Synthesizing Expressions using Facial Feature Point Tracking: How Emotion is Conveyed

Tadas Baltrušaitis, Laurel D. Riek, and Peter Robinson
Computer Laboratory
University of Cambridge
tb346@cam.ac.uk, Laurel.Riek@cl.cam.ac.uk, Peter.Robinson@cl.cam.ac.uk

ABSTRACT

Many approaches to the analysis and synthesis of facial expressions rely on automatically tracking landmark points on human faces. However, this approach is usually chosen because of ease of tracking rather than its ability to convey affect. We have conducted an experiment that evaluated the perceptual importance of 22 such automatically tracked feature points in a mental state recognition task. The experiment compared mental state recognition rates of participants who viewed videos of human actors and synthetic characters (physical android robot, virtual avatar, and virtual stick figure drawings) enacting various facial expressions. All expressions made by the synthetic characters were automatically generated using the 22 tracked facial feature points on the videos of the human actors. Our results show no difference in accuracy across the three synthetic representations, however, all three were less accurate than the original human actor videos that generated them. Overall, facial expressions showing surprise were more easily identifiable than other mental states, suggesting that a geometric approach to synthesis may be better suited toward some mental states than others.

Categories and Subject Descriptors

H.1.2 [**User/Machine Systems**]: Human factors

General Terms

Human Factors

Keywords

Affective Computing, Facial expression analysis, Facial expression synthesis, Emotion

1. INTRODUCTION

Humans express their inner emotional or affective states through nonverbal expressive behaviour [3]. Facial expressions are of special importance when communicating affect and thus have been studied extensively. They have been investigated scientifically for nearly 140 years and are generally well understood [14, 15]. Facial expressions not only signal the basic emotions but also affective or mental states [6]. They also provide an important channel for human-human interaction and can reveal intent, signal affection, and help people regulate turn-taking during conversation.

People can discriminate facial expressions of emotion even when the input signal is very limited. For example, Bassili [8] performed a point-light experiment which showed that even when the only available information available is the dynamics of a facial expression of emotion without any appearance cues people still perform above chance in an emotion recognition task.

Perhaps this result inspired some of the first attempts in automated facial expression analysis, which used a geometry driven approach to recognition [21, 16]. While recent trends in the field have led researchers toward using more appearance driven [7] or combined geometry/appearance driven approaches [4].

One common approach of geometry driven facial expression analysis is the use of located facial features to detect Action Units (AU) from the Facial Action Coding System (FACS) which are the building blocks of facial expressions [15]. AUs can then be used to infer emotional displays and gestures, which can then be used to infer mental states.

One advantage of a geometric approach is that it lends itself extremely well to expression synthesis [26]. Probably the most straightforward approach to facial expression synthesis is to use automatically tracked data and translate it to AUs or MPEG-4 facial animation parameters (FAPS) [20]. Raouzaiou et al. [22] identified possible mappings from FAPS to AUs. Bee et al. [9] amongst others used AUs to synthesize animation on a virtual character. MPEG-4 driven animation has been extensively used by several researchers [1, 2, 22].

However, moving from expression recognition to interpretation and ultimately to synthesis requires discrimination between configurations of the face that are psychologically significant against those which have only a morphological value [1, 11]. In particular, it is necessary to understand how well people can extract emotional information conveyed via geometry driven synthesis. In addition to helping improve the synthesis, this understanding can also help to improve automatic facial expression recognition.

In this work, we wanted to determine how much affective information automatically-tracked facial feature points convey when synthesized via different representations. In particular, we wondered if more photo-realistic representations of synthetic characters lead people to be more accurate in their ability to identify emotions.

Other similar experiments were carried out by Ahlberg et al. [2] and Costantini et al. [12], who were interested in evaluating the emotional expressivity of synthetic faces in comparison to real actor data that generated the synthetic expressions. Ahlberg et al. used synthetic faces that were animated by 22 tracked facial feature points through head tracking equipment and IR-reflecting markers creating MPEG-4 FAPs and using them on two different facial animation engines. They then used the two synthetic forms and the videos of actors in an emotion judgement study. Their results showed significant recognition differences between the synthetic and real facial expressions, but no differences between synthetic representations. Costantini et al. compared emotion recognition of real and synthetic facial expressions. The synthetic expressions were generated in two ways: a scripted approach (manual animation), and a data-driven one (generated from actor performances). In absolute terms the scripted approach was more recognisable than data-driven one, although the data-driven one was more similar to that of the videos of real actors. One limitation of the mentioned studies is that they only used basic emotions, which are not the ones that are most common in day-to-day interaction [23].

