FEATURE ARTICLE

Investigating Emotion Style in Human Faces Using Clustering Methods

Julia Kubiak Melgare, Pontifical Catholic University of Rio Grande do Sul, 90619-900, Porto Alegre, Brazil

Rossana Baptista Queiroz, University of Vale do Rio dos Sinos, 93022-750, São Leopoldo, Brazil

Soraia Raupp Musse, Pontifical Catholic University of Rio Grande do Sul, 90619-900, Porto Alegre, Brazil

For the past decade, performance-driven animation has been a reality in games and movies. While capturing and transferring emotions from human beings to avatars is a reasonably solved problem, it is accepted that humans express themselves in different ways, with personal styles, even when performing the same action. This article proposes a method to extract the style of human beings' facial movement when expressing emotions in posed images. We hypothesize that personal facial styles may be detected by clustering methods based on the similarity of individuals' facial expressions. We use the K-Means and Gaussian mixture model clustering methods to group emotion styles. In addition, extracted styles are considered to generate facial expressions in virtual humans and are tested with users. After an evaluation using both quantitative and qualitative criteria, our results indicate that facial expression styles do exist and can be grouped using quantitative computational methods.

Each person's unique way of acting and moving in real life can influence how virtual characters, who should imitate real people, are animated in virtual environments. This is present in the performance-driven animation (PDA) area when transferring data from real actors to virtual humans.[1]

Meanwhile, regarding human facial expressions, there are several studies that suggest they are universal and a physiological response to our emotions.[2] According to Ekman,[3] there are seven universally recognized emotions (Disgust, Anger, Fear, Sadness, Happiness, Surprise, and Contempt), which directly trigger the same set of facial muscles and produce very similar expressions in any human being. However, more recent studies have since refuted this long-standing universality hypothesis, as can be seen in discussions proposed by Jack *et al.*[4] and Elfenbein and Hess.[5] In this study, we also argue that expressions by different subjects within one emotional category can be different, although recognizable. In order to classify and discretize human facial expressions, Ekman proposed the facial action coding system (FACS),[6] which codifies facial expressions by abstracting the set of muscles that lead facial movements. The system defines a set of action units (AUs), which represent muscle contraction and relaxation. Through the AUs, the system defines how the contraction of each facial muscle (alone and in combination with other muscles) changes the appearance of the face. Although the FACS had been conceived to describe Ekman's observations in psychology, most researchers of computer graphics and computer vision use the FACS system to compose (or decompose) facial expressions.

Furthermore, human motion can be characterized in at least two main aspects: 1) content and 2) style.[7] Content represents a task that is performed with an objective purpose (e.g., walk, jump, and throw a ball), while style describes how each individual performs the defined action, that is, how a person walks, jumps, or throws a ball, and this style may differ from person to person regarding the manner through which the actions are performed. In our study, the concept of emotion style would be a person's unique way of expressing emotions and how it differs from other individuals, even when expressing the same emotion. Much work has been conducted in the perceptual area, aiming to study people's perception of facial expressions.[5] In this work, we are mainly interested in providing a method that allows the extraction of facial emotion styles quantitatively, and validating them

0272-1716 © 2021 IEEE
Digital Object Identifier 10.1109/MCG.2021.3125904
Date of publication 9 November 2021; date of current version 15 July 2022.

qualitatively. Indeed, it is a challenging area because some facial expressions have considerable overlap, according to Kohler *et al.* [8]

Nevertheless, why is studying people's emotional style relevant? This research type is relevant in the context of interfaces for entertainment applications, such as games and gamified products, where the avatar's emotion is important in the included narrative. Indeed, it is important that virtual characters display more variety of expressions to communicate the same emotional state in a way that is more similar to how human beings do, in order to make their nonverbal behavior more natural and believable. In particular, we are not interested in the imitation process, which currently can be covered using PDA, but synthesizing new animations, based on a particular style and expressing more variation, similar to what humans do.

Possible applications are the motion and style transfer to animated characters in games or movies, which can be more realistic and better promote engagement and trust to the users. We can use personalized emotion styles to represent the user through an avatar in virtual environments or explore specific emotion styles to better produce empathetic and convincing virtual humans. Understanding individual emotion styles can also contribute to improving facial expression and emotion detection algorithms, or even in creating applications to help and train people to understand and recognize different styles of expressing emotions.

In this work, we intend to investigate the extraction of facial emotion styles from still pictures. We considered the acted content to be all six basic emotions, according to Ekman's universal emotions,[3] i.e., Happiness, Fear, Disgust, Anger, Surprise, and Sadness; while style states for the variations observed within the expression of these emotions. We use two datasets: first, we experimented with 11 subjects who were asked to express the same emotions but in their own style, not imitating some reference, which defines a posed dataset. We also used the FacesDB dataset[9] (also posed facial expressions) and measured their expressions using FACS. This dataset contains images of 36 individuals demonstrating, among many other expressions, those six basic emotions mentioned previously.

RELATED WORK

Studies that have been made on the subject of motion style in the literature mostly address body motion or facial motion style for expression transferring to 3D characters. In this section, we address some of these works and other facial expression related studies that attempt similar experiments to the ones made in our work.

Kunz and Lautenbacher[10] have done a similar study to the one we propose; however, they attempt to detect individual facial patterns of expressing pain instead of emotions. Their method for achieving this was also similar. First, they assessed facial expressions during pain in two different individuals' samples and then had each of these measured by trained professionals using FACS. After that, they applied both a hierarchical clustering analysis and K-Means clustering using $k = 4$ to group individuals by their manner of expressing pain.

Elfenbein and Hess[5] provided evidence for the presence of cultural differences in the appearance of facial expressions. They evaluate ten different emotions, including the six universal ones. They expected more substantial differences for emotions that are more frequently used as signals for a social audience (Anger, Happiness, and Sadness) and least likely for expressions typically invoked by nonsocial elicitors that have reliable reflex components, such as Disgust and Surprise, or that are very similar in appearance across mammals, such as Fear. Considering the six universal emotions, they found the most significant variations in Sadness, Happiness, and Anger (decreasing order). However, contrary to their expectations, the study found a weak yet significant difference for surprise and no significant variants in Disgust and Fear. Nevertheless, Disgust was not well recognized across both groups and presented significant differences, even in-group analysis, while Happiness was well recognized even with significant cultural variants.

Meanwhile, learning the motion style of people for 3D character animation has turned into an exciting research field for the past few years. Facial motion style is a novel research field in the computer facial animation area. The goal is to learn the motion style of an actor and either synthesize novel motions or transform existing motions as if the original actor performed them. Ma *et al.*[11] introduced a constraint-based Gaussian process model that can effectively learn the editing style from a small set of facial editing pairs, which can be applied to automate the editing of the remaining facial animation frames or transfer editing styles between different animation sequences, being able to reduce the manual efforts necessary in most of the current facial animation editing practices.

