Evaluation of Selected APIs for Emotion Recognition from Facial Expressions

Krzysztof Kutt[0000−0001−5453−9763], Piotr Sobczyk, and Grzegorz J. Nalepa[0000−0002−8182−4225]

Jagiellonian Human-Centered Artificial Intelligence Laboratory (JAHCAI) and Institute of Applied Computer Science, Jagiellonian University, Kraków, Poland
krzysztof.kutt@uj.edu.pl, gjn@gjn.re

Abstract. Facial expressions convey the vast majority of the emotional information contained in social utterances. From the point of view of affective intelligent systems, it is therefore important to develop appropriate emotion recognition models based on facial images. As a result of the high interest of the research and industrial community in this problem, many ready-to-use tools are being developed, which can be used via suitable web APIs. In this paper, two of the most popular APIs were tested: Microsoft Face API and Kairos Emotion Analysis API. The evaluation was performed on images representing 8 emotions—anger, contempt, disgust, fear, joy, sadness, surprise and neutral—distributed in 4 benchmark datasets: Cohn-Kanade (CK), Extended Cohn-Kanade (CK+), Amsterdam Dynamic Facial Expression Set (ADFES) and Radboud Faces Database (RaFD). The results indicated a significant advantage of the Microsoft API in the accuracy of emotion recognition both in photos taken en face and at a 45 degree angle. Microsoft's API also has an advantage in the larger number of recognised emotions: contempt and neutral are also included.

Keywords: Affective computing · Facial expression · Emotions · Benchmark datasets.

1 Introduction and Motivation

American psychologist Albert Mehrabian published in 1968 a rule known as "7-38-55" [16]. According to it, the content of an utterance alone conveys 7% of the emotional state, the tone of voice accounts for 38%, while non-verbal communication (facial expressions, gestures) accounts for as much as 55%. This indicates that the ability to recognise emotions from the face is highly important [4]. On the other hand, it should be pointed out that emotional expressions are fairly universal, i.e., they are recognised by members of different cultures – even those that have had no contact with each other [5]. Both of these factors point to the need to include emotion processing from facial expressions in information systems that deal with emotional information processing, i.e., affective computing (AfC) systems [18].

Deciphering emotions from facial expressions is a complex issue. Extensive time has been devoted to it by Paul Ekman, whose efforts were summarised in the Facial Action Coding System (FACS), developed in 1978 [3] and later updated in 2002 [6]. In the FACS system, facial expressions are encoded by more than 40 action units (AUs), which are the visually smallest muscle movements that cause changes in face expression. According to Ekman, each basic emotion is characterised by a unique expressive pattern, which may vary slightly for variants of the same emotion. As an example, there are over 60 facial expressions for anger, although all of them have two features in common: lowered eyebrows and tightened lips [3]. FACS has some limitations, among others, it is pointed out that action units are local patterns, whereas facial expressions are the result of the interaction of structures. FACS also does not take into account the temporal perspective of facial changes [18].

An alternative approach in recognizing emotions from facial expressions was proposed in [7]. This system is based on recognising emotions from the flow of movement of the whole face, rather than synthesising them from the movement of individual muscles. For this purpose, a geometric model of the facial shape was used, which was then superimposed on the face and the detected emotions were assigned based on the movement patterns of the grid. In addition to the two systems mentioned above, there are a number of other solutions (see, e.g., [2,9]). Among them, the most interesting from the point of view of creating practical systems seem to be commercial solutions, in which access to an appropriate API is provided, to which one sends photos/videos containing emotional expressions and in response receives feedback on emotions detected by the system. Thanks to this, implementation of a solution based on facial emotion recognition does not require large resources to run the model, e.g., on mobile phones or in wearable systems – only constant access to the Internet is needed.

Our research is aimed at preparing a toolkit for developing personalised intelligent affective systems [11]. The core assumption is to use affordable wearables as the basis for the whole mechanism. This requires a lightweight framework, which will not have high hardware requirements and will be adequately robust for the assumed applications [17]. This straight leads to the choice of using one of the available APIs for emotion recognition. The aim of the work described in this paper was to verify two such solutions – the key to their selection was the availability of a free version that allows testing. The results obtained will be the basis for further work, in which the usability of the selected API will be evaluated on data collected in BIRAFFE (*Bio-Reactions and Faces for Emotion-based Personalization for AI Systems*) series of experiments [12,13].

