ORIGINAL ARTICLE
Journal Section

Facial emotion recognition with a reduced feature set for video game and metaverse avatars

Darren Bellenger, Minsi Chen and Zhijie Xu[1*]

[1]Department of Computer Science, University of Huddersfield, UK

Correspondence
Darren Bellenger, Department of Computer Science, University of Huddersfield, UK
Email: darren.bellenger@hud.ac.uk

Funding information
CGI Group

This paper presents a novel real-time facial feature extraction algorithm, producing a small feature set, suitable for implementing emotion recognition with online game and metaverse avatars. The algorithm aims to reduce data transmission and storage requirements, hurdles in the adoption of emotion recognition in these mediums. The early results presented show a facial emotion recognition accuracy of up to 92% on one benchmark dataset, with an overall accuracy of 77.2% across a wide range of datasets, demonstrating the early promise of the research.

KEYWORDS
emotion recognition, metaverse, virtual reality, online games

1 | INTRODUCTION

As a society, we increasingly live our lives within online mediums, a prime example being the growing popularity of online games and metaverse platforms [1]. Future metaverses may eventually integrate into all aspects of our lives, from avatars embodying our physical appearance or how we want to be perceived to look, to ownership of virtual assets and currency [2]. Dionisio et al. [3], Hughes [4] define the metaverse as a "network of 3D virtual worlds focused on social connection". For this paper the term will also be used to encompass other types of virtual worlds [5].

This paper presents a novel method of integrating facial emotion recognition into metaverse platforms. This utilises facial data already held within modern 3d avatars [6], thus allowing emotion recognition to be implemented within metaverse platforms reaching a global audience. Whilst implementing emotion recognition within online mediums raises ethical concerns [7], there are many benefits to people's lives, gained from the use of this technology.

The method is illustrated in the example shown in Figure 1. Here users are conversing in a metaverse, embodied within personalised 3d avatars. Four users have access to hardware that allows for facial expressions to be visualised onto their avatars, functionality already available in some online games [8, 9]. The hardware for achieving facial capture could be either an inexpensive webcam [10], or via the facial capture technology on next generation Virtual Reality (VR) headsets (HMDs) [11, 12, 13, 14].

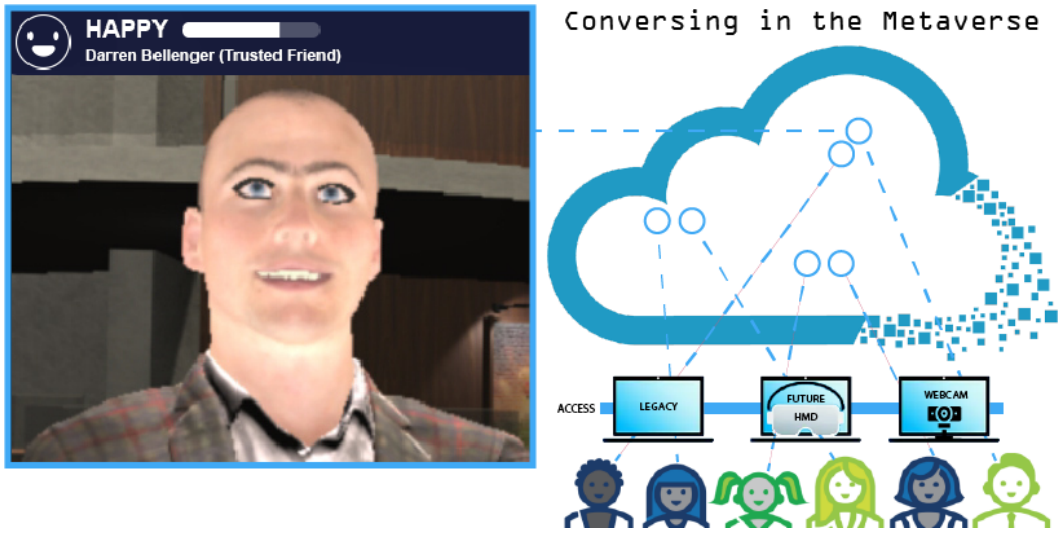

FIGURE 1 An example scenario using the method where a large number of users are online

With this novel method users could be offered an onscreen indicator of the emotion others are exhibiting, gained from real-time avatar data, without adversely affecting metaverse performance. With some metaverses hosting tens of thousands of concurrent users [15], visualising and then storing facial avatar data, could realise many use cases, discussed in the next section. The proposal contrasts with existing examples of metaverse-based emotion recognition, where prediction is limited to the local personal computer setup the user is accessing [16, 17, 18]. Along with this, these examples only involve a single user accessing a metaverse, there being no attempt at scaling these solutions. The problem of scalability is an issue this proposal addresses, thus solving a barrier to the adoption of emotion recognition within games and metaverses.

2 | THE CASE FOR EMOTION WITHIN GAMES AND THE METAVERSE

There are compelling use cases for integrating emotion recognition within online games and metaverses, including:

- **Evolution of games**. Research has covered the capacity for using emotion within games to increase addictiveness, help create online connections and relationships [19, 20].
- **Wellness technologies**. Utilising emotion within a metaverse may help improve emotion recognition in children, particularly those with autism, compared with traditional methods [21, 22, 23], and building on early VR research [24, 25, 26].
- **Improved training simulations**. Real-time tracking of learner emotion may allow for later analysis of stress levels

[27]. This tracking may also improve immersion within soft skills simulation training, which has become increasingly important within the military training arena [28].

- **Opinion mining**. Customer's attitude to products could be tracked and analysed (sentiment analysis), either during online games advertising, or during simulation of, for example, a supermarket setup [29, 30].

The method may better suit new advances in VR HMDs, designed to provide facial input into games and metaverses [11]. The Oculus Half-Dome research project is one such example of a new design of HMD that incorporates sensors that provide data on the changing facial expressions of a user [12]. This data is reduced from that provided by a human image, and is aimed at providing the minimum data required to accurately visualise an avatar's face [31]. Other companies are integrating EEG and ECG biosensors into HMDs, which also use a reduced feature set [32], allowing researchers to also investigate into using this data to predict emotion [33].

Other advances in online technologies may also adopt the proposal. Nivida have developed a replacement for traditional video conferencing, where high-quality avatars of people are animated using facial landmark data generated from a webcam [34, 35, 36]. This technology could benefit from this proposal, offering emotion prediction from this avatar data. Future Augmented Reality (AR) applications are another area where this proposal may be of benefit, due to the reduction in the amount of stored emotional data [37]. This is in part due to the lower data storage and processing power of AR devices currently [38].

The researcher's main driver in integrating emotion recognition into a metaverse, is to improve its usefulness as an educational environment. In particular for people with emotion recognition difficulties, such as autism. The educational benefits of metaverses are already wide-scale [39], from improving trust within teams [40], to acting as a social leveller [41] and a social experimentation platform [42]. Students are now being noted to see no separation between playing and learning [40]. Earlier research has looked at improving social skills [24] and emotion recognition [25], yet without using real-time emotion recognition. In relation to the general use of metaverses within healthcare it can lower costs and allow for experiential learning [43]. The researcher intends on investigating the effects of metaverse based emotion recognition on teenage children (13-19), who currently enjoy online games and metaverses from home [44]. These will invariably be on the autistic spectrum, given autistic teenagers are known to be heavy users of online games and metaverses [45], being shown to enjoy them [26].

In attempting to improve education for people with emotion recognition difficulties, there are specific medical issues, exhibited in everyday life, that this proposal may alleviate. A lack of social skills can hinder future careers in the workplace [46, 47]. Emotional Intelligence (EI) is affected, which is increasingly important in business, given its transferability [48, 49]. Careers that require reading and responding to emotion [50] may be affected, as well as the ability to raise one's own self-esteem [51]. People may even by viewed as impersonal [52] if they lack positive reactions in meetings [53]. Those with emotion recognition deficit may struggle to sustain conversations [54], or maintain an attention span [49].

With the proposal being usable on both HMDs or webcam input, there is the option of using inexpensive webcams [10, 55]. This can help in the promotion of metaverse based education in areas of deprivation [56, 57, 58]. New metaverse platforms are already allowing users to use a variety of access methods, from expensive HMDs to lower cost effective methods [59, 60].