Another relevant study was perfomed by Courgeon et al. [13] who evaluated the importance of wrinkles in the perception of eight emotions and found the perceived expressivity and preference of avatars increases with the realism of wrinkles, but emotion recognition rates do not.

Our experimental design was similar to that of Afzal et al. [1] who compared emotion recognition accuracy across several geometry driven synthetic displays that varied in their photo-realism (point-light, stick figure, and virtual avatar). Afzal et al. found that people are better at recognizing emotions in stick-figure representations, followed by point-light displays, followed by avatars.

We postulated that their results may have been due to the fact that their avatar representation was of low quality, and thus in our experiment used a far more sophisticated graphical avatar, as well as an extremely realistic android robot. Our results show no difference in participants' emotion recognition accuracy across the three synthetic representations, however, all three were less identifiable than the original human actor videos that generated them. Overall, facial expressions showing surprise were more easily identifiable than other mental states, suggesting that a geometric approach to synthesis may be better suited toward some mental states than others.

2. METHODOLOGY

The purpose of our experiment was to determine how much affective information automatically tracked facial feature points convey when synthesized on different representations. The mental states we used (boredom, confusion, interest, surprise, happiness) were the same ones chosen by Afzal et al. [1] in order to compare our results more readily. The selected mental sates were not balanced in terms of Arousal, Pleasure, and Dominance dimensions [24]: three with positive Pleasure (interested, happy, surprised), four with positive Arousal (interested, happy, surprised, confused), and two with positive Dominance (interested, happy). Videos of actors enacting the selected mental states were used to generate videos on three synthetic representations: a physical android robot head, virtual avatar, and virtual stick figure animations. In order to ensure adequate coverage of the various mental states, we chose three actors to represent each one.

Our design was a $5 \times 4 \times 3$, within subjects, video-based experiment. The variables we manipulated were the five mental states, four representations, and three actors. Thus, we had 60 video stimuli in total. (See Fig. 1 for examples).

We expected that expressions made by different synthetic expressions would lead to different recognition rates, as they allow for better expressivity. In addition, Garau et al. [17] hypothesise that, at least with relation to eye gaze, consistency between visual appearance of the avatar and the type of behaviour is exhibits seems to be necessary: low realism appearance demands low realism behaviour and high realism appearance requires high realism behaviour. Because the animation were created using acted data providing high realism of behaviour the more realistic forms (avatar and robot) were expected to perform better in the emotion recognition task.

2.1 Stimuli Creation

2.1.1 Stimuli Source

The videos of the five mental states used in our experiment were taken from the Mind Reading DVD [5]. The DVD consists of videos of actors expressing various mental states. Each video was judged by a panel of ten judges, and was accepted into the DVD only if eight of the ten judges agreed about the mental state label. The DVD consists of 24 categories of mental states, with subcategories showing different shades of each mental state. These videos provided us with the ground-truth labels for our stimuli.

We chose to include complex mental states (boredom, interest, and confusion) alongside basic ones (surprise and happiness). We selected these for comparison with earlier results of Afzal et al. [1], and those are the ones they used in their experiment due to their relevance in learning contexts. Furthermore, while many researchers in synthesis focus on representing the six basic mental states (anger, joy, surprise, disgust, sadness, and fear), they comprise only a small subset of the mental states that people can experience, and are arguably not the most frequently occurring ones in day-to-day interaction [23]. Apart from *bored*, all mental states included in our experiment are in the top ten of the most commonly occurring expressions in both symmetric and asymmetric interaction [23].

Because our android robot looks like an aged, caucasian male, we needed to find actors on the DVD that were similar in appearance, as it can influence perceived mental states [19]. We also needed to provide sufficient coverage of the five selected mental states (three each). Consequently, we selected three caucasian, older male actors for our stimuli.

2.1.2 Extracting Facial Features

For feature point tracking we used the NevenVision (now Google) facial feature tracker. The tracker uses a generic face template to bootstrap the tracking process, initially locating the position of 22 facial landmarks (see Fig 2). It uses Gabor wavelet image transformations and neural networks for the tracking of subsequent images in the video. It is fully automatic and requires no manual labelling. The tracker deals with a wide range of physiognomies and tracks users that wear glasses and/or have facial hair. It is robust to a certain amount of out-of-plane head motion, and is good at detecting head pose. When evaluated on Boston University dataset [10] the NevenVision tracker the absolute errors of orientation estimation are as follows: roll $\mu = 3.09°, \sigma = 2.58°$, pitch $\mu = 5.73°, \sigma = 7.94°$, and yaw $\mu = 5.19°, \sigma = 4.74°$. All tracked points of visual sequences used in the experiment were visually inspected to validate that the tracking was not giving wildly incorrect results, a formal validation would be desirable for future experiments. In addition, it was used well-validated by other similar uses, such as the MindReader software developed by el Kaliouby and Robinson [16].