The rest of paper is organized as follows. General approach for emotion recognition from facial expressions is discussed in Sect. 2. Then, in Sect. 3, the datasets selected for experiments are outlined. Evaluation results are presented in Sect. 4. The paper is concluded in Sect. 5.

2 Emotion Recognition from Facial Expressions

The general automatic emotion recognition systems based on facial expressions consists of detecting a face in an image, extracting its features, and finally classifying them [1]. Various challenges arise in this process, for example, the face may not be captured centrally but from a semi-profile; individual differences in face shape and appearance make generalisation difficult; the classifier may be overfitting or underlearning due to insufficient teaching examples [1]. Assuming that a facial expression is a manifestation of a perceived emotion, the problem of emotion recognition can be reduced to the problem of pattern recognition. Supervised learning methods are usually used for this purpose [10].

The first step is to detect the location of the face in the image. In the case of facing the camera, this can be done using the Viola-Jones detector [21]. Its algorithm is based on extracting Haar features for each part of the image, i.e., small areas containing a vertical, horizontal or diagonal line. This is necessary to calculate whether a photo contains a face, but it is very inefficient due to the need to check a large number of features. To solve this problem, so-called cascades of classifiers were introduced: first checking which features have the biggest influence on the final classification and then grouping them into subsequent steps of the matching algorithm. In this way, if a potential face is not found in a window in the first step, the rest of the features no longer need to be checked in that window. [21] suggested that this face search algorithm has very good performance. To avoid the effect of face rotation on emotion recognition, the detected face needs to be normalised, i.e., transformed to a standard size and orientation.

Once the face is found, the next step is feature extraction based on determining the position of landmarks such as the contours of the eyes and eyebrows, the corners of the mouth, and the tip of the nose. A geometric grid is superimposed on the face, constructed from the landmarks of the universal neutral face model. Differences between the models are recalculated by the classifier to determine a specific emotion. Information about the position and orientation of key features provides input to the classifier algorithms, which return action units or final recognised emotional states as output. Most approaches to learning a classification model boil down to supervised learning, where emotions are labelled in a learning set [1,10].

There are many solutions on the market that analyse emotions based on facial expressions, working along the above described scheme. The industry is constantly growing, offering commercial applications to study the perception of advertisements, the user's sense of satisfaction, and even to check the level of anger in public places such as stadiums or airports to prevent possible dangerous events. Since creating software that recognises emotions requires having a large amount of data for machine learning, companies offering such solutions primarily aim to hide the business logic of the application. This is most often done through the use of microservice architecture: the software that recognises emotions based on a face photo runs as a service on the company's server, accessible by the API, which makes it possible to integrate this solution into your

own application while having no knowledge of the specific operation of the system. It is important to note that API-based systems are usually characterised by limited availability (daily or monthly transaction limit). In addition to application access, systems may also differ in the number of emotions detected, moreover, each system has its own capabilities and limitations. For the purpose of this research, two comprehensive systems for emotion recognition based on facial expressions were selected, differing in their specifications and the emotion model used: Microsoft Face API and Kairos Emotion Analysis API. Only free variants of the above-mentioned tools were used in the experiments.

2.1 Microsoft Face API

Microsoft Cognitive Services[1] is a collection of various services including image recognition, photo identification, voice verification or intelligent recommendation systems. Among these services is the Microsoft Face API[2] for emotion detection based on facial expressions. The API takes as input an image in which it seeks a face. Then, for each face it finds—based on its expression—it returns the confidence level for each emotion in the recognised set and the coordinates of the rectangle bounding the face field in JSON format. The detected emotions are: anger, contempt, disgust, fear, joy, sadness and surprise and—additionally—neutral. During interpretation of the returned results, the emotion with the highest score should be understood as the detected emotion. All emotion scores are summed to one. Microsoft agrees not to publish the submitted data or give access to it to other users, although it reserves the right to use the images to improve its services.

2.2 Kairos Emotion Analysis API

Kairos is a company offering services related to demographic data analysis—including emotion recognition—through the use of vision systems and machine learning. The Kairos Emotion Analysis API[3] runs as a web service available in a REST architecture. The service allows recognition of emotions not only from a photo, but also from videos. A photo containing a face is uploaded to the Kairos Emotion Analysis API. The found facial features and expressions are processed by algorithms, returning in response values corresponding to the recognised emotions and the locations of facial feature points in the image. The service recognises six basic emotions: anger, disgust, fear, joy, sadness, surprise. The confidence level of the detected emotions ranges from 0 to 100 and, as with the Microsoft API, the one with the highest score should be interpreted as the recognised emotion.