3 | RELATED WORKS

An important part of any algorithm is the input required to predict an emotion. Historically algorithms required human image-based input, from a picture, video or webcam, to predict an emotion [61]. Tian et al. [62] outline this input takes the form of landmark point data from permanent features of the face, "brows, eyes, mouth", along with optional transient features "lines and furrows, shown in Figure 2. Khadoudja and Caplier [63] note transient features indicate movement that is difficult to detect with point data, such as the wrinkling involved with negative emotions.

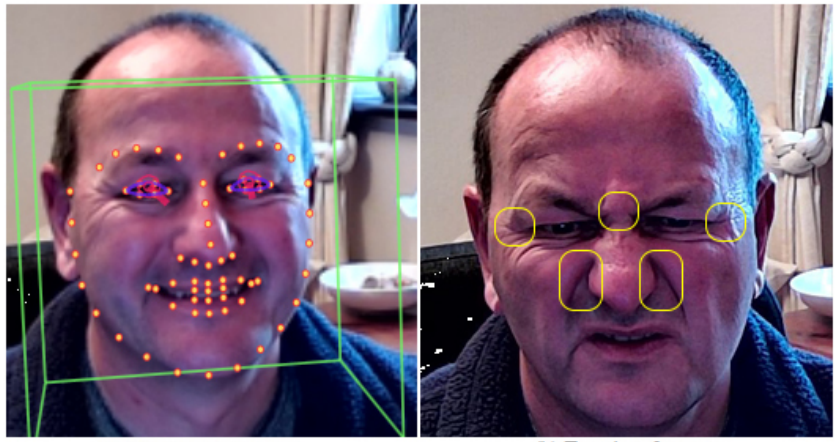

FIGURE 2 Examples of data required by emotion recognition algorithms [62].

Historical research that uses permanent and transient features do claim to achieve very high accuracy levels. Liliana [64] reports an accuracy of 92.81% training a machine learning model with the CK+ dataset [65]. Stöckli et al. [66] performed a validation study of the FACET algorithm within the commercial iMotions software, reporting 96% accuracy on images, with a lower accuracy of 67% for videos. In another validation study of commercial recognition software, called FaceReader, Lewinski et al. [67] reported an accuracy of 85% validating against two datasets, ADFES and WSEFEP [68, 69].

However, these accuracy levels seem to be attained by concentrating on validating against one or two emotion dataset libraries, rather than testing on a larger range. More insight maybe gained from validation studies that review algorithms over either a larger range of libraries, or libraries that are not in popular use. Dupré et al. [70] validated 8 commercially available recognition software, with the highest accuracy found to be quite low at 61.9%. This study included reviewing both Facet and FaceReader, discussed earlier. Facet was reported to have the highest accuracy of 61.9%, with FaceReader at 57.3%. This is far below the high values cited earlier from Stöckli et al. [66] and Lewinski et al. [67]. In Dupré et al. [70], the accuracy of another commercial software called CrowdEmotion, whch focuses of video recognition, had an accuracy of 48.3%.

The size of this image input data would be difficult to transmit over the internet within a game or metaverse platform. An additional difficulty is that historical algorithms perform poorly when used with images of a variety of avatars. This issue was reported by Lou et al. [71] in their research into utilising an existing algorithm to recognise the emotion of

avatars. This is understandable, as historical algorithms are not designed to work with subjects as avatars, that may not exhibit transient features, or even may not be entirely humanoid in appearance.

One recent piece of new research is very similar to that proposed in this paper [72]. Siam et al. [72] created a reduced feature set using similar calculations on landmark data, to that proposed in this paper. It reports to be as accurate as 97% on one benchmark dataset with an overall accuracy of 85% across 3 datasets. The main difference with this paper is that Siam et al. [72] utilises the Google FaceMesh 468 point landmark system [73], whilst this research uses the smaller and more commonly used Dlib 68 point system [74]. Additionally this paper's aims to show accuracy across a large range of datasets, and have a reduced feature set designed to correlate directly to the facial properties of a common metaverse avatar [6].

It is this research's hypothesis that predicting emotion from the facial properties of metaverse avatars, as opposed to directly via human images or webcam, will allow for the above hurdle to be overcome. Additionally this should overcome an additional issue with historical algorithms, that they perform poorly when presented with subjects as avatars [71].

4 | SELECTING A METHOD FOR CAPTURING FACIAL LANDMARK DATA

To obtain facial landmark data the research utilised the inexpensive concept of a live webcam video feed, sometimes referred to as Face Over Internet Protocol (FOIP) [8, 75]. This was due to the technology being widely available and already used, within games and metaverses for visualising avatar facial features [10, 55]. This decision eschewed newer and more expensive technologies, such as HMDs incorporating sensors [12, 76]. Additionally, it was important to select an appropriate method for detecting a face from a video image and then detect facial landmarks. As part of the research, a number of options were investigated, with a final decision made on using Dlib's shape predictor, which can be used in real-time and has been used within games [77, 10]. Dlib is a deeper learning method for detecting landmarks and is available on a wide range of devices [78], using a 68 point model for detecting landmarks [74].

An older method was also considered, the Facemark API available within the popular OpenCV image library, which also allowed for a 68 point trained landmark model to be used [79]. Whilst this solution has been used within online games and metaverses already [80, 55], it was seen to be exhibit issues with subjects showing chin dimpling. Whilst OpenCV Facemark API is slower than Dlib's shape predictor, it does have the advantage of working better with smaller image sizes [77], but with modern webcams this was not considered important.

Other interesting options exist, based on other deeper learning methods. The Cambridge Face Tracker (CFT) generates accurate 68 point facial landmarks in real-time [81]. This deep learning method provides additional information, such as head and eye tilt and rotation [82]. It has become popular through it being used as the landmark functionality behind the OpenFace platform [83]. Whilst investigating CFT for possible use, it was seen to be effective in disregarding chin dimpling, but poor in recognising lip depression. Another deep learning method, based on Tensorflow, has been used to generate accurate landmarks, even when the head is rotated/tilt at extreme angles [84]. As such, this method performs much better than others, when a face is occluded. But the method is slow and thus inappropriate for real-time applications such as games and metaverses.

Discussed earlier, the new Google FaceMesh 468 point landmark system [73] is starting to be used in research, although the researcher has not seen this used commercially as yet. This research should also be able to accomodate FaceMesh, as well as the Dlib 68 point model, given that FaceMesh incorporates the same Dib points.

In summary, in selecting a method for capturing facial landmark data, the research ultimately decided to utilise Dlib, as its general popularity may ensure a better take up of the new method. It was noted to be more accurate than OpenCV and works in real-time, with both OpenCV and Dlib faster than other options [77].

5 | LINKING MODERN AVATAR DATA TO EMOTION

For recognising emotions this research focuses on the Facial Action Coding System (FACS), a comprehensive anatomically based system for describing visually discernible facial movement [85]. Pioneered by Paul Ekman, FACS breaks down facial expressions into individual components of muscle movement, called Action Units (AUs). The prescence of action units in a specific arrangement infers the presence of an emotion from Ekman's emotional model [86]. An advantage found with FACS is that it has been shown to generate consistent results across different ethnicities [87]. As such, the use of the Ekman emotion set, referred to in research as an emotional model [88], has become widespread.

A key driver for the positional hypothesis, is that data held within some current types of modern avatar, do mirror Ekman's action units (see Figure 3). This leads to the question of whether, if this facial data is already being held (to visualise an avatar's face in real-time) and transmitted over the internet, whether that data can be used for predicting emotion.

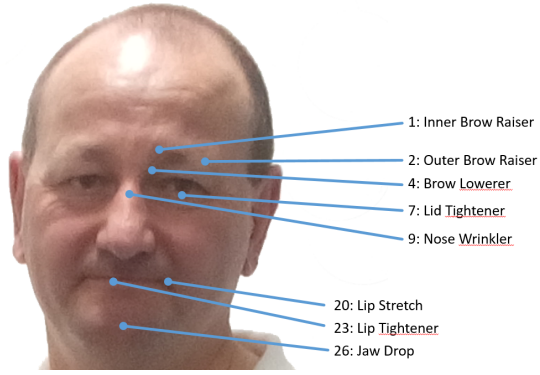

FIGURE 3 Examples of Ekman action units most relevant to facial emotion recognition.