The problem with most facial expression recognition from video approaches is that they rely on a neutral initial frame, or a known neutral configuration of the face. Most of the video stimuli used in our experiment do not start in a neutral pose and we do not have a labelled neutral pose of the actor. This posed a significant challenge when detecting and synthesising action units.

Thus, in order to make our system work with non-neutral ini-

Figure 1: Still frames from eight of our 60 video stimuli. From left to right: the human actor, the physical android robot, the virtual avatar, and the stick figure animation. The top four videos show *surprised*, and the bottom four show *happy*.

tial frames, we split the Mind Reading DVD videos based on actors. Then we ran the FaceTracker on every video of the same person, which allowed us to determine their average facial appearance. Some examples of facial appearance features are: distance between inner eye corner to inner brow, mouth width, mouth height, distance from the point between the eyes to the centre of the mouth, etc.

In addition to providing us with the information about the actor's facial configuration, this approach also provided us with information about the way they can change their expression. We accomplished this by recording the maximum amount their mouth can open, the highest their eyebrows can be raised, etc. This made it easier to use the expressions to drive animations on our synthetic representations.

2.1.3 Choosing Synthetic Representations

While we wanted to design our experiment to be comparable to Afzal et al. [1], we also wanted to understand how photo-realistic synthetic characters might affect people's ability to extract affective information. Thus, while three of our stimuli groups were the same (human actor, stick figure drawing, and avatar), we used a photo-realistic android robot instead of a point-light representation.

Stick figure drawings are the least photo-realistic of all our synthetic representations, but remain the most true to the original geometric facial-features upon which they are based.

Slightly more photo-realistic are the virtual avatars used in our experiment, that are driven via Source SDK[1]. It supports animation of characters based on FACS, allowing for rather realistic motion.

Our most photo-realistic representation was an android robot head from Hanson Robotics[2] made to the likeness of Charles Babbage. The robot has 27 degrees of freedom in the face and neck, is very lifelike and contains wrinkles on the skin. Thus, the robot is able to enact very human-like expressions. For this experiment, only 22 of the motors were used; corresponding to the tracked facial landmarks.

2.1.4 Generating Synthetic Representations

The stick figure animations used in the experiment were synthesised using the same rules described in Afzal et al. [1]. Every

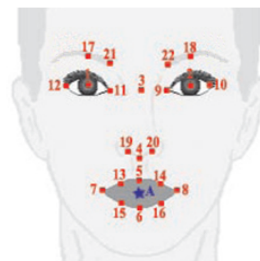

Figure 2: The feature points tracked by FaceTracker.

tracked facial feature point was drawn on black background. In addition, the inner and outer eyebrows were connected, the outline of the lips was added, and the outline of the eyes was drawn based on the corners of the eyes. (See Fig. 1, last panel)

To synthesize movements on both the avatar and android, our software first recognizes the activation of specific action units using a rule-based approach similar to those used by Pantic et al. [21] and el Kaliouby and Robinson [16]. In addition to AU activation we incorporate the intensity of AU during each frame, so that the onset, apex, and offset would all be visible. Knowing the facial properties of the person whose video is used for synthesis allowed us to model the intensity of AUs as well.

Our AU recognition approach was not limited to frontal images of the face, which differs from other approaches presented in the literature. This was possible because we knew the parameters for each actor's face a priori. Furthermore, our approach is able to deal with out-of-plane head motion as features for each of the actors were extracted at various orientations.

The extracted AUs were then used to animate the virtual avatar directly using FacePoser from Source SDK. This was possible to do without any translation because the animation system is AU-based.

For our android robot, motion was generated via a direct AU-to-motor mapping. However, because the physical motion of the robot was slightly slower than the real-time movement of the AUs, we accelerated the timing of the final video slightly to be the exact same length as video of the other representations.