[1] See: https://azure.microsoft.com/en-us/services/cognitive-services/.
[2] See: https://azure.microsoft.com/en-us/services/cognitive-services/face/.
[3] See: https://www.kairos.com/docs/api/.

3 Datasets

Standardised sets of face images labelled by experts were used for the experiments. It was decided to select the 4 most popular datasets:

- *Cohn-Kanade (CK)* [8]. It is a dataset created at Carnegie Mellon University in 2000 consisting of a series of posed photographs, taken at short intervals, starting with a neutral facial expression and ending with a face corresponding to a certain emotion. The last photo in the series is also coded using the updated FACS system [6] and labelled with the name of the emotion. The label itself refers to the facial expression the subjects were asked to express, not to the one actually presented. The collection contains a total of 585 image series, representing the 6 basic emotions. Ninety-seven students aged between 18 and 30 were selected to create it, 65% of whom were female. In terms of ethnicity, 15% were African-American, 3% Asian or Hispanic, and the rest Euro-American. The images were taken straight on with uniform lighting and are available in greyscale at 640x490 pixel resolution.
- *Extended Cohn-Kanade (CK+)* [15]. This is the second version of the CK dataset released in 2010. In addition to posed photographs, a certain group of spontaneous expressions was included. Compared to CK, the group of subjects and the number of photo series taken has been increased. The photos are also coded in FACS and labelled with emotions, but the labels have been verified by experts. The facial expressions captured represent 8 states: 6 basic emotions, contempt and neutrality. A 593 image series covering the photographs of 123 people was recorded. The images are provided in 640x490 pixel resolution. Most of them are in grayscale, but there are also some in colour.
- *Amsterdam Dynamic Facial Expression Set (ADFES)* [20]. This dataset, created at the University of Amsterdam, consists of 648 MPEG-2 videos lasting approximately 6 seconds. The videos were recorded simultaneously from two angles: straight ahead and from a 45°angle. In addition, the database includes frames extracted from videos recorded frontally, demonstrating people expressing emotion at their peak. These are colour JPEG images with a resolution of 720x576 pixels. The people recorded are of different ethnicities, being Europeans, Africans and Turks between the ages of 18 and 25. The subject group consisted of 10 women and 12 men. They were trained so that their facial expressions corresponded to a prototype of a certain emotion according to the FACS system [6]. A major advantage of the ADFES database is the number of emotions presented – in addition to the 6 basic ones, each of the subjects presented a neutral face, contempt, as well as pride and shame.
- *Radboud Faces Database (RaFD)* [14]. The RaFD collection, created in 2010 at Radboud Universiteit in Nijmegen, consists of photos of 67 models (Europeans and Africans), of which 38 are men, 19 – women, 4 – boys and 6 – girls. Each person was trained by an expert to be able to make facial expressions corresponding to eight emotional states according to FACS [6]: 6 basic, contempt and neutral. The photograph of each emotional expression

was simultaneously taken from different perspectives: en face, in profile and from a 45°angle. The entire collection consists of 8040 colour images with a resolution of 681x1024 pixels.

4 APIs' Evaluation

A series of experiments were conducted using the free version of both APIs. Images were sent one at a time, response sent by the API was received, and then the detected emotion was compared with the emotion with which the image was labelled. For each of the tests—one test is an evaluation of one API on a single emotion tested on images from one set—accuracy score was calculated by dividing the number of correctly assigned labels to all images used in the test. The Kairos API does not recognise a neutral emotion, so for the purposes of testing it was assumed that the "no emotion" response returned by this API would be treated as a neutral emotion.

Due to the large number of images in the learning sets and the limitations of the free versions of the API, for the purpose of the tests, from each of the available datasets a subset was selected on which the evaluation was performed:

- for each of the series of images (one subject expressing one emotion), a single image was selected in which the emotion reached its peak of expression,
- in CK+ dataset, only images annotated by FACS experts were considered.

Finally, the evaluation was carried out on the following number of pictures:

- CK dataset: sadness – 128, joy – 103, surprise – 103, fear – 66, anger – 45, and disgust – 42;
- CK+ dataset: surprise – 83, joy – 69, disgust – 59, anger – 45, sadness – 28, fear – 25, and contempt – 18;
- ADFES dataset: 21 photos for surprise and 22 images for each of the other conditions: anger, contempt, disgust, fear, joy, sadness and neutral;
- RaFD dataset: 67 *en face* images for each condition: surprise, anger, contempt, disgust, fear, joy, sadness and neutral. A second set of the same size was also selected, in which the subjects were photographed at 45°.