Avatar technology has evolved over the past decade, and many avatars used within games and metaverses, embody data settings to change the facial appearance in finite detail. Different terms are used for these settings, such as "bone positions" or "blend shapes", dependant on the technology platform. The second of these, is the term used for one modern avatar system available for use within the popular Unity3D platform. The UMA Avatar System is a popular system used by Unity3D developers and Figure 4 shows how it links very closely to Ekman action units [6].

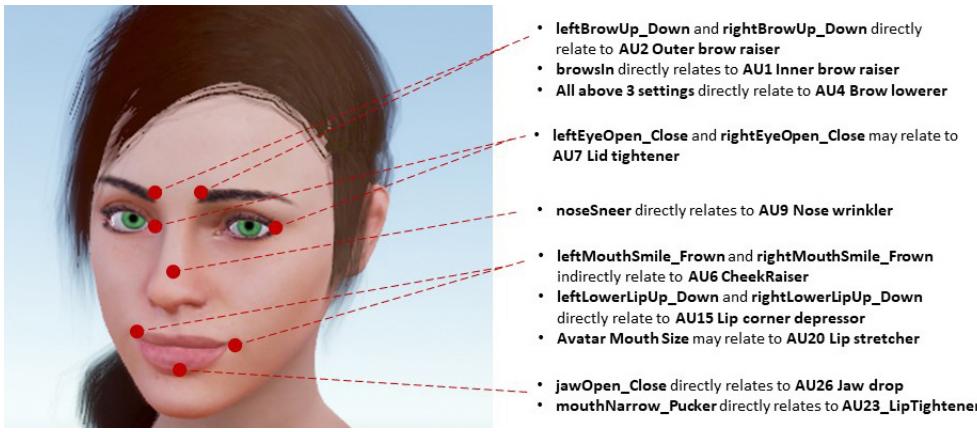

FIGURE 4 Mapping UMA Avatar System to relevant Ekman action units

The concept of mapping action units onto an avatar is not new. The FACSvatar project recognises action units in real-time from a webcam and maps these onto an avatar [89]. But it is somewhat cumbersome and quite slow and has not, as yet, been integrated into a metaverse or game. FACSvatar also does not go as far as recognising and recording emotion. Other research projects have also looked at using emotion within a metaverse, but these do not attempt to predict an emotion from avatar data, but directly from webcam input [90].

6 | PROPOSED REDUCED FEATURE SET AND FEATURE EXTRACTOR

The proposed new feature set will be based around a minimal set of 11 Ekman action units values for facial emotion recognition. These are namely action units 1/2/4/6/7/9/15/20/23/25/26 which relate to values used when controlling modern avatar facial features, see Figure 4. It is proposed to create a feature extractor that will work well within a game or metaverse, as shown in Figure 5. It is of paramount importance to create a feature extractor that calculates values that both work for visualising emotion directly onto the avatar's face, as well as providing input for training/predicting an emotion. This prediction would be achieved by using these values to train a machine learning model, discussed in the following section.

Whilst an existing feature extractor algorithm could be used to generate action unit values, the most consistently accurate of these use transient features [83]. As such, they may not be an appropriate method of generating our proposed reduced set of action unit values. Therefore this research attempts to use a simple set of calculations to make up a feature extraction algorithm. These calculations surface an approximate strength for each action unit. Initial draft calculations are outlined below, utilising the 68 point landmark model [74]. In making the decision to design a simpler set of calculations, research was undertaken into other algorithms that predicted emotion without using transient features. But a number were found to either not to predict in real-time, or be based on having an initial neutral image of a subject as a reference point [91, 92].

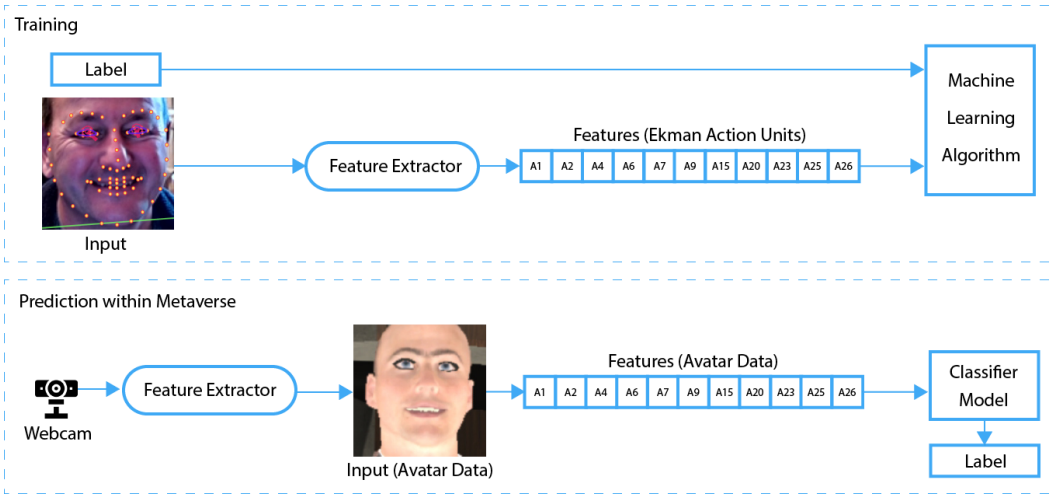

FIGURE 5 Use of the feature extractor in training and then prediction within a metaverse

6.1 | Brow, eye and nasal action units

The main brow action units (1, 2 and 4) are heavily involved in predicting 5 of the 6 Ekman emotions, namely anger, disgust, fear, sadness and surprise. Along with these, action unit 7 is also involved in predicting anger. The calculations are outlined below, and based on landmarks shown in Figure 6. A number of the calculations heavily use the ATAN2 function, which is the arctangent of two numbers [93]. Additionally the Euclidean distance calculation is also used, as below:

$$Distance(d) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

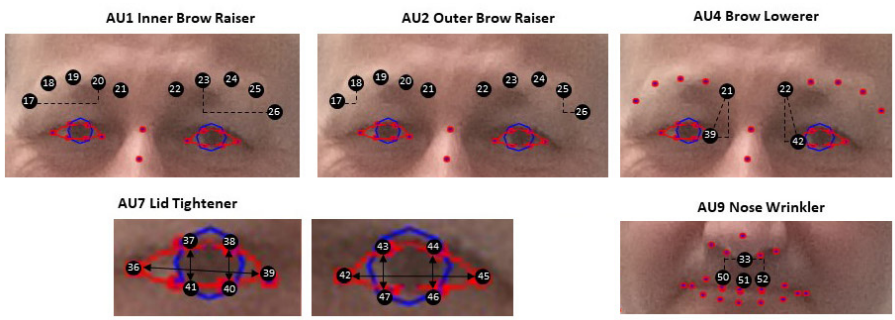

FIGURE 6 Landmark points used in calculating strength of action units 1 to 9.

6.1.1 | Action Unit 1/2/4 - Inner/Outer Brow Raiser and Brow Lowerer

The ATAN2 of points 17-20 and 23-26 are calculated and added together to indicate action unit 1. The ATAN2 of points 17-18 and 25-26 are used together for action unit 2. Finally the ATAN2 of points 21-39 and 22-42 are calculated and added to indicate action unit 4. These can be seen in Figure 6).

$$AU1 = \text{atan2}(y^{17} - y^{20}, x^{20} - x^{17}) + \text{atan2}(y^{26} - y^{23}, x^{26} - x^{23})$$
$$AU2 = \text{atan2}(y^{17} - y^{18}, x^{18} - x^{17}) + \text{atan2}(y^{26} - y^{25}, x^{26} - x^{25})$$
$$AU4 = \text{atan2}(y^{39} - y^{21}, x^{21} - x^{39}) + \text{atan2}(y^{42} - y^{22}, x^{42} - x^{22})$$

6.1.2 | Action Unit 7 - Eyelid Tightener

This action unit, whilst not hard to calculate using the same process as those above, does have issues with being stored within an avatar. Most avatars only store the relative openness of each eye, not whether the certain part of the eyelid is tightening. Instead the average general openness of the eyes is used, based on the commonly found mouth aspect ratio function [94]. As can also be seen in Figure 6, this uses values from the range of eye points 36 to 47.

$$AU7 = \frac{d(p_{48},p_{66}) + d(p_{62},p_{66}) + d(p_{63},p_{65})}{2 \times d(p_{36},p_{45})}$$

6.1.3 | Action Unit 9 - Nose wrinkler

The ATAN2 of points 50-33 and 52-33 are calculated and added together. The higher the resultant value (beyond a minimum threshold) the greater the indication of nose wrinkling. It is accepted that this method will be not as accurate as analysing transient features [95].

$$AU9 = \text{atan2}(y^{50} - y^{33}, x^{33} - x^{50}) + \text{atan2}(y^{52} - y^{33}, x^{52} - x^{33})$$

6.2 | Mouth, cheek and jaw action units

The remaining action units (6, 15, 20, 23 and 26) are involved in predicting all 6 Ekman emotions. Action unit 6 (Cheek Raiser) is the primary indicator for happiness, with action unit 26 (Jaw Drop) being a primary indicator for surprise and fear. Action unit 23 (Lip Tightener) contributes towards an indication of anger, with action unit 20 (Lip Stretcher) contributing to fear. Finally action unit 15 (Lip Corner Depressor) is a primary indicator for disgust, and contributes towards an indication of sadness. The calculations are outlined below, and based on landmarks shown in Figure 7.