Due to the importance of temporal information in perceiving emotions [8] the avatar and stick figure representations were ani-

1 http://developer.valvesoftware.com/wiki/Facial_Expressions_Primer
2 http://hansonrobotics.wordpress.com/

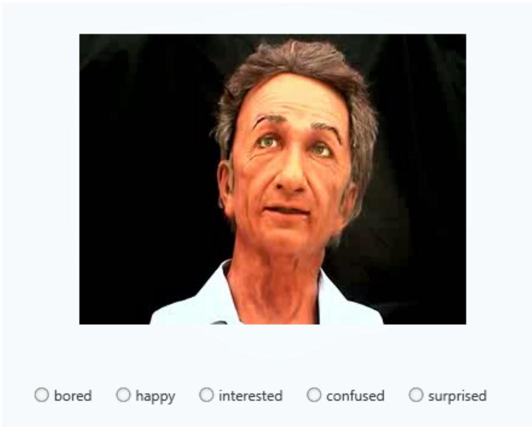

Figure 3: The labelling interface used in our experiment.

mated with a frame by frame correspondence to the original videos, due to technical limitations this was not possible on the robot.

2.2 Participants

Participants were recruited using a graduate e-mail list across the university. In total 15 people took part in the experiment, ten of them were female and five male. Their ages ranged from 21 to 34 (mean age 24.5, s.d.=3.4). All participants had normal or corrected to normal vision, and all considered themselves fluent English speakers. They were all given a £10 gift voucher for their participation in the experiment. All of the participants except for one were university students (both graduate and undergraduate).

2.3 Task

After seeing each video stimulus, participants had to choose one out of five possible labels (see Fig. 3). Participants could not choose a label before the video finished playing, and to move on to the next video participants had to choose only one of the five mental state labels, distractor labels were not present.

2.4 Procedure

Upon arrival, each participant was given a brief training labelling task to learn and practice before participating in the main experiment. The videos used in training did not appear in the main experiment to avoid learning effects.

In order to ensure participants knew the meanings of the labels they would be using in the experiment, each was given a sheet of dictionary definitions of all labels. Participants could refer to this sheet at any point during the trials.

The main experiment consisted of three trial blocks, each block consisting of 20 videos randomly selected from the total of 60. For a given stimulus, participants first saw a crosshair, then saw the video played twice, and then were asked to select a label. Participants were not allowed to replay the video.

The experimenter left the room during the main trial, allowing the participants to complete the task on their own.

2.5 Measures

Our dependent variable was label accuracy as measured via the Baron-Cohen taxonomy [5] (i.e., "ground truth"). Our independent variables were representation (actor, virtual stick figure, virtual avatar, or android robot), emotion (bored, confused, surprised, interested, and happy), and actor (actor 1, actor 2, actor 3).

3. RESULTS

Before conducting our overall analysis, we first checked to ensure actor choice did not influence overall labelling accuracy. We ran a χ^2 (chi-square) test, which showed a significant association between actor and correct labelling of the stimulus material, $\chi^2(2) = 7.15$, $p < .05$. However, Cramer's V was 0.089, indicating a weak association between actors [18]. Because the association was weak, in our subsequent analysis we feel confident not taking the actor into account in our subsequent analysis.

A two-way, repeated measures ANOVA was performed with representation and emotion as factors and the number of correct answers as the dependent variable. Mauchly's test indicated that the assumption of sphericity had been violated for the main effects of representation $*$ emotion, $\chi^2(77) = 109.5$, $p < .05$. Therefore degrees of freedom on this interaction were corrected using Greenhouse-Geisser estimates of sphericity ($\epsilon = .42$ for the main effect of representation $*$ emotion).

The mean accuracy for representations over all emotions was as follows: original 80%, robotic head 35%, avatar 36%, and stick figures 48%. There was a significant main effect of the type of representation on the number of correct answers $F(3, 42) = 48.13$, $p < .001$. Contrasts revealed that number of correct answers for original videos were significantly higher than those of robotic head, $F(1, 14) = 127.3$, $r = .95$, $p < .001$, avatars, $F(1, 14) = 132.3$, $r = .95$, $p < .001$, stick figures, $F(1, 14) = 35.4$, $r = .85$, $p < .001$. No significant differences were found when comparing the synthetic representations. All reported contrasts are Bonferroni corrected.

The mean accuracy for emotions across all representations were as follows: bored 48%, confused 46%, happy 50%, interested 42%, and surprised 62.8%. There was a significant main effect of the type of emotion on the number of correct answers $F(4, 56) = 5.05$, $p < .05$. Contrasts revealed that number of correct answers for *surprised* was significantly higher than those for *bored*, $F(1, 14) = 17.1$, $r = .67$, $p < .05$, *confused*, $F(1, 14) = 33.3$, $r = .84$, $p < .001$, and *interested*, $F(1, 14) = 18.8$, $r = .76$,$p < .001$. No significant differences were found between the other mental states. All reported contrasts are Bonferroni corrected.