Based on the obtained accuracy results shown in Fig. 1 and 2, and on the analysis of the confusion matrix (not included in the article, due to limited space), the following observations can be drawn:

1. Microsoft Face API dominates for most emotions and most datasets. The Kairos Emotion Analysis API achieves higher performance only for anger.
2. Only the Microsoft Face API handled photos that were not taken *en face*.
3. In the CK and CK+ datasets, the Kairos Emotion Analysis API most often had problems detecting faces: for these datasets, either no face or no emotion message appeared most often.
4. Both systems perform well in identifying joy and surprise, while anger, fear and sadness receive generally low accuracy scores.

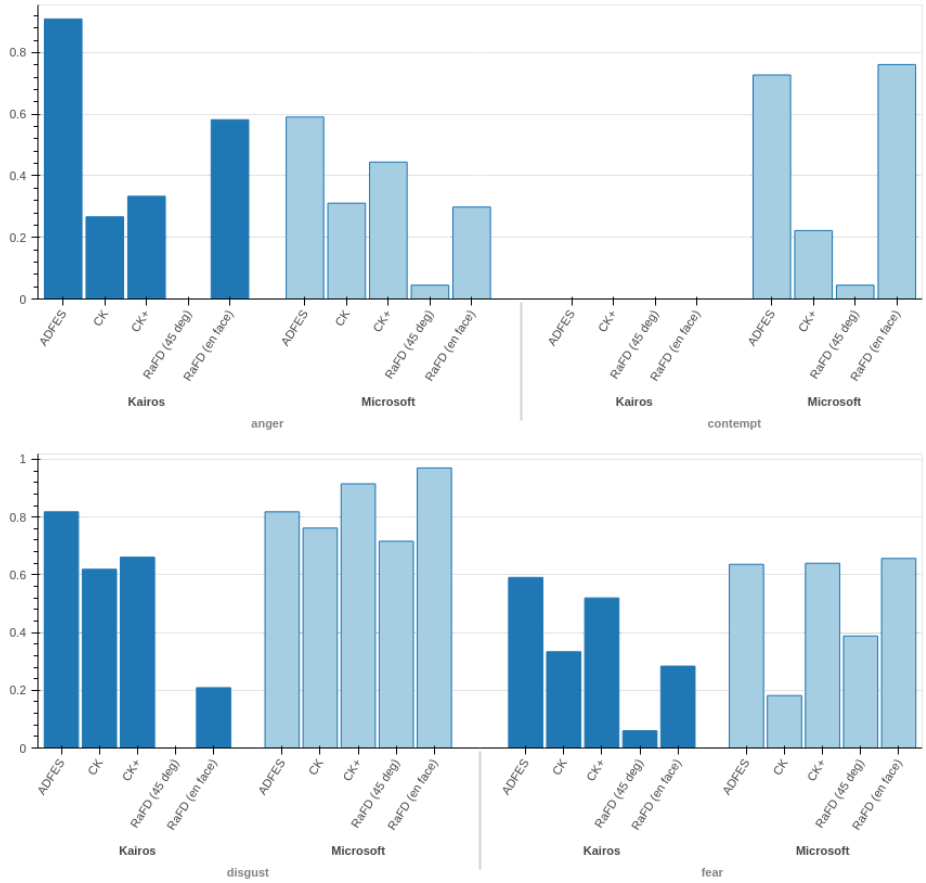

Fig. 1. Accuracy scores for each tested scenario: 8 emotions, 2 frameworks, 5 datasets (cont. on Fig. 2).

5. Fear is often mistaken for surprise and sadness for no emotion.
6. The model used in the Kairos system does not include contempt, so this API is unable to recognise this emotion (hence the 0 values for all tests).
7. ADFES dataset includes also faces tagged as pride and shame. Both systems are unable to recognise these emotions so this is not included in the figures, however we have considered these images in the evaluation. The confusion matrix analysis indicates that pride is often recognised as joy and shame as neutrality. This pattern is confirmed by Russell's two-dimensional emotion model [19] – pride and joy are adjacent in this space, whereas shame can be confused with neutrality because it lacks a specific facial muscle representation [20].