6.2.1 | Action Unit 6 - Cheek Raiser

The concept behind the mouth aspect ratio is again used to derive this action unit, using points from 48 to 66. The current calculation produces acceptable values but needs refinement.

$$AU6 = \frac{d(p_{48},p_{66}) + d(p_{66},p_{54})}{d(p_{48},p_{51}) + d(p_{51},p_{54})}$$

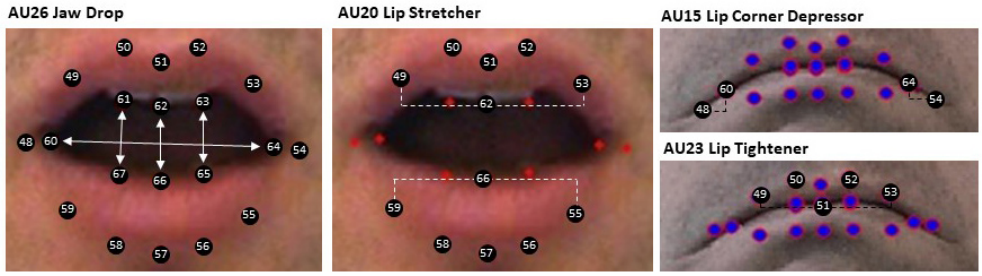

FIGURE 7 Landmark points used in calculating strength of action units 6, 15 and 20 to 26.

6.2.2 | Action Unit 15 - Lip Corner Depressor

The ATAN2 of points 60-48 and 64-54 are calculated and added to create an indication of lip corner depression.

$$AU15 = \text{atan2}(y^{48} - y^{60}, x^{60} - x^{48}) + \text{atan2}(y^{54} - y^{64}, x^{54} - x^{64})$$

6.2.3 | Action Unit 20 - Lip stretcher

The ATAN2 of points 62-49 and 62-53 are calculated and added together, along with the highest value of either the ATAN2 of points 59-66 plus 55-66 or points 57-59 plus 57-55. This creates an indication of lip stretching. The current calculation produces acceptable values but needs refinement.

$$AU20 = \text{atan2}(y^{59} - y^{65}, x^{65} - x^{59}) + \text{atan2}(y^{55} - y^{67}, x^{55} - x^{67}) + \text{atan2}(y^{59} - y^{66}, x^{66} - x^{59}) + \text{atan2}(y^{55} - y^{66}, x^{55} - x^{66}) + \text{atan2}(y^{59} - y^{67}, x^{67} - x^{59}) + \text{atan2}(y^{55} - y^{65}, x^{55} - x^{65})$$

6.2.4 | Action Unit 23 - Lip tightener

In researching lip movement, a decision was taken to only evaluate the upper lip, as this has the largest differentiation in movement [96]. Therefore, the ATAN2 of the upper lip points are calculated and added together to create an indication of lip tightener. The current calculation produces acceptable values but needs refinement.

$$AU23 = \text{atan2}(y^{49} - y^{50}, x^{50} - x^{49}) + \text{atan2}(y^{53} - y^{52}, x^{53} - x^{52}) + \text{atan2}(y^{61} - y^{49}, x^{61} - x^{49}) + \text{atan2}(y^{63} - y^{53}, x^{53} - x^{63}) + \text{atan2}(y^{58} - y^{59}, x^{58} - x^{59}) + \text{atan2}(y^{56} - y^{55}, x^{55} - x^{56}) + \text{atan2}(y^{60} - y^{51}, x^{51} - x^{60}) + \text{atan2}(y^{64} - y^{51}, x^{64} - x^{51}) + \text{atan2}(y^{57} - y^{60}, x^{57} - x^{60}) + \text{atan2}(y^{57} - y^{64}, x^{64} - x^{57}) + \text{atan2}(y^{62} - y^{49}, x^{62} - x^{49}) + \text{atan2}(y^{62} - y^{53}, x^{53} - x^{62}) + \text{atan2}(y^{57} - y^{60}, x^{57} - x^{60}) + \text{atan2}(y^{57} - y^{64}, x^{64} - x^{57})$$

6.2.5 | Action Unit 26 - Jaw drop

This calculation is derived from the commonly found mouth aspect ratio function [94]. The distances between 61-67, 62-66, 63-65, are added together, they are then divided by double the distance 36-45. The value indicates how open the mouth appears to be. At this point an opportunity arose to set calculate action unit 25 (Lips Parted). This is a simple Boolean value, which within other research was pinpointed as an additional indicator of disgust [97]. This led

to improve accuracy during later analysis.

$$AU26 = \frac{d(p_{61},p_{67})+d(p_{62},p_{66})+d(p_{63},p_{65})}{2\times d(p_{36},p_{45})}$$

7 | USING THE FEATURE EXTRACTOR TO TRAIN A MACHINE LEARNING MODEL

In the previous section we outlined the creation of a feature extractor, based around a small set of calculations. This feature extractor can now be used to train a machine learning model. With current algorithms, machine learning methods are used with facial landmarks and transient features, allowing for emotions to be classified and predicted [98]. This research followed the same methodology, initially using a support vector machine (SVM) model with the feature extractor, for predicting emotion.

Training the machine learning model requires requires a dataset library of facial images with each image annotated with the analysed emotion. A number of such image dataset libraries exist that have been extensively used in other research [65, 99, 100, 101, 68, 102, 103]. The feature extractor produces 11 action unit values (discussed in Section 6) for each dataset image, which are then linear interpolated, so that each value is on a scale between 0 and 1, reflecting the strength of each action unit [104]. The final linear interpolated values are then used as the inputs to the proposed machine learning model.

7.1 | Investigating the viability of integrating the model within a metaverse platform

The research initially set out to see if using the proposed design, within a metaverse platform, was viable. To facilitate this, an early draft machine learning model was trained using a moderated subset of the CK+ emotion library [65], and produced quite high early accuracy, as shown in Table 1. With the initial CK+ testing, an emotional accuracy of 73% was recorded. This accuracy was maintained when compared to a sample of Helen images [99]. The accuracy lowered for other data sets, but was still promising.

TABLE 1 Accuracy of initial model, analysed against CK+,Helen,Yale, Jaffe, ADFES, Oulu-CASIA and FEI image libraries. [65, 99, 100, 101, 68, 102, 103]

Accuracy	CK+	Helen	Yale	Jaffe	VisGraf	ADFES	Oulu	FEI
Anger	67	50	n/a	0	31	18	29	n/a
Disgust	75	50	n/a	33	44	56	45	n/a
Fear	60	n/a	n/a	18	28	45	25	n/a
Happiness	100	76	80	41	61	100	75	67
Sadness	27	25	29	3	14	09	01	n/a
Surprise	100	60	60	91	58	86	76	n/a
Neutral	100	87	73	77	81	100	n/a	83
Overall	**76**	**75**	**65**	**36**	**45**	**59**	**42**	**75**
Overall-Neutral	**73**	**73**	**62**	**29**	**39**	**52**	**42**	**67**

As stated earlier the draft model was trained with a moderated subset of the CK+ library, follow on testing with other libraries was performed with their full versions, with moderation only required on the Helen data set [99, 100, 101, 68, 102, 103]. At this early stage, moderation of Helen data set and original CK+ was necessary, as the libraries do contain clear ambiguous entries. This removal of ambiguity, allowed the research to also focus on refining the design of the calculations.