The data was inspected for outliers amongst the participants, no participant stood out with either a very high or very low accuracy.

Accuracy of the labelling task was not affected by gender $\chi^2(1) = 1.81$, $p > .05$.

4. DISCUSSION

The purpose of our experiment was to investigate how much affective information automatically tracked facial feature points convey when synthesized across a range of synthetic representations that differ in their photo-realism.

As expected, the video stimuli featuring human actors had better labelling accuracy than stimuli featuring synthetic representations. Contrary to our expectation more photo-realistic synthetic representations did not lead to higher accuracy. Instead, we found no significant difference between the synthetic representations.

This result may be due to several different factors. Firstly, according to some of the participants, it was hard to tell the emotion of the virtual avatar or the robotic head because of conflicting emotional cues (i.e., the mouth region suggested one emotion while the eye area suggested another). Participants were thus sometimes diverted to areas of the face that were not particularly well tracked by our facial feature tracker, and thus poorly synthesised. In particular, the eye region was the most difficult, as our tracker has no way to detect the squinting and widening of the eyes. Thus, it might be beneficial to have geometrically simpler avatars if it will be impossible to properly animate more complex ones.

Our results are supported by Slater et al. [25] who argues that higher realism in avatar's appearance may lead to higher expectations for behavioural realism, which might not have been met in our stimulus videos, especially in the eye region.

Additionally, some participants reported not paying attention to the gaze of the stick figures, however, they were influenced by the gaze cues conveyed by the virtual avatar and robotic head. This was especially true for the *confused* and *interested* mental states. This exposes one limitation of our facial feature tracking system as it does not track iris locations very accurately.

An additional factor, at least for the virtual avatar and robotic head representations, may have been the lack of head motion. With the exception of orientation, out of plane head motion was not portrayed on the robotic head and the virtual avatar. This may have affected accuracy, because according to Ekman and Oster [15], head position can play a role in recognizing some mental states, such as *interest*.

Our results support the fact that complex mental states are more difficult to recognise than basic emotions, as this disparity was seen across all representations. This suggests that while some emotions, such as *surprise*, can be well recognised using just geometrically tacked feature points, synthesizing more complex mental states may require more complex algorithms which incorporate other features, such as appearance, head position, and gaze information.

When comparing our results to those found by Afzal et al. [1], we find a contradiction. Their results showed a statistically significant difference in accuracy across representations (stick-figure was best, followed by point-light, followed by avatars), whereas we did not. This may be explained by the fact that their animation system seemed somewhat limited by animation quality, thus leading to lower accuracy rates.

One thing we found in common with the results of Afzal et al. [1] was that participants were better at recognising *surprise* across all representations, which suggests that geometric approach to facial expression recognition and synthesis is better for some expressions than others.

The results of this experiment have implications for facial expression synthesis, as they illustrate the importance of consistency when representing expression. In particular, it is important to ensure that all parts of the face “tell the same story", including gaze, head position, and orientation.

An interesting extension to the experiment might be the systematic manipulation of timing and intensities of expressions and seeing how they influence the perceived emotions or naturalness of behaviour in avatars and robot head. In future studies it would be beneficial to evaluate the perceptual realism of the synthetic forms used.

5. ACKNOWLEDGMENTS

We acknowledge funding support from EPSRC and Thales Research and Technology (UK), and the Qualcomm Research Studentship in Computing. We would also like to thank Graham Titmus and Per Ola Kristensson.

6. REFERENCES

[1] S. Afzal, T. M. Sezgin, Y. Gao, and P. Robinson. Perception of Emotional Expressions in Different Representations Using Facial Feature Points. In *ACII 2009*.

[2] J. Ahlberg, I. S. Pandzic, and L. You. *MPEG-4 Facial Animation*, chapter Evaluating MPEG-4 Facial Animation Players, pages 287–291. John Wiley & Sons, 2003.

[3] N. Ambady and R. Rosenthal. Thin Slices of Expressive behavior as Predictors of Interpersonal Consequences : a Meta-Analysis. *Psychological Bulletin*, 111(2):256–274, 1992.

[4] A. Asthana, J. M. Saragih, M. Wagner, and R. Goecke. Evaluating AAM Fitting Methods for Facial Expression Recognition. In *ACII 2009*.