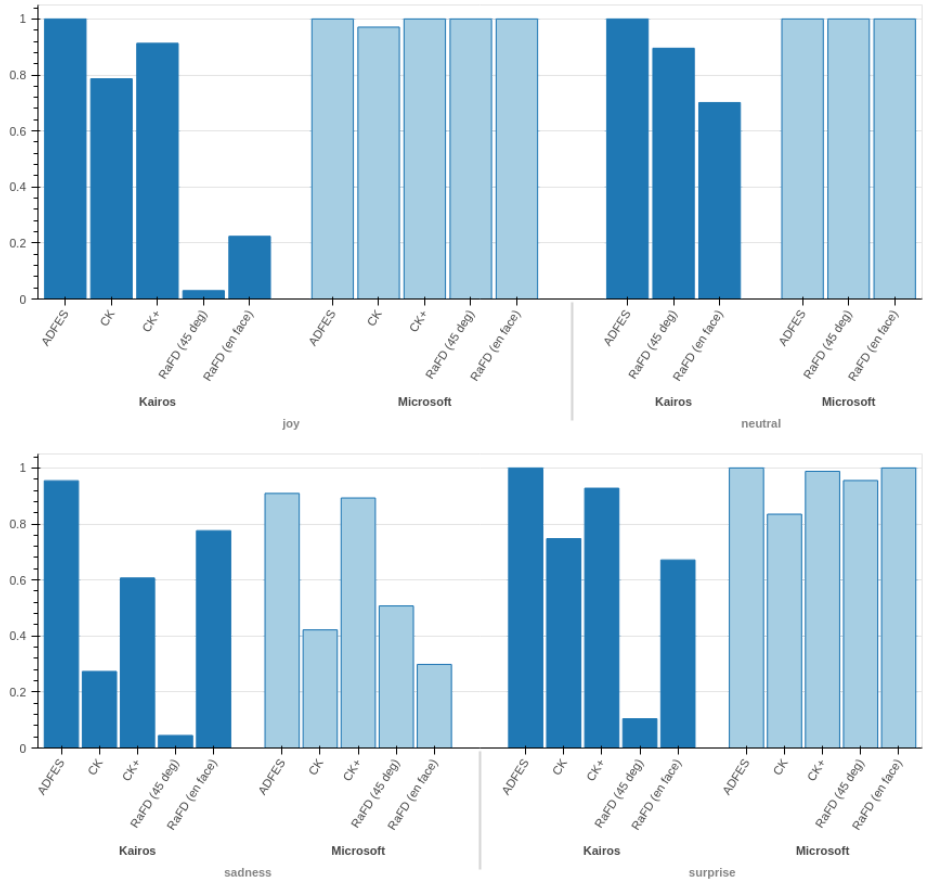

Fig. 2. Accuracy scores for each tested scenario: 8 emotions, 2 frameworks, 5 datasets (cont. from Fig. 1).

5 Summary and Future Work

This paper presents a detailed comparison of two popular web APIs for emotion recognition from facial expressions: Microsoft Face API and Kairos Emotion Analysis API. The evaluation was performed on images taken from 4 benchmark datasets: Cohn-Kanade (CK), Extended Cohn-Kanade (CK+), Amsterdam Dynamic Facial Expression Set (ADFES) and Radboud Faces Database (RaFD). The results indicated a significant advantage of the Microsoft API in the accuracy of emotion recognition both in photos taken en face and at a 45 degree angle. Microsoft's API also has an advantage in the larger number of recognised emotions: contempt and neutral are also included.

In the benchmark datasets used in this study, subjects were explicitly asked to produce a specific emotional expression, making the facial changes highly ex-

pressive. As part of future work, we plan to evaluate both APIs on the data obtained in the BIRAFFE1 and BIRAFFE2 experiments [12,13]. In contrast to the benchmark datasets used in the present work, in the BIRAFFE datasets emotional expressions appeared as a *by-product* of the execution of the experimental protocol related to the evaluation of emotional stimuli and to the playing of affective games. This will make these expressions more similar to natural ones and present in everyday life, so that it will be possible to obtain information about the effectiveness of APIs in more ecological conditions, which is important for the design of affective user interfaces.