The machine learning model and associated feature extractor were then ready to be integrated. A prototype metaverse was developed, shown in Figure 8, which used an "every other frame" concept for emotion prediction. The metaverse was viewed on a computer setup that was based around an Intel i7-6700K processor, with 16GBRam and an Nvidia GTX 1070 graphics card. This specification cannot be termed modern, given the processor and graphics card are quite dated. From a pure visual standpoint the metaverse was seen to not be affected by the overhead of emotion recognition, when emotion prediction was being used.

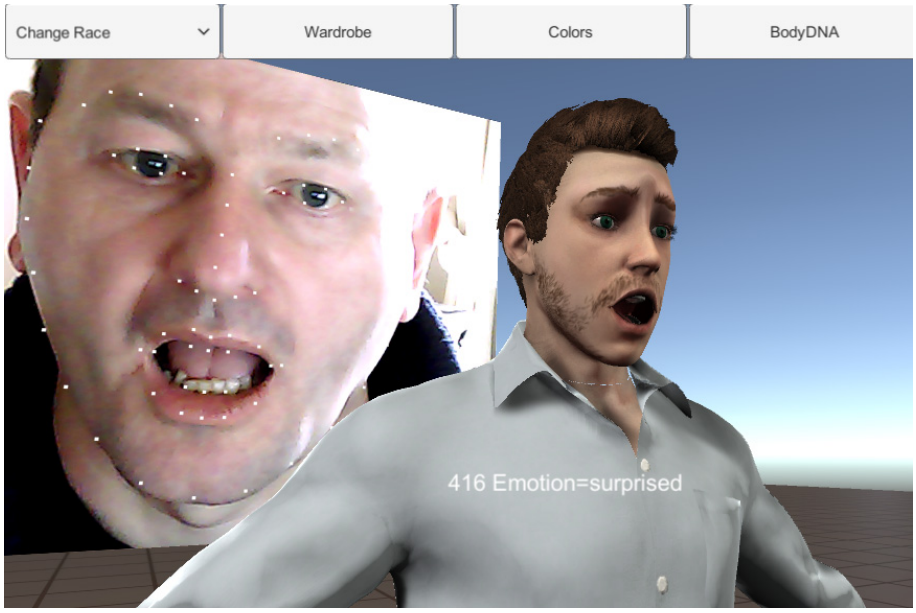

FIGURE 8 Prototype view of facial emotion mapped onto a personalised 3d avatar

The prototype metaverse was developed using the popular Unity 3D development environment [105], which allowed for measuring the time to predict an emotion. The timings were a little disappointing with an average of 0.071 seconds, although it is the image processing that was found to take up over 95% of that time. Taking this into account, the timings show that the additional processing required for emotion prediction is a negligible burden on any game or metaverse that is already utilising facial input for avatars [8].

Whilst this test admittedly raises the question of whether integrating facial emotion prediction into a game or metaverse, where there is currently no facial input, is achievable. But given most modern webcams work at 30FPS for image capture [79], this speed opens the possibility of limiting the processing of facial input to one of every 2 to 4 frames, if one was attempting to reach the high frame rate seen in modern games [106].

7.2 | Improving the model

A final stage involved further improving the model by training against full versions of 6 large libraries: CK+, Jaffe, VisGraf, ADFES, Oulu and FEI. The results are shown in Table 2 with additional testing performed on the smaller Yale library. The average accuracy across the 7 libraries was 77.2%. At this point, a decision was made to move from SVM to a Multilayer Perceptron (MLP) neural network. This new approach was a large factor in the marked increase in accuracy. Another factor in the improvement was that the use of more datasets led to a more balanced training set overall, with there being less concentration of samples in any one emotion.

It is important to note that when analysing the ADFES video library, a threshold accuracy of over half of a videos frames needed to be correctly recognised, before a video was denoted as being correctly analysed as a whole. This threshold may seem low, but with many videos it is found that only a portion of a video shows an actor, or actress, expressing an emotion.

TABLE 2 MLP model trained on a larger set of libraries

Accuracy	CK+ (all)	Yale	Jaffe	VisGraf	ADFES	Oulu	FEI	ADFES(V)
Anger	71	n/a	37	39	68	63	n/a	45
Disgust	76	n/a	45	11	86	45	n/a	86
Fear	88	n/a	41	56	86	57	n/a	55
Happiness	96	80	68	81	100	88	92	100
Sadness	57	20	29	36	36	39	n/a	23
Surprise	100	87	73	81	90	85	n/a	77
Neutral	96	47	87	67	91	n/a	89	91
Overall	**92**	**58**	**54**	**53**	**80**	**63**	**90**	**68**
Overall-Neutral	**85**	**62**	**49**	**50**	**78**	**63**	**92**	**64**

A benefit of using such a large number of dataset libraries is the increase in diversity offered. A number of libraries have a much greater amount of actors, that come from diverse cultures, not represented any single library. Jaffe was created from volunteers who were solely Japanese females, whilst Oulu-CASI contains a large amount of Chinese volunteers (both male and female). Using many dataset libraries can introduce issues though, both Jaffe and Oulu-CASI use a much smaller image size, which may have an affect on accuracy, in comparison to other libraries.

7.3 | Analysing the results

The final accuracy of 77.2% is higher than that found in commercial emotion recognition software from some recent validation studies [70, 66]. It is though admittedly lower than validation studies that purport to achieve over 90% [107, 64]. It must be stressed that validation studies showing very high accuracy levels, generally analyse one or two image dataset libraries, an accuracy this research could have achieved by reporting on the CK+ and FEI. Additionally, the 68% accuracy achieved when analysing the ADFES video library compares favourably with other research studies [70, 108]. This indicates the novel feature extraction algorithm may work as well on videos as it does on images.

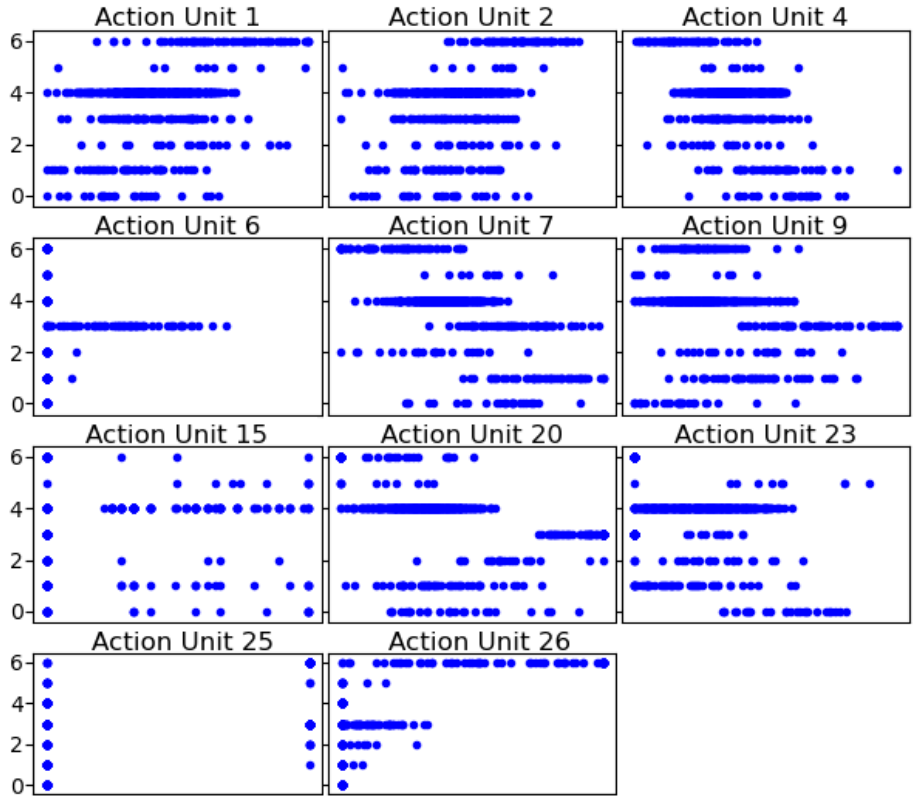

FIGURE 9 Action unit values for the trained model

With regards to poor accuracy amongst certain emotions, it should be noted that in other studies, commercial algorithms have been found to have an uneven distribution of accuracy [66]. To take a closer look at this, distribution charts for each action unit indicator were created early in the research, shown in Figure 9. Whilst reviewing the charts, do note that the larger number of Neutral(4) readings is due to the unbalanced nature of the starting training data, in favour of this setting. One though, can clearly see that the spread of values across action unit 15 does not seem to provide any useful demarcation that could help in identifying an emotion. As this action unit is used for indicating Disgust(1) and Sadness(5), this may be a main reason behind the lower accuracy on these emotions. Additionally, Fear(2), which embodies action Units 1,2,4, 20 and 26, does not have a clear visible demarcation in any unit, thus possibly contributing to its lower accuracy.