Given these related works presented, our main contribution with this work is the exploration of different manners of expressing the six basic emotions through the use of computational clustering methods, as well as the study of how such styles can be applied to the facial expressions of virtual humans in computer animation.

METHOD

This section describes the proposed methodology, which is organized in four sections: 1) we detail the two

FIGURE 1. Graphic representation of the emotions used in the experiment.[6]

datasets used in this work; 2) we describe our method to extract data to compute the facial styles; 3) we propose a manner of clustering individuals, based on their expression style for each emotion, along with the methods used to evaluate such clustering; and lastly, 4) we attempt to represent the different styles found by the most well-evaluated clusterings in a virtual human.

Datasets

First, it is important to present some concepts. In general, facial expression datasets can be created with many different approaches. First, by trained experts, e.g., having Ekman training, where subjects learn how to express emotions with a specific facial expression that follows the definition created by Paul Ekman;[3] second, facial expressions can be posed, if the participants are asked to perform different expressions, but they are not trained, i.e., expressions can be individualized, without training, but they are not spontaneous. Finally, we can have spontaneous expressions in a natural way. This last presents a challenge with respect to creating a dataset because we need the ground-truth regarding participants' emotion.

In this work, we investigate two datasets, both posed. First, we decided to create our own database, which was already used in previous work.[12] Second, we used a dataset from the literature,[9] which claims to be posed and not trained, like ours, but having more subjects. The next sections detail such datasets.

Dataset FACES-Lab

We created our database, from now on called FACES-Lab.[12] We recorded the footage of individuals by asking each of them to express each of the six basic emotions in their way, while sitting in front of a camera, as commonly used in the area. Although there are many datasets in the area, we created our own to assure that the individuals were instructed to act as spontaneously as possible (i.e., not imitating a pattern), even if this dataset is classified as a posed one. Six videos were recorded for each subject (one for each emotion) using a 720p/30 fps camera. The videos begin with the subject in a neutral expression until they are asked to express a certain emotion and hold it for five seconds. Since there were no stimuli for the subjects to express their emotions, Figure 1 was available to those who asked to see an example of what each expression should look like; however, we made it clear that they should not attempt to mimic the expressions in the images and instead try to do it in their own style. The group consisted of 11 subjects, ten males and one female, around the same age group (ranging from 18 to 29 years old). The subjects were lay people, not actors. Figure 2 shows one of our subjects presenting all of the six basic emotions, as an example of our dataset. It is possible to note the difference between performed expressions in Figures 1 and 2, to show that expressions were different from the Ekman reference.[3]

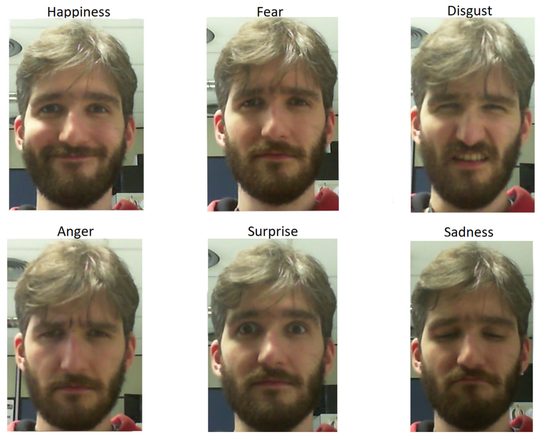


FIGURE 2. One of our subjects expressing all of the six basic emotions.

Dataset FacesDB

In addition to FACES-Lab, we included the FacesDB dataset,[9] which contains images of 36 individuals, composed of 20 men and 16 women, with the majority aged between 20–50 years. The images in this dataset consist of 14 samples for each subject, six of which are the basic emotions we needed, and the remaining samples are facial expressions with mouth and eyes open and closed, along with two samples corresponding to the lateral profiles of the subjects.

Extracting Data From Footage and Defining the Motion Style

To get quantitative data from the recorded footage, we use OpenFace,[a] an open-source facial recognition software that uses deep neural networks that

[a]https://cmusatyalab.github.io/openface/

TABLE 1. All of the 17 facial expressions detected by OpenFace and their respective code (AU) on the FACS.

Facial Expression	Corresponding Action Unit
Inner brow raiser	AU1
Brow raiser	AU2
Brow furrow	AU4
Eye widen	AU5
Cheek raiser	AU6
Lid tighten	AU7
Nose wrinkler	AU9
Upper lip raiser	AU10
Lip corner puller	AU12
Dimpler	AU14
Lip corner depressor	AU15
Chin raiser	AU17
Lip stretcher	AU20
Lip tightener	AU23
Lips part	AU25
Jaw drop	AU26
Blink	AU45

includes, among other features, the detection of 17 AUs[6] in photos and videos. Table 1 lists all of the AUs that the software can detect. In our work, we took all of them into account except for AU45, which represents blinking, an AU that does not affect the perception of any of the six basic emotions used in this work Ekman's EMFACS[13] and is not relevant because we are working with still images.

Using OpenFace, we can obtain the intensity of each AU on each image. When working with FACES-Lab, since we have videos, we extracted the frame with the highest average AU intensity from the ones present in the specific emotion, to be processed in the next step. Also, such frames were visually evaluated in order to verify if they represent the expected emotion. For the FacesDB dataset, this extraction was not necessary since the footage is already in the required image format, and thus, they were processed directly in the software.

In this work, we propose that all subjects express their six basic emotions in six pictures. The pictures are processed using OpenFace and generate six facial emotion styles, one for each of the basic emotions. Of course, there are many ways and intensities that a person can express his/her emotions. In this work, we are working with one picture of each subject for each emotion, but it is plausible to consider that our method can work with many samples for each person.

Let $\vec{ES_i} = \{\vec{H_i}, \vec{F_i}, \vec{D_i}, \vec{A_i}, \vec{Su_i}, \vec{S_i}\}$ be the vector that states the emotion style of individual i, where parameters state for six basic emotions. Each one is a vector that comprises all AUs' intensity values, as extracted from the individual face, using OpenFace. Thus, for each individual from the studied datasets, our method extracts 11 $\vec{ES}$ from FACES-Lab and 36 from FacesDB.