References

1. Cohn, J.F., De La Torre, F.: Automated face analysis for affective computing. In: Calvo, R.A., D'Mello, S., Gratch, J., Kappas, A. (eds.) The Oxford Handbook of Affective Computing, pp. 131–150. Oxford University Press (2015). https://doi.org/10.1093/oxfordhb/9780199942237.013.020, https://doi.org/10.1093/oxfordhb/9780199942237.013.020
2. Deshmukh, R.S., Jagtap, V.: A survey: Software api and database for emotion recognition. In: 2017 International Conference on Intelligent Computing and Control Systems (ICICCS). pp. 284–289 (2017). https://doi.org/10.1109/ICCONS.2017.8250727
3. Ekman, P., Friesen, W.V.: Facial Action Coding System: A Technique for the Measurement of Facial Movement. Consulting Psychologists Press, Palo Alto (1978)
4. Ekman, P., Rosenberg, E. (eds.): What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS). Oxford University Press (2005)
5. Ekman, P., V. Friesen, W.: Constants across cultures in the face and emotion. Journal of personality and social psychology **17**(2), 124–129 (1971)
6. Ekman, P., V. Friesen, W., C. Hager, J.: Facial Action Coding System (FACS): Manual. A Human Face, Salt Lake City (2002)
7. Essa, I.A., Pentland, A.P.: Coding, analysis, interpretation, and recognition of facial expressions. IEEE Transactions on Pattern Analysis and Machine Intelligence **19**(7), 757–763 (1997)
8. Kanade, T., Cohn, J.F., Tian, Y.: Comprehensive database for facial expression analysis. In: Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition (FG'00). pp. 46–53. Grenoble, France (2000)
9. Khanal, S.R., Barroso, J., Lopes, N., Sampaio, J., Filipe, V.: Performance analysis of microsoft's and google's emotion recognition API using pose-invariant faces. In: Proceedings of the 8th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion, DSAI 2019, Thessaloniki, Greece, June 20-22, 2018. pp. 172–178. ACM (2018). https://doi.org/10.1145/3218585.3224223, https://doi.org/10.1145/3218585.3224223
10. Konar, A., Chakraborty, A. (eds.): Emotion Recognition: A Pattern Analysis Approach. John Wiley & Sons, New Jersey (2015)
11. Kutt, K., Drążyk, D., Bobek, S., Nalepa, G.J.: Personality-based affective adaptation methods for intelligent systems. Sensors **21**(1), 163 (2021). https://doi.org/10.3390/s21010163, https://doi.org/10.3390/s21010163

12. Kutt, K., Drążyk, D., Jemioło, P., Bobek, S., Giżycka, B., Fernández, V.R., Nalepa, G.J.: BIRAFFE: Bio-reactions and faces for emotion-based personalization. In: AfCAI 2019: Workshop on Affective Computing and Context Awareness in Ambient Intelligence. CEUR Workshop Proceedings, vol. 2609. CEUR-WS.org (2020)
13. Kutt, K., Drążyk, D., Szelążek, M., Bobek, S., Nalepa, G.J.: The BIRAFFE2 experiment. study in bio-reactions and faces for emotion-based personalization for AI systems. CoRR **abs/2007.15048** (2020), https://arxiv.org/abs/2007.15048
14. Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D., Hawk, S., van Knippenberg, A.: Presentation and validation of the radboud face database. Cognition & Emotion **24**(8), 1377–1388 (2010)
15. Lucey, P., Cohn, J.F., Kanade, T., Saragih, J., Ambadar, Z., Matthews, I.: The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In: Proceedings of the Third International Workshop on CVPR for Human Communicative Behavior Analysis (CVPR4HB 2010). pp. 94–101. San Francisco, USA (2010)
16. Mehrabian, A.: Communication without words. Psychology Today **2**(4), 53–56 (1968)
17. Nalepa, G.J., Kutt, K., Bobek, S.: Mobile platform for affective context-aware systems. Future Generation Computer Systems **92**, 490–503 (mar 2019). https://doi.org/10.1016/j.future.2018.02.033, https://doi.org/10.1016/j.future.2018.02.033
18. Picard, R.W.: Affective Computing. MIT Press, Cambridge, MA (1997)
19. Russell, J.A.: A circumplex model of affect. Journal of Personality and Social Psychology **39**, 1161–1178 (1980)
20. van der Schalk, J., Hawk, S., Fischer, A., Doosje, B.: Moving faces, looking places: Validation of the amsterdam dynamic facial expression set (adfes). Emotion **11**(4), 907–920 (2011)
21. Viola, P., Jones, M.: Robust real-time object detection. In: International Journal of Computer Vision. vol. 57 (2001)