7.4 | Future improvements

After analysing the results of the final trained model a number of recommendations can be made, for future improvements. These center on the training data, image capture and calculations, these are:

1. **Algorithm improvement**. The calculations for indicating the presence of some of the action units should be improved. In particular, action unit 15 but also action units 2 and 4 could be re-appraised.

2. **Continue increasing diversity**. Look into expanding the training data with further image dataset libraries, to increase the diversity of the model further.

3. **Improve facial landmark capture**. Whilst the Dlib method of facial landmark capture performed well. There were still some issues around chin dimpling affecting the accuracy of landmarks around the mouth. Looking at implementing ways of mitigating against this, may increase accuracy.

There is scope for improving the feature extraction algorithm along with expanding the training data. Pursuing these may allow the research to approach the accuracy purported by other research.

8 | CONCLUSION AND FUTURE WORK

In this paper we introduced a novel method for predicting facial emotion using a reduced feature set, designed for use with metaverse-based avatars. The machine learning was trained and then quantitatively evaluated using a range of image datasets, producing initial promising results.
Earlier in section 7.5 future improvements were identified and covered revisiting and fine tuning the calculations that make up the feature extraction algorithm, with a view to improving accuracy in an attempt to match the accuracy reported in other research. Accompanying this improvement, gaining access to more culturally diverse image datasets, for both model training and evaluation, will increase the validity of the research.

Future work may also incorporate integrating the method into an example metaverse to showcase its feasibility. Additionally a qualitative review could then be conducted to gain an early insight into the integration of emotion recognition into metaverse platforms.

references

[1] Jones N, The promise and perils of life lived online. Knowable Magazine; 2021. https://knowablemagazine.org/article/technology/2021/the-promise-and-perils-life-lived-online.

[2] Mileva G, 50+ Metaverse Statistics | Market Size Growth (2022). Influencer Marketing; 2022. https://knowablemagazine.org/article/technology/2021/the-promise-and-perils-life-lived-online.

[3] Dionisio JDN, III WGB, Gilbert R. 3D Virtual Worlds and the Metaverse: Current Status and Future Possibilities. ACM Comput Surv 2013 jul;45(3). https://doi-org.libaccess.hud.ac.uk/10.1145/2480741.2480751.

[4] Hughes I. The Metaverse: Is it the Future? ITNOW 2022 02;64(1):22–23. https://doi.org/10.1093/itnow/bwac011.

[5] Sanchez J, A Social History of Virtual Worlds. UploadVR; 2009. https://uploadvr.com/how-sinespace-second-life-linden-lab-vr/.

[6] UMA Steering group, Unity Multipurpose Avatar (UMA); 2021. https://github.com/umasteeringgroup/UMA.

[7] Morris D, Meta Leans In to Tracking Your Emotions in the Metaverse. CoinDesk; 2022. https://www.coindesk.com/layer2/2022/01/19/meta-leans-in-to-tracking-your-emotions-in-the-metaverse/.

[8] Robbaz, Star Citizen - FOIP Face Tracking 2 - Space Delivery; 2018. https://www.youtube.com/watch?v=kKrTKlnKYro.

[9] McWhertor M, EverQuest 2's new facial recognition tech lets you role-play a Froglok like never before. Polygon; 2012. https://www.polygon.com/gaming/2012/6/1/3057820/everquest-2-soemote-facial-recognition.

[10] Enox, Dlib FaceLandmark Detector. Enox; 2016. https://enoxsoftware.com/dlibfacelandmarkdetector/.

[11] Brodsky S, How Face Tracking could Make VR Better. Lifewire; 2021. https://www.lifewire.com/how-face-tracking-could-make-vr-better-5116169.

[12] Oculus, Introducing the Team Behind Half Dome — Facebook Reality. Oculus; 2018. https://www.oculus.com/blog/introducing-the-team-behind-half-dome-facebook-reality-labs-varifocal-prototype/.

[13] HTC VIVE, Bringing your Facial Expressions to Life in VR; 2021. https://www.youtube.com/watch?v=fBLwshNHBRo.

[14] Pita P, Transfer Emotions and Facial Expressions to VR. VR Times; 2017. https://virtualrealitytimes.com/2017/03/13/transfer-emotions-and-facial-expressions-to-vr/.

[15] Wagner J, VRChat User Concurrency Hit Nearly 90,000 Last New Year's Eve! New World Notes; 2022. https://nwn.blogs.com/nwn/2022/01/vrchat-concurrency-2021.html.

[16] Carlos-Roca LR, Torres IH, Tena CF. Facial recognition application for border control. IEEE; 2018. p. 1–7. https://ieeexplore.ieee.org/document/8489113.

[17] Marín-Morales J, Higuera Trujillo JL, Greco A, Guixeres J, Llinares C, Scilingo E, et al. Affective computing in virtual reality: emotion recognition from brain and heartbeat dynamics using wearable sensors. Scientific Reports 2018 09;8. https://www.nature.com/articles/s41598-018-32063-4.

[18] Granato M, Gadia D, Maggiorini D, Ripamonti LA. An empirical study of players' emotions in VR racing games based on a dataset of physiological data. Multimedia tools and applications 2020;79(45-46):33657–33686. https://link.springer.com/article/10.1007/s11042-019-08585-y.

[19] Geher G, Video Games and Emotional States. Psychology Today; 2018. https://www.psychologytoday.com/us/blog/darwins-subterranean-world/201809/video-games-and-emotional-states.

[20] Isbister K. How games move us: emotion by design. First mit press paperback ed. Cambridge, Massachusetts: The MIT Press; 2017.

[21] Bareket R. Playing It Right! Kansas, US: Autism Asperger Publishing Co; 2006.

[22] Beyer J, Gammeltoft L. Autism and play. London: Jessica Kingsley; 1999.

[23] Baron-Cohen S, Golan O, Ashwin E. Can emotion recognition be taught to children with autism spectrum conditions? Philosophical Transactions of the Royal Society B: Biological Sciences 2009;364(1535):3567–3574. https://royalsocietypublishing.org/doi/10.1098/rstb.2009.0191.

[24] Faita C, Brondi R, Tanca C, Carrozzino M, Bergamasco M. Natural User Interface to Assess Social Skills in Autistic Population. In: De Paolis LT, Bourdot P, Mongelli A, editors. Augmented Reality, Virtual Reality, and Computer Graphics Springer International Publishing; 2017. p. 144–154. https://link.springer.com/chapter/10.1007/978-3-319-60928-7_12.

[25] Newbutt N. The development of virtual reality technologies for people on the autism spectrum 2014;p. 230–252. https://www.igi-global.com/gateway/chapter/99571.

[26] Newbutt N, Sung C, Kuo HJ, Leahy MJ, Lin CC, Tong B. Brief Report: A Pilot Study of the Use of a Virtual Reality Headset in Autism Populations. Journal of Autism and Developmental Disorders 2016;46(9):3166–3176. https://link.springer.com/article/10.1007/s10803-016-2830-5.

[27] Wilson J, Simulation and mission rehearsal relies on state-of-the-art computing. Military+Aerospace Electronics; 2020. Retrieved from https://www.militaryaerospace.com/computers/article/14183981/military-simulation-mission-rehearsal-computing-computer.

[28] Alelo, Tactical Iraqi Language & Culture Training System (TILTS). Alelo; 2011. https://www.alelo.com/tilts/.

[29] Meißner M, Pfeiffer J, Peukert C, Dietrich H, Pfeiffer T. How virtual reality affects consumer choice. Journal of business research 2020;117:219–231. https://www.sciencedirect.com/science/article/abs/pii/S0148296320303684.

[30] Lau KW, Lee PY. Shopping in virtual reality: a study on consumers' shopping experience in a stereoscopic virtual reality. Virtual reality : the journal of the Virtual Reality Society 2018;2019;;23(3):255–268. https://link.springer.com/article/10.1007/s10055-018-0362-3.