[5] S. Baron-Cohen, O. Golan, S. Wheelwright, and J. J. Hill. Mind Reading: the interactive guide to emotions. London: Jessica Kingsley Limited, 2004.

[6] S. Baron-Cohen, A. Riviere, M. Fukushima, D. French, J. Hadwin, P. Cross, C. Bryant, and M. Sotillo. Reading the Mind in the Face: A Cross-cultural and Developmental Study. *Vis Cogn*, 3:37–60, 1996.

[7] M. S. Bartlett, G. Littlewort, M. G. Frank, C. Lainscsek, I. R. Fasel, and J. R. Movellan. Automatic Recognition of Facial Actions in Spontaneous Expressions. *Journal of Multimedia*, 1(6):22–35, 2006.

[8] J. N. Bassili. Facial motion in the perception of faces and of emotional expression. *Journal of Experimental Psychology: Human Perception and Performance*, 4(3):373 – 379, 1978.

[9] N. Bee, B. Falk, and E. André. Simplified Facial Animation Control Utilizing Novel Input Devices: A Comparative Study. In *IUI 2009*, pages 197–206.

[10] M. L. Cascia, S. Sclaroff, and V. Athitsos. Fast, Reliable Head Tracking under Varying Illumination: An Approach Based on Registration of Texture-Mapped 3D Models. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(4):322–336, 2000.

[11] J. Cohn and K. Schmidt. The timing of facial motion in posed and spontaneous smiles. *IJWMP*, 2:1 – 12, March 2004.

[12] E. Costantini, F. Pianesi, and P. Cosi. Evaluation of Synthetic Faces: Human Recognition of Emotional Facial Displays. In *Proceedings of Tutorial and Research Workshop "Affective Dialogue Systems"*, 2004.

[13] M. Courgeon, S. Buisine, and J.-C. Martin. Impact of Expressive Wrinkles on Perception of a Virtual Character's Facial Expressions of Emotions. In *IVA*, pages 201–214, 2009.

[14] C. Darwin. *The Expression of The Emotions in Man and Animals*. London, John Murray, 1872.

[15] P. Ekman, W. V. Friesen, and P. Ellsworth. *Emotion in the Human Face*. Cambridge University Press, second edition, 1982.

[16] R. el Kaliouby and P. Robinson. Real-Time Inference of Complex Mental States from Facial Expressions and Head Gestures. In *Real-Time Vision for Human-Computer Interaction*, pages 181–200. Springer US, 2005.

[17] M. Garau, M. Slater, V. Vinayagamoorthy, A. Brogni, A. Steed, and M. A. Sasse. The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment. In *CHI '03*, pages 529–536, 2003.

[18] F. J. Gravetter and L. B. Wallnau. *Statistics for the Behavioral Sciences*. Cengage Learning, 2006.

[19] U. Hess, R. B. Adams, and R. E. Kleck. The face is not an empty canvas: how facial expressions interact with facial appearance. *Phil Trans of the Royal Soc B*, 364(1535):3497–3504, December 2009.

[20] I. S. Pandzic and R. Forchheimer, editors. *MPEG-4 Facial Animation: The Standard, Implementation and Applications*. John Wiley & Sons, Inc., New York, NY, USA, 2003.

[21] M. Pantic and L. Rothkrantz. Facial Action Recognition for Facial Expression Analysis from Static Face Images. *IEEE Trans Syst Man Cybern B Cybern*, 34(3):1449–1461, June 2004.

[22] A. Raouzaiou, N. Tsapatsoulis, K. Karpouzis, and S. Kollias. Parameterized facial expression synthesis based on MPEG-4. *EURASIP J. Appl. Signal Process.*, 2002(1):1021–1038, 2002.

[23] P. Rozin and A. B. Cohen. High frequency of facial expressions corresponding to confusion, concentration, and worry in an analysis of naturally occurring facial expressions of Americans. *Emotion*, 3(1):68 – 75, 2003.

[24] J. A. Russell and A. Mehrabian. Evidence for a three-factor theory of emotions. *Journal of Research in Personality*, 11(3):273 – 294, 1977.

[25] M. Slater and A. Steed. Meeting people vitually: experiments in shared virtual environments. pages 146–171, 2002.

[26] Q. Zhang, Z. Liu, B. Guo, D. Terzopoulos, and H.-Y. Shum. Geometry-Driven Photorealistic Facial Expression Synthesis. *IEEE Transactions on Visualization and Computer Graphics*, 12:48–60, 2006.