Grouping Individuals by Emotion

Once we have computed $\vec{ES}$ of the 47 subjects in the datasets, we could start the process of clustering them to group emotion styles. To cluster the subjects, we used K-Means[14] and Gaussian mixture models[15] (GMM), with the input data being the AU intensities contained within $\vec{ES}$ of each subject, normalized between 0% and 100%. K-Means was chosen because it was used in other related work,[10,12] and it is also a very known method in scientific literature. GMM was chosen since, as its name implies, each cluster created by the method is modeled after a different Gaussian distribution, an approach that is more flexible for modeling data than other clustering methods, such as K-Means. In comparison to the latter, GMM also considers the variation of data when clustering, instead of taking into account only the mean. Furthermore, we processed the Akaike information criterion (AIC)[16] and Bayesian information criterion (BIC)[17] for each emotion as a way of helping us select the best number of clusters and components to use when applying the K-Means and GMM methods, respectively. The AIC and BIC are both penalized-likelihood criteria, trying to deal with the tradeoff between the goodness of fit and the model's simplicity. Usually, BIC penalizes model complexity more heavily. In general, it might best to use AIC and BIC together in model selection.[18] The results obtained with those information criteria for each method and emotion can be seen in Table 2, and both AIC and BIC agree for all emotions. The number of clusters, however, can be different considering the two clustering algorithms. We can see that when using K-Means, the ideal number for k would be 5 for Happiness, 4 for Fear, 3 for Disgust, and so forth. Meanwhile, with GMM, the recommended number of components n would be 4 for all emotions except Surprise, which is 3.

The Python library Scikit Learn[b] was used for all the analyses cited above. We tested both clustering methods using 1–21 clusters for each emotion and computed the AIC and BIC values.

For evaluating the obtained clusters, we use the Silhouette Score method,[19] which is a measure of how similar an object is to its cluster (cohesion) compared to

[b]https://scikit-learn.org/

TABLE 2. Results obtained upon using the AIC and BIC for recommended number of clusters for each emotion.

Emotion	Clustering method	AIC	BIC
Happiness	K-Means	5	5
Fear	K-Means	4	4
Disgust	K-Means	3	3
Anger	K-Means	3	3
Surprise	K-Means	3	3
Sadness	K-Means	3	3
Happiness	GMM	4	4
Fear	GMM	4	4
Disgust	GMM	4	4
Anger	GMM	4	4
Surprise	GMM	3	3
Sadness	GMM	4	4

other clusters (separation) and results a range from -1 to +1 (a higher value indicates that a particular object matches to its own cluster and poorly matches to other clusters). In addition, we also use the longest common subsequence (LCS) distance[20] to evaluate the clusters by comparing the AUs present in each facial expression of each individual and calculating an average distance between each cluster (including the cluster's distance to itself). Results obtained with both approaches are discussed in the "Experimental Results" section.

Representing Emotion Styles in a Virtual Human

Once the clustering is complete, we use a 3D facial rig based on FACS[c] to represent the styles found among the most well-evaluated clusterings for each emotion, as discussed in the "Experimental Results" section. The rig allows the character's facial expressions to be changed by altering the intensity of each desired AU, which are the same ones used by the OpenFace software (see Table 1). This allows us to transfer the AUs' characteristics of a certain cluster directly to the rig, giving us a graphical generalization of the styles found for each emotion. Since the cluster is made up of faces that have been clustered together by their similarity in some AUs, it is plausible that not all faces are equal. That said, when customizing a virtual character with AU data from a certain cluster, it is necessary to create the generalization of the style of each cluster. In order to achieve this, we define three different operations to be applied on the AU data of the clusters and transferred to the rig, based on data generalization methods applied in cartography.[21] These operations are as follows.

- Smooth—transfer the mean intensity of the most frequent AUs in each cluster (i.e., the ones present in all subjects) to the rig, giving us a visualization of a generalization of the cluster's style.
- Exaggerate—increase the mean AU intensity to be transferred to the 3D model by a certain percentage (e.g., 25%, 50%, and 75%) while not trespassing a value greater than the highest intensity observed for each AU in the cluster.
- Simplify—when calculating the mean intensity of each AU, consider a sample containing only the n most representative subjects of the cluster (i.e., only the n subjects closest to the cluster's centroid), where the value of n can be defined by the user (for our experiments, we used $n = 3$).

To alter the model's facial expressions and render them to the images seen in this article, we use the 3D computer animation software Maya.[d] Table 3 presents the generated facial expressions obtained when applying the operations described previously to the happiness clusters acquired using GMM. The results generated for the remaining five emotions can be seen in Tables 31–35 in the Supplemental material, while our method for evaluating how well these facial expressions represent the clusters is described in the "Public Survey" section.

PUBLIC SURVEY

As a means of evaluating the operations defined in the previous section, we conduct a public survey with lay people in order to determine which of these operations are more appropriate for the generalization and representation of emotion styles in a virtual character, according to human perception.

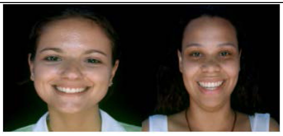
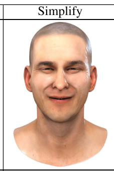
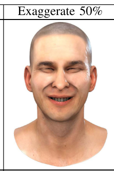
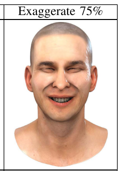
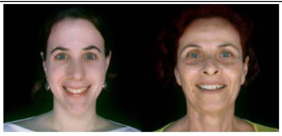
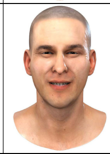
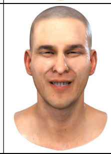
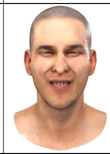
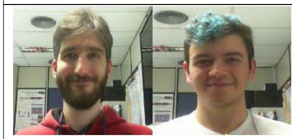
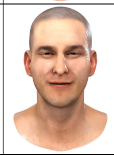
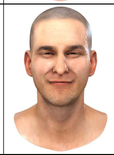
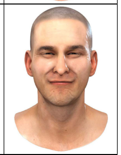
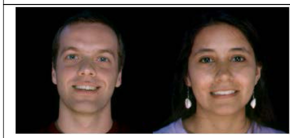
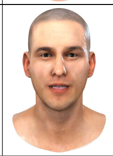
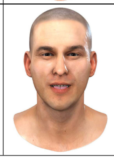
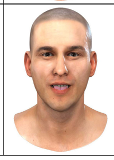
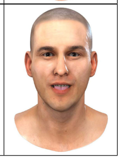
The questionnaire is composed of two parts: the first one contains two control questions while the second part contains a total of 21 questions that evaluate the similarity between the generated facial expressions and the cluster that they attempt to represent, using the most well-evaluated clustering of each emotion.