[31] Heaney D, Facebook Working On Quest 3 & 4, Zuckerberg Wants Face & Eye Tracking. UploadVR; 2021. https://uploadvr.com/zuckerberg-quest-3-4-eye-face-tracking/.

[32] Neurosky, Enhancing AR/VR Devices with EEG and ECG Biosensors. NeuroSky; 2018. http://neurosky.com/2018/01/enhancing-arvr-devices-with-eeg-and-ecg-biosensors/.

[33] Lou J, Wang Y, Nduka C, Hamedi M, Mavridou I, Yu H. Realistic Facial Expression Reconstruction for VR HMD Users. IEEE Transactions on Multimedia 2019 08;PP:1–1. https://ieeexplore.ieee.org/document/8792194.

[34] BBC Click, Nvidia's Video Calling 'Puppets' - BBC Click; 2020. https://www.youtube.com/watch?v=TQy3EU8BCmo.

[35] BBC Click, WFH: The Zoom Boom - BBC Click; 2020. https://www.youtube.com/watch?v=lFt3p76iBBA.

[36] NVIDIA Developer, Inventing Virtual Meetings of Tomorrow with NVIDIA AI Research; 2020. https://www.youtube.com/watch?v=NqmMnjJ6GEg.

[37] Handford F, The Promise of Emotion-Enabled Augmented Reality (AR). Affectiva; 2020. https://blog.affectiva.com/the-promise-of-emotion-enabled-augmented-reality-ar.

[38] GlobalData Thematic Research, Smart Glasses: Technology Trends. Verdict; 2020. https://www.verdict.co.uk/smart-glasses-technology-trends/.

[39] Heiphetz A, Woodill G. Training and collaboration with virtual worlds. McGraw-Hill; 2010.

[40] Steele CB. Building Collaborative Learning Environments: The Effects of Trust and Its Relationship to Learning in the 3-D Virtual Education Environment of Second Life. Steele Shark Press; 2013.

[41] Koles B, Nagy P. Virtual worlds as digital workplaces: Conceptualizing the affordances of virtual worlds to expand the social and professional spheres in organizations. Organizational Psychology Review 2014;4(2):175–195.

[42] Witt KJ, Oliver M, McNichols C. Counseling via Avatar: Professional Practice in Virtual Worlds. International Journal for the Advancement of Counselling 2016;38(3):218–236.

[43] Maddox T, VR Enabling Better Health. Medium; 2018. https://medium.com/edtech-trends/report-vr-enabling-better-health-e5f9037fd0a7.

[44] Madison House Autism Foundation, Video Games and Autism: Helpful or Harmful? Medium; 2017. https://www.madisonhouseautism.org/video-games-and-autism-helpful-or-harmful/.

[45] NHS, Autism and ADHD associated with video game 'addiction'. Nicswell; 2013. https://www.nicswell.co.uk/health-news/autism-and-adhd-associated-with-video-game-addiction.

[46] Satell G, Don't Teach Your Kid to Code. Teach Them to Communicate. Medium; 2018. Retrieved from https://medium.com/s/story/these-are-the-skills-your-kids-will-need-for-the-future-hint-its-not-coding-9b5d47f372f1.

[47] Deming DJ. The growing importance of social skills in the labor market. The quarterly journal of economics 2017;132(4):1593–1640.

[48] Piercy G, Steele Z. The Importance of Social Skills for the Future of Work 2016 01;16:32–42.

[49] Sicile-Kira C, Sicile-Kira J. A Full Life with Autism: From Learning to Forming Relationships to Achieving Independence. Basingstoke, UK: Chipmunkapublishing; 2012.

[50] Fairweather A. How to manage difficult people. London: How To; 2014.

[51] Fairweather A. How to be a motivational manager. Oxford: How To; 2007.

[52] Evenson E. Powerful phrases for dealing with difficult people: Over 325 ready-to-use words and phrases for working with challenging personalities. New York: AMACOM; 2014.

[53] Gibbons S, You and Your Business Have 7 Seconds to Make A First Impression: Here's How to Succeed. Forbes; 2018. https://www.forbes.com/sites/serenitygibbons/2018/06/19/you-have-7-seconds-to-make-a-first-impression-heres-how-to-succeed.

[54] Howarth S. No Matter What. Brentwood, UK: Chipmunkapublishing; 2009.

[55] Enox, OpenCV for Unity. Enox; 2016. https://enoxsoftware.com/opencvforunity/.

[56] Sarwar N, One-Third Of Humanity Has Never Touched The Internet, New Report Says. Screenrant; 2021. https://screenrant.com/what-percentage-of-humans-are-online-internet-access/.

[57] Damm D, Augmented Reality Is Helping End Poverty by Delivering High-Quality Education in Myanmar and Beyond. Singularity Group; 2019. https://www.su.org/blog/how-augmented-reality-is-improving-education-in-myanmar.

[58] 360ed Visioneering Learning, 360ed, First Myanmar AR/VR EduTech Social Enterprise. 360ed Visioneering Learning; 2018. https://youtu.be/s5XUv-DR9nQ.

[59] Haselton T, Facebook just showed us how we'll work in the metaverse — here's what it was like. CNBC; 2021. https://www.cnbc.com/2021/08/19/facebook-introduces-virtual-reality-conference-rooms.html.

[60] Warren T, Microsoft Teams enters the metaverse race with 3D avatars and immersive meetings. The Verge; 2021. https://www.theverge.com/2021/11/2/22758974/microsoft-teams-metaverse-mesh-3d-avatars-meetings-features.

[61] EDPS TechDispatch. Facial Emotion Recognition 2021;https://edps.europa.eu/system/files/2021-05/21-05-26_techdispatch-facial-emotion-recognition_ref_en.pdf.

[62] Tian Y, Kanade T, Cohn JF. Recognizing action units for facial expression analysis. IEEE transactions on pattern analysis and machine intelligence 2001;23(2):97–115. https://ieeexplore.ieee.org/document/908962.

[63] Khadoudja G, Caplier A. Positive and Negative Expressions Classification Using the Belief Theory. International Journal of Tomography & Statistics 2011 07;17. https://hal.archives-ouvertes.fr/hal-00565679.

[64] Liliana DY. Emotion recognition from facial expression using deep convolutional neural network. Journal of Physics: Conference Series 2019 apr;1193:012004. https://doi.org/10.1088/1742-6596/1193/1/012004.

[65] Lucey P, Cohn JF, Kanade T, Saragih J, Ambadar Z, Matthews I. The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression. IEEE; 2010. p. 94–101. https://ieeexplore.ieee.org/document/5543262.

[66] Stöckli S, Schulte-Mecklenbeck M, Borer S, Samson A. Facial expression analysis with AFFDEX and FACET: A validation study. Behavior Research Methods 2017 12;50. https://link.springer.com/article/10.3758/s13428-017-0996-1.

[67] Lewinski P, den Uyl TM, Butler C. Automated Facial Coding: Validation of Basic Emotions and FACS AUs in FaceReader. Journal of neuroscience, psychology, and economics 2014;7(4):227–236. https://doi.org/10.1037/npe0000028.

[68] van der Schalk J, Hawk ST, Fischer AH, Doosje B. Moving Faces, Looking Places: Validation of the Amsterdam Dynamic Facial Expression Set (ADFES). Emotion (Washington, DC) 2011;11(4):907–920.

[69] Olszanowski M, Pochwatko G, Kuklinski K, Scibor-Rylski M, Lewinski P, Ohme RK. Warsaw set of emotional facial expression pictures: a validation study of facial display photographs. Frontiers in psychology 2015;2014;;5:1516–1516. https://doi.org/10.3389/fpsyg.2014.01516.

[70] Dupré D, Krumhuber EG, Küster D, McKeown GJ. A performance comparison of eight commercially available automatic classifiers for facial affect recognition. PLOS ONE 2020 04;15(4):1–17. https://doi.org/10.1371/journal.pone.0231968.

[71] Lou J, Cai X, Dong J, Yu H. Real-time 3D Facial Tracking via Cascaded Compositional Learning. ArXiv 2020;abs/2009.00935. https://arxiv.org/abs/2009.00935.

[72] Siam AI, Soliman NF, Algarni AD, Abd El-Samie FE, Sedik A. Deploying Machine Learning Techniques for Human Emotion Detection. Computational intelligence and neuroscience 2022;2022:8032673–16. https://doi.org/10.1155/2022/8032673.

