOL. 2016. HUGELOL – About. Retrieved March 26, 2016 from http://hugelol.com/more

6. Michele Knobel and Colin Lankshear. 2007. Online memes, affinities, and cultural production. In *A new literacies sampler*, Michele Knobel and Colin Lankshear (eds.). Peter Lang Publishing, New York, 199-227.