Table 4 presents the two questions of the survey (first part) in more detail. These questions were presented to determine if participants were capable of discerning facial expressions manifesting different emotions in varying levels of intensity, an ability that would be necessary to give a valid response to the questionnaire. Indeed,

[c]https://www.turbosquid.com/3d-models/3d-model-of-rig-based-facs/1005479

[d]https://autodesk.com/maya

TABLE 3. Example of the results obtained with our defined operations for cluster style representation.

Cluster	Smooth	Simplify	Exaggerate 50%	Exaggerate 75%
 C0				
 C1				
 C2				
 C3				

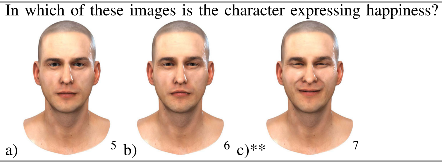
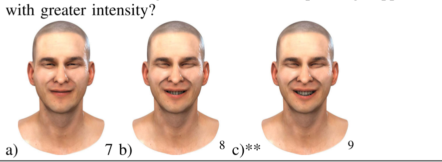
Note: In this table, we show the results obtained with the Happiness GMM cluster AU data.

such questions aim to evaluate our method by showing the same style (cluster) generalization to the participants, with varying characteristics (Simplify) and intensities (Smooth and Exaggerate 50% and 75%). Q1 was formulated to determine whether the participant was able to discern between different emotions, so we ask to indicate which of the presented images was expressing Happiness. For the alternatives, we hand-picked images from our generated dataset that express Anger (Anger K-Means C0 with 75% exaggeration), Sadness (Sadness K-Means C3 Smooth) and Happiness (Happiness GMM C0 Smooth), respectively. These specific expressions were chosen because they were among the most expressive styles found, which made the emotions more clear for the participants. Since we are not evaluating our method in this first section, but rather certifying that the participants are apt for providing a valid response, we were concerned with selecting images that would be appropriate for differentiating the emotions, regardless of the cluster or method used to generate them. For Q2, we wanted to test if the participant was able to differentiate between varying intensities of the same emotion. For the alternatives, we also hand-picked images from our generated dataset, but in this case made sure to choose images from the same style (Happiness GMM C0 Smooth, Happiness GMM C0 with 50% exaggeration, Happiness GMM C0 with 75% exaggeration, with Happiness GMM C0 being the most expressive style found for Happiness—a characteristic that made the variation in intensities between generated images more visible). Thus, for these control questions, we objectively had a correct alternative, which is Q1-c) and Q2-c), as shown in Table 4. It is important to note that the alternatives for Q1 and Q2 were presented to the participants in random order, as to avoid order bias. The subjects who were unable to provide the correct answer for any of these questions were not accounted for in the final results.[e–i]

As for the second part of the survey, an example of the format of the questions can be seen in Figure 3. We include a question for each of the clusterings with the best quantitative evaluation for each emotion (as seen in the "Experimental Results" section), which results in a total of 21 questions. The objective is to find which of the operations better represent a cluster, given pictures of two of the most representative subjects in such cluster for comparison. In the given example, the participants could choose between the Smooth, Simplify, and Exaggerate 50% and 75% options, respectively. For some clusters, however, the results obtained when using Exaggerate 50% and 75% were the same (as can be seen in all K-Means Anger clusters in Table 31 in the Supplemental material, for example). This happens because we do not allow the exaggerated value to surpass the maximum intensity observed for each AU in the cluster in question, as mentioned in representing emotion styles in a virtual human. For these cases where

[e] Anger K-Means C0 with 75% exaggeration.
[f] Sadness K-Means C3 Smooth.
[g] Happiness GMM C0 Smooth.
[h] Happiness GMM C0 with 50% exaggeration.
[i] Happiness GMM C0 with 75% exaggeration.

TABLE 4. Control questions presented in the first section of the public survey.

Q1	In which of these images is the character expressing happiness?  a) [5] b) [6] c)** [7]
Q2	In which of these images is the character expressing happiness with greater intensity?  a) [7] b) [8] c)** [9]

*Note: Alternatives marked with ** are the correct answers. The alternatives for these questions were presented to the participants in random order at the time of the survey. The numbers beside each image are referencing footnotes that show which cluster and method were used to produce the facial expression.*

maximum exaggeration had already been achieved with 50%, we presented only three alternatives (Smooth, Simplify, and Exaggerate 50%), along with "None" and "I don't know." This factor was accounted for when analyzing the survey results, as explained in the "Public Survey Results" section. It is important to note that, for this part of the survey, the alternatives for the questions were not shown in a randomized order, so that the "None" and "I don't know" options would always be listed last, encouraging the participants to look at all available images before choosing one of these latter options. As for the order that the questions were presented, we made sure to alternate between the six emotions, in order to prevent participants from analyzing too many images of the same emotional category at once, which could cause visual fatigue and affect the results (e.g., {Happiness C0, Anger C0, Disgust C0, Fear C0, Surprise C0, Sadness C0, Happiness C1...} instead of {Happiness C0, Happiness C1, Happiness C2, Happiness C3, Anger C0, Anger C1...}).

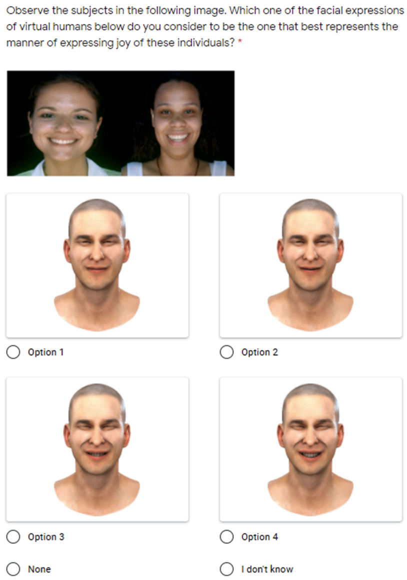


FIGURE 3. Example of the format of the questions presented in the second section of the questionnaire. The participants were asked to compare the available 3D model facial expressions to that of the photos and choose the one that seemed most similar.

EXPERIMENTAL RESULTS

First, we present the clustering results obtained using the K-Means and GMM clustering methods. After that, we present the Silhouette Score and LCS distances evaluation of the results. It is important to notice that the virtual character is not imitating a specific person's style, but representing one of the defined ways to generalize the clusters (e.g., the Smooth operation). Lastly, we present the results obtained in the public survey used to evaluate our method of representing emotion styles with the 3D model.

Clustering Results

This section provides both qualitative and quantitative assessments (regarding AUs) of the clusters proposed in this work. It is crucial to notice that we are not arguing that 3, 4, or 5 is the expected number of emotional styles existent in life. Our contribution is to propose a method of capturing and extracting emotional style in faces, providing a study on the grouping of such styles. For each emotion, we present the differences in the two grouping methods using K-Means and GMM (using the Jaccard index) and access qualitatively the styles obtained with images and generalizations of the facial expressions of that style using a 3D model. Additional information cited in this section is available in the Supplemental material.

Happiness

When using K-Means clustering on the Happiness emotion, the number of clusters was five. The classification results for all 47 subjects can be seen in Table 7 in the Supplemental material. In addition, Figure 4 illustrates the five obtained clusters, where the two selected subjects are the ones closest to their cluster's centroid [on the left, Figure 4(i)]. On the right [Figure 4(ii)], we can see a plot showing the average AU intensity of the most frequent AUs in each cluster (i.e., the ones that were detected in all of the clusters' members), with each of the colors representing one of the 16 total AUs. This same data are also used when synthesizing facial expressions using the

TABLE 5. Average silhouette score for each clustering result obtained on all six emotions for both the K-Means and GMM methods.

Emotion	K-Means	GMM
Happiness	0.1539	0.1002
Fear	0.1674	0.1615
Disgust	0.1551	0.0954
Anger	0.0981	0.0686
Surprise	0.1847	0.1888
Sadness	0.1675	0.0948

Smooth operation, which will be seen throughout the clustering results.

As can be seen in Figure 4(i) and 4(ii), the clustering method seems to group expressions with similar intensities, but not only that, our method groups facial styles. For example, C0 and C1 seem to represent similar intensities, but eyes are more closed in C1 than in C0, as evidenced by the fact that AU7 (Lid Tighten), was present in all of C1's subjects with an average intensity of 44.5% (std=11.1).

More specifically, although all of the clusters except C2 presented the two AUs that define happiness according to the EMFACS[13] (AU6 + AU12), there were many other unique AUs present in each cluster that differentiates them between one another: Clusters C0, C1, and C2 grouped subjects who smile with visible teeth, while clusters C3 and C4 contain subjects who smile only by pulling the corner of their lips. Subjects in cluster C1 have a more intense and expressive smile and also tightened their eyelids, as indicated by the presence of AU7 (Lid Tighten) in their most present AUs. Individuals on cluster C0 had their cheeks and eyes remain neutral while having more action on the brow area, as indicated by the presence of AU1 (Inner Brow Raiser) and AU2 (Brow Raiser) in their most representative AUs. Members of cluster C2 have a less expressive smile and do not have their cheeks raised, having their expressions be seemingly less spontaneous, as indicated by the fact that their most frequent AUs pertain only to the mouth region (AU12—Lip Corner Puller, AU25—Lips Part, AU26—Jaw Drop). Lastly, with clusters C3 and C4, it can be argued that they should be combined into one cluster, as it seems like there is no visual difference between their members, and their set of most frequent AUs are similar.

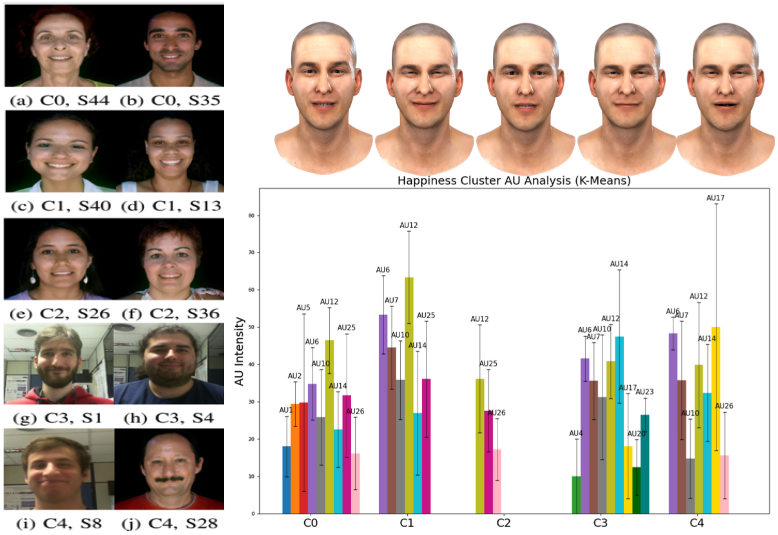
When using GMM, the number of components was 4. Table 8 in the Supplemental material presents the results obtained with this method, while Figure 5 shows a visual representation of the four groups obtained, where the subjects selected are the ones with the highest probability of belonging to their group, alongside a graph of the average intensity of the AUs present in all of the subjects the clusters.

Upon inspecting Figure 5(i) and 5(ii), we can note three clusters (C0, C1, and C3) that have members who smile with visible teeth, and the differences among them remain the same. Cluster C0 contains the subjects who have an expressive smile and tighten their lids (AU7), while C1 has subjects with lower intensities and no AU14 (Dimpler), and C3 has the subjects with less spontaneous expressions again. With the K-Means method, we previously had two groups with subjects who smiled without visible teeth. However, when using GMM, since the number of clusters was 4, the subjects who smile only by pulling the corners of their lips were united in the same cluster C3, which seems more coherent upon a visual inspection.

To compare the results obtained with both clustering methods, we used the Jaccard index to measure the percentage of similarities between the K-Means and GMM clusters. For this emotion, the average similarity between the clusters was 53%. We can note that both clustering techniques were able to categorize similar patterns, since cases where some subjects

TABLE 6. Average intercluster and intracluster LCS distances for each clustering result obtained for all six emotions using both the K-Means and GMM methods.

	K-Means		GMM	
Emotion	Intercluster distance	Intracluster distance	Intercluster distance	Intracluster distance
Happiness	4.7061 (std=0.9133)	2.4747 (std=0.894)	4.5659 (std=0.6837)	3.2044 (std=0.3343)
Fear	5.4252 (std=0.6569)	3.8863 (std=0.5621)	5.4837 (std=0.7657)	3.7943 (std=0.514)
Disgust	4.7706 (std=0.6275)	3.518 (std=0.4562)	4.6163 (std=0.704)	3.3166 (std=0.2729)
Anger	5.2229 (std=0.192)	4.4978 (std=0.4016)	5.2279 (std=0.4123)	4.2066 (std=0.2394)
Surprise	5.1627 (std=0.2541)	3.6648 (std=0.4178)	5.1987 (std=0.2847)	3.6354 (std=0.361)
Sadness	5.4378 (std=0.5105)	4.3352 (std=0.4899)	5.1868 (std=0.4197)	4.1391 (std=0.7612)



(i) The two most representative members of each cluster. For each image from (a) to (j), we specify the cluster **C** and subject id **S**.

(ii) Average intensity of the most frequent AUs in each cluster.

FIGURE 4. Results obtained when clustering the emotion Happiness with K-Means, using $k = 5$. On the left, we show the two most representative members of each cluster (chosen by proximity to centroid), and on the right, we show a graph of the average AU intensity for the most frequent AUs in each cluster (that is, the ones that were detected in all of the clusters' members), with error bars and each color representing one of the 16 AUs. Values obtained from OpenFace are normalized between 0% and 100% over the whole dataset before calculating the average. Our 3D model expressing these AUs in their corresponding average intensity (Smooth operation) can also be seen.

belonged to the same cluster in both methods could be noted: $\{$S2, S10, S11$\}$ in cluster C1 of both K-Means and GMM, $\{$S6, S15, S20, S26$\}$ in cluster C2 of both K-Means and GMM, and $\{$S1, S4, S9$\}$ in cluster C3 of K-Means and GMM, for example. These similarities between the results of both methods were highlighted in bold in Tables 7 and 8 in the Supplemental material. Figure 6 serves as a visual reference for these subjects as well. Also, it is interesting to note that when we decreased the number of clusters from 5 (using K-Means) to 4 (using GMM), K-Means clusters C3 and C4 were almost merged to form cluster C3 in the GMM, with the only exception being that S7 went to another group (GMM cluster C1). A visual reference of S7 can also be seen in Figure 6.

Fear

For the Fear emotion, the number of clusters used was four for both methods. Table 9 shows the results obtained when using K-Means and Table 10 shows results obtained using GMM (both tables can be found in the Supplemental material). The subjects highlighted in bold in both tables are the ones that were grouped in the same clusters by both methods. Upon observing Tables 9 and 10, it can be noted that the resulting clusters are almost the same. We used the Jaccard index to measure the similarity between the groups obtained with both methods, and the results had an average of being 86.75% similar.

In order to present all the four manners of expressing Fear found by the clustering methods, we chose to use the results obtained by the K-Means method as an example in Figure 7, where we show the four clusters obtained as well as the average AU intensity for the most present AUs in each of them. The equivalent plotting for the GMM method can be found in Figure 21 in the Supplemental material and will not be discussed due to its similarity with the K-Means results.

When observing the subjects' pictures and the AU intensity graph, we can make a few statements: All clusters are visually different from each other, each with unique characteristics in the subjects' expressions, as supported by the graph in Figure 7(ii). Subjects in K-Means cluster C0 and GMM cluster C2 were characterized mainly by having their mouths open and eyebrows raised, as can be seen in Figure 7(i)- (a) and (b) and as evidenced by the cluster's most frequent AUs (AU1—Inner Brow Raise, AU5—Eye Widen, and AU26—Jaw Drop). Subjects in K-Means cluster C1 and GMM cluster C1 were similar to C0, with the difference being that they did not open their mouths and raise their brows more, as can be seen in Figure 7(i)-(c) and (d) and as evidenced by their set of most frequent AUs in Figure 7(ii), which contains AU1, AU2 (Brow Raiser), and

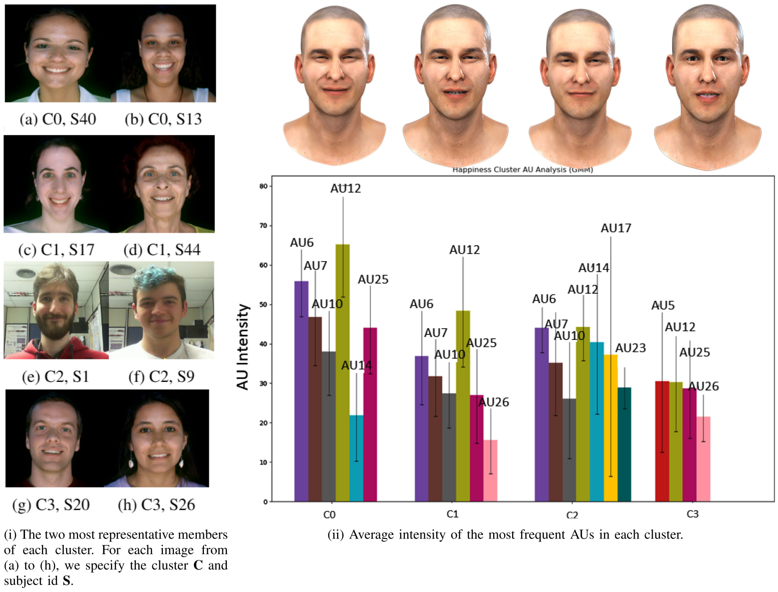


FIGURE 5. Results obtained when clustering the emotion Happiness with GMM, using $n = 4$. On the left, we show the two most representative members of each cluster (chosen by proximity to centroid), and on the right, we show a graph of the average AU intensity for the most frequent AUs in each cluster (that is, the ones that were detected in all of the clusters' members), with error bars and each color representing one of the 16 AUs. Values obtained from OpenFace are normalized between 0% and 100% over the whole dataset before calculating the average. Our 3D model expressing these AUs in their corresponding average intensity (Smooth operation) can also be seen.

AU5, but does not contain AU26. K-Means cluster C2 and GMM cluster C3 were characterized by having subjects with their brows furrowed and lids tightened, while also being the style that has the most number of present AUs, as can be seen in Figure 7(i)-(e) and (f) and as evidenced by the presence of AU4 (Brow Furrow) and AU7 (Lid Tighten) in their most frequent AUs. Lastly, K-Means cluster C3 and GMM cluster C0 had

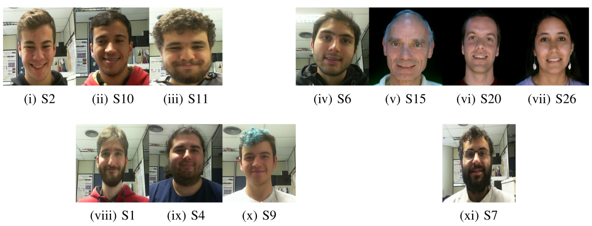


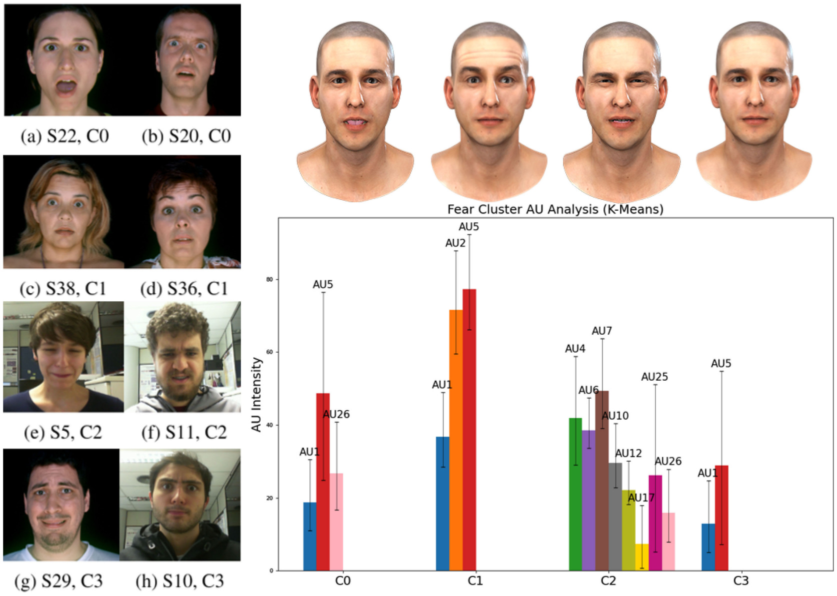
FIGURE 6. Pictures of the subset of subjects [from (i) to (xi): Subjects ID **S** in the pictures] that were grouped together in the same cluster by both methods when expressing Happiness. For example, S2, S10, and S11 were in the same cluster when using K-Means and in the same cluster when using GMM, as were the subsets composed of {S6, S15, S20, S6} and {S1, S4, S9}. S7, meanwhile, was a subject who was moved to another cluster when decreasing the number of clusters from 5 (used in K-Means) to 4 (using GMM), since he was in C4 using K-Means, but moved to C1 when using GMM.

mostly members who either raised their inner brows or furrowed their brows while also widening their eyes, as can be seen in Figure 7(i)-(g) and (h) and as evidenced by their most frequent AUs in Figure 7(ii) (AU1—Inner Brow Raise, and AU5—Eye Widen).

Disgust

For the Disgust emotion, the recommended number of clusters was 3 when using K-Means and 4 for GMM. The complete results for the clustering using K-Means and GMM methods can be seen in Tables 11 and 12 in the Supplemental material. Upon inspecting the tables, one can remark that while there are similarities between the results from both methods, they are not as evident as the previous emotions. When measuring the clusters' similarity with the Jaccard index, the average was 48%. Figures 8 and 9 present the visual results obtained by both methods as well as graphs of each cluster's most representative AUs.

Even though each of the clustering methods used a different number of clusters to group the subjects, we can note three distinct styles of expressing Disgust among them. Examples of these three distinct styles can be seen in cases where some subjects were put in the same clusters by both methods: {S10, S25, S44} in K-Means cluster C0 and GMM cluster C3, {S8, S13, S38} in K-Means cluster C1 and GMM cluster C1, and {S20, S23, 26} in K-Means cluster C2 and GMM cluster



(i) In this example, we show the labels given by the K-Means method. (ii) Average intensity of the most frequent AUs in each cluster. We used the result given by the K-Means method as an example.

FIGURE 7. Pictures of subjects [from (a) to (h)] expressing Fear, who were grouped in the same cluster when using K-Means and GMM. For means of explanation, we chose to use the results obtained using K-Means as an example. On the left, we show the two most representative members of each cluster (chosen by proximity to centroid). On the right, we show a graph of the average AU intensity for the most frequent AUs in each cluster (that is, the ones that were detected in all of the clusters' members), with error bars and each color representing one of the 16 AUs. Values obtained from OpenFace are normalized between 0% and 100% over the whole dataset before calculating the average. Our 3D model expressing these AUs in their corresponding average intensity (Smooth operation) can also be seen.

C2. Figure 10 serves as a visual reference for the cases just described. These groups of subjects were also highlighted in Tables 11 and 12.

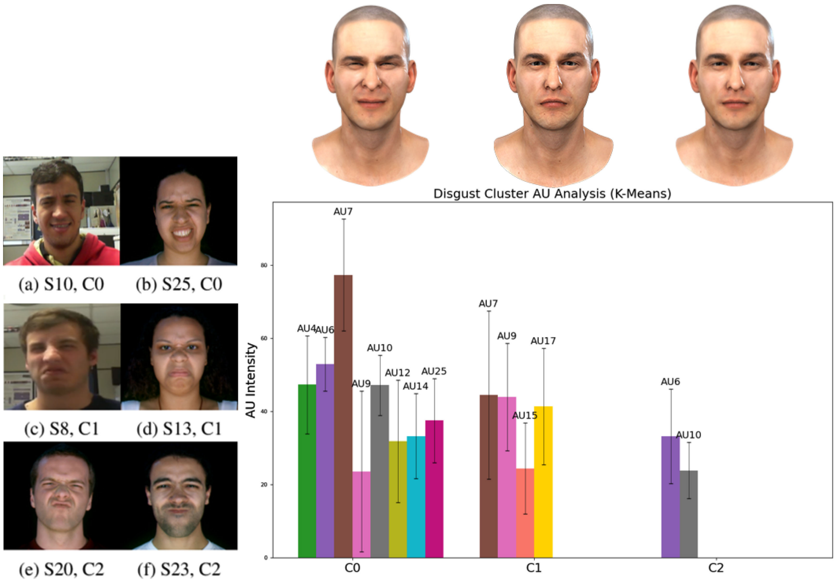
Regarding the three styles of expressing Disgust, some statements can be made: Subjects in K-Means cluster C0 and GMM cluster C3 mainly expressed the brow furrow, lid tighten, and upper lip raise AUs, as evidenced by their presence on these clusters' most frequent AUs in Figures 8(ii) and 9(ii), and as seen in Figure 10(i)–(iii). Subjects in K-Means cluster C1 and GMM cluster C1, unlike the previous groups, did not raise their upper lip, instead showing a combination of AU15 (Lip Corner Depressor) and AU17 (Chin Raiser) [as can be noted in Figures 8(ii), 9(ii), and 10(iv)–(vi)]. Lastly, individuals in K-Means cluster C2 and GMM cluster C2 were the only ones who did not possess AU7 (Lid Tighten) in their most frequent AUs [as can be noted in Figures 8(ii), 9(ii), and 10(vii)–(ix)].

Anger

For the Anger emotion, the number of clusters used with K-Means was three and four for GMM. The results for each method can be seen in Tables 13 and 14 in the Supplemental material. The average of similarities between the clusters obtained with both methods was 55%. The visual results of the clustering methods, as well as the average intensity of the most frequent AUs in each group, can be seen in Figure 11 (K-Means) and Figure 12 (GMM).

Once again, by observing the tables and the results, we were able to determine three different styles of expressing Anger amongst the subjects. These styles can be illustrated by some examples of subjects who were put in the same clusters by both methods: {S40, S41, S46} in K-Means cluster C0 and GMM cluster C0, {S4, S9, S22} in K-Means cluster C1 and GMM cluster C3, and {S14, S15, S37, S42} in K-Means cluster C2 and GMM cluster C1. Figure 13 presents a visual reference for the groups mentioned. They are also highlighted in bold in Tables 13 and 14.

Upon observing the subjects in Figure 13, we can make a few observations regarding the three different styles found: Two out of the three styles of Anger found had the common characteristic of presenting the brow furrow (AU4) AU, which is frequent in the Anger emotion.[13] Subjects in K-Means and GMM cluster C0 were characterized by having visible teeth (AU10—Upper Lip Raiser) and tightening their lids (AU7), as evidenced by their most frequent AUs in



(i) The two most representative members of each cluster. For each image from (a) to (f), we specify the cluster **C** and subject id **S**.

(ii) Average intensity of the most frequent AUs in each cluster.

FIGURE 8. Results obtained when clustering the emotion Disgust with K-Means, using $k = 3$. On the left, we show the two most representative members of each cluster (chosen by proximity to centroid), and on the right, we show a graph of the average AU intensity for the most frequent AUs in each cluster (that is, the ones that were detected in all of the clusters' members), with error bars and each color representing one of the 16 AUs. Values obtained from OpenFace are normalized between 0% and 100% over the whole dataset before calculating the average. Our 3D model expressing these AUs in their corresponding average intensity (Smooth operation) can also be seen.

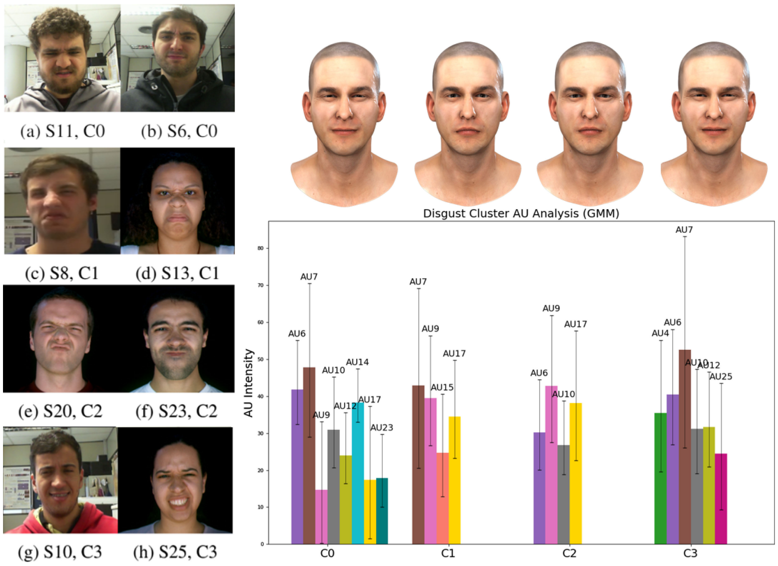
Figures 11(ii) and 12(ii), and exemplified in Figure 13(i)–(iii). Subjects in K-Means cluster C1 and GMM cluster C3, aside from expressing the AU common to all groups (AU4), were less expressive in the lower part of their face, keeping their lips and nose neutral, as evidenced by the fact that their most frequent AUs consist only of AU4 (Brow Furrow) in Figure 11(ii) and AU4 and AU9 (Nose Wrinkler) in Figure 12(ii), and as can be seen in Figure 13(iv)–(vi). Lastly, members of K-Means cluster C2 and GMM cluster C1 presented a combination of lip tightener (AU23) and dimpler (AU14), as evidenced by their most frequent AUs in Figures 11(ii) and 12(ii), and exemplified in Figure 13(vii)–(x).

Surprise

For the Surprise emotion, the number of clusters used was three for both methods. The results obtained can be seen in Tables 15 and 16 in the Supplemental material. Observing the tables, it is possible to see that the results obtained are almost the same, with the only difference being that S26 was placed in cluster C2 when using K-Means and in C0 when using GMM (S26 was highlighted in both tables). When we calculated the Jaccard index between the clusters, the average of similarity was 96%.

With these three different clusters, we were able to identify three styles of expressing Surprise among the subjects. To present them, we chose to use the results obtained by the K-Means method as an example in Figure 14, where we show the three clusters obtained as well as the average AU intensity for the most present AUs in each cluster. The equivalent plotting for the GMM method can be found in Figure 22 in the Supplemental material and will not be discussed due to its similarity with the K-Means results.

Regarding these three styles and S26, we were able to make some observations: Members of K-Means cluster C0 and GMM cluster C1 were overall more expressive, presenting the highest average intensities for inner brow raise (AU2, avg=56.95% and std=24.32) and eye widen (AU5, avg=66.34% and std=22.67), and although AU26 (Jaw Drop) is not present in their most frequent AUs, it was still detected in 93.1% of the subjects (avg=25% and std=9.75), as evidenced in Figure 14 (ii) and exemplified in Figure 14(i)-(a) and (b). Members of K-Means cluster C1 and GMM cluster C0 were similar to the previous group; however, all of them presented AU26, as evidenced by Figure 14(ii) and exemplified in Figure 14(i)-(c) and (d). Lastly, C2 in both K-Means and GMM was characterized by having



(i) The two most representative members of each cluster. For each image from (a) to (h), we specify the cluster **C** and subject id **S**.

(ii) Average intensity of the most frequent AUs in each cluster.

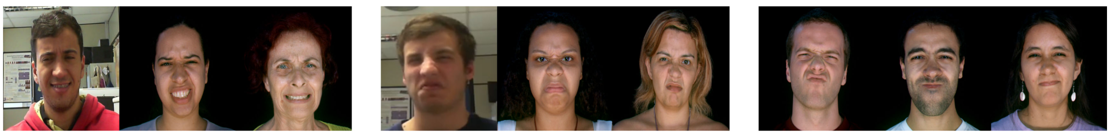
FIGURE 9. Results obtained when clustering the emotion Disgust with GMM, using $n = 4$. On the left, we show the two most representative members of each cluster (chosen by proximity to centroid), and on the right, we show a graph of the average AU intensity for the most frequent AUs in each cluster (that is, the ones that were detected in all of the clusters' members), with error bars and each color representing one of the 16 AUs. Values obtained from OpenFace are normalized between 0% and 100% over the whole dataset before calculating the average. Our 3D model expressing these AUs in their corresponding average intensity (Smooth operation) can also be seen.

subjects who did not open their mouth, expressing Surprise only by raising their brows, as evidenced by Figure 14(ii) and exemplified in Figure 14(i)-(e) and (f).

Sadness