[73] Grishchenko I, Ablavatski A, Kartynnik Y, Raveendran K, Grundmann M, Attention Mesh: High-fidelity Face Mesh Prediction in Real-time. arXiv; 2020. https://arxiv.org/abs/2006.10962.

[74] Pandey S, Dlib 68 points Face landmark Detection with OpenCV and Python. Study Tonight; 2020. https://www.studytonight.com/post/dlib-68-points-face-landmark-detection-with-opencv-and-python.

[75] Edwards R, Why FOIP will change the face of content creation in games. AltChar; 2018. https://www.altchar.com/game-news/why-foip-will-change-the-face-of-content-creation-adBpU8I7E5A9.

[76] Fove Inc, FOVE Launches v1.0 of Its VR Platform With Major New Eye Tracking Features. Cision PR Newswire; 2020. https://www.prnewswire.com/news-releases/fove-launches-v1-0-of-its-vr-platform-with-major-new-eye-tracking-features-301150620.html.

[77] Gupta V, Face Detection – OpenCV, Dlib and Deep Learning (C++ / Python). Learn OpenCV; 2018. https://learnopencv.com/face-detection-opencv-dlib-and-deep-learning-c-python/.

[78] King DE. Dlib-ml: A Machine Learning Toolkit. Journal of Machine Learning Research 2009;10:1755–1758. https://www.jmlr.org/papers/volume10/king09a/king09a.pdf.

[79] Mallick S, Facemark: Facial Landmark Detection using OpenCV. Learn OpenCV; 2018. Retrieved from https://learnopencv.com/facemark-facial-landmark-detection-using-opencv/.

[80] Singh G, Introduction to Using OpenCV With Unity; 2018. https://www.raywenderlich.com/5475-introduction-to-using-opencv-with-unity.

[81] Thomas P, Baltrusaitis T, Robinson P, Vivian A. The Cambridge Face Tracker: Accurate, Low Cost Measurement of Head Posture Using Computer Vision and Face Recognition Software. Translational Vision Science and Technology 2016 10;5. https://tvst.arvojournals.org/article.aspx?articleid=2565308.

[82] Baltrusaitis T, Robinson P, Morency L. 3D Constrained Local Model for rigid and non-rigid facial tracking. Proceedings / CVPR, IEEE Computer Society Conference on Computer Vision and Pattern Recognition IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2012 06;https://ieeexplore.ieee.org/document/6247980.

[83] Baltrusaitis T, Robinson P, Morency L, OpenFace: An open source facial behavior analysis toolkit. University of Cambridge; 2016. https://www.cl.cam.ac.uk/research/rainbow/projects/openface/wacv2016.pdf.

[84] Agarwal V, Facial Landmark Detection for Occluded Angled Faces. Towards Data Science; 2020. https://towardsdatascience.com/robust-facial-landmarks-for-occluded-angled-faces-925e465cbf2e.

[85] Ekman P, Friesen WV. Manual for the Facial Action Code. Palo Alto, CA: Consulting Psychologist Press; 1978.

[86] Velusamy S, Kannan H, Anand B, Sharma A, Navathe B. A method to infer emotions from facial Action Units. IEEE; 2011. p. 2028 – 2031. https://ieeexplore.ieee.org/document/5946910.

[87] Ekman P, Friesen WV, O'Sullivan M, Chan A, Diacoyanni-Tarlatzis I, Heider K, et al. Universals and cultural differences in the judgments of facial expressions of emotion. Journal of Personality and Social Psychology 1987;53(4):712–717.

[88] Constantine L, Hajj H. A survey of ground-truth in emotion data annotation. IEEE; 2012. p. 697–702.

[89] van der Struijk S, Huang HH, Mirzaei MS, Nishida T. FACSvatar: An Open Source Modular Framework for Real-Time FACS Based Facial Animation. In: Proceedings of the 18th International Conference on Intelligent Virtual Agents New York, NY, USA: Association for Computing Machinery; 2018. p. 159–164. https://doi.org/10.1145/3267851.3267918.

[90] Jackson P, Michon PE, Geslin E, Carignan M, Beaudoin D. EEVEE: the Empathy-Enhancing Virtual Evolving Environment. Frontiers in Neuroscience 2015 02;9. https://www.frontiersin.org/articles/10.3389/fnhum.2015.00112/full.

[91] Ma J, Li X, Ren Y, Yang R, Zhao Q. Landmark-Based Facial Feature Construction and Action Unit Intensity Prediction. Mathematical Problems in Engineering 2021 03;2021:1–12. https://www.hindawi.com/journals/mpe/2021/6623239/.

[92] Perveen N, Mohan C. Configural Representation of Facial Action Units for Spontaneous Facial Expression Recognition in the Wild; 2020. p. 93–102. https://www.researchgate.net/publication/340073457_Configural_Representation_of_Facial_Action_Units_for_Spontaneous_Facial_Expression_Recognition_in_the_Wild.

[93] Medcalc, ATAN2 function. Medcalc; 2020. https://www.medcalc.org/manual/atan2_function.php.

[94] Kir Savaş B, Becerkli Y. Real Time Driver Fatigue Detection Based on SVM Algorithm. IEEE; 2018. p. 1–4. https://ieeexplore.ieee.org/document/8751886.

[95] Thorne JM, Chatting DJ. Prometheus: Facial Modelling, Tracking and Puppetry. In: Hall P, Willis P, editors. Vision, Video, and Graphics (VVG) 2003 The Eurographics Association; 2003. https://diglib.eg.org/handle/10.2312/vvg20031023.

[96] Ozel M, Face the FACS: Lower face cheat sheet; 2020. https://melindaozel.com/lower-face-cheat-sheet/.

[97] Barrett LF, Adolphs R, Marsella S, Martinez AM, Pollak SD. Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements. Psychological science in the public interest 2019;20(1):1–68. https://journals.sagepub.com/doi/10.1177/1529100619832930.

[98] Cosker D, Krumhuber E, Hilton A. A FACS valid 3D dynamic action unit database with applications to 3D dynamic morphable facial modeling. In: 2011 International Conference on Computer Vision IEEE; 2011. p. 2296–2303. https://ieeexplore.ieee.org/document/6126510.

[99] Le V, Brandt J, Lin Z, Bourdev L. Interactive Facial Feature Localization. Springer Link; 2012. https://link.springer.com/chapter/10.1007/978-3-642-33712-3_49.

[100] UCSD Computer Vision, Yale Face Database. University of California San Diego; 2001. http://vision.ucsd.edu/content/yale-face-database.

[101] Lyons M, Kamachi M, Gyobi J. The Japanese Female Facial Expression (JAFFE) Dataset 1998 04;https://zenodo.org/record/3451524#.YBlLUXmnyUk.

[102] Zhao G, Huang X, Taini M, Li S, Pietikäinen M. Facial expression recognition from near-infrared videos. Image Vision Comput 2011 08;29:607–619. https://www.sciencedirect.com/science/article/abs/pii/S0262885611000515.

[103] Thomaz CE, Giraldi GA. A new ranking method for principal components analysis and its application to face image analysis. Image and Vision Computing 2010;28(6):902–913. https://www.sciencedirect.com/science/article/pii/S0262885609002613.

[104] DesLauriers M, Linear Interpolation; 2018. https://mattdesl.svbtle.com/linear-interpolation.

[105] Shah V, Reasons Why Unity3D Is So Much Popular In The Gaming Industry. Medium; 2017. https://medium.com/@vivekshah.P/reasons-why-unity3d-is-so-much-popular-in-the-gaming-industry-705898a2a04.

[106] Stewart S, What Is The Best FPS For Gaming? Gamingscan; 2021. https://www.gamingscan.com/best-fps-gaming/.

[107] Hassouneh A, Mutawa AM, Murugappan M. Development of a Real-Time Emotion Recognition System Using Facial Expressions and EEG based on machine learning and deep neural network methods. Informatics in Medicine Unlocked 2020;20:100372. https://www.sciencedirect.com/science/article/pii/S235291482030201X.

[108] Wenzslaus A. A Comparative Investigation into the use of Cognitive Services API and Image Processing in Python for Emotion Detection 2019 09;http://dx.doi.org/10.13140/RG.2.2.28143.87201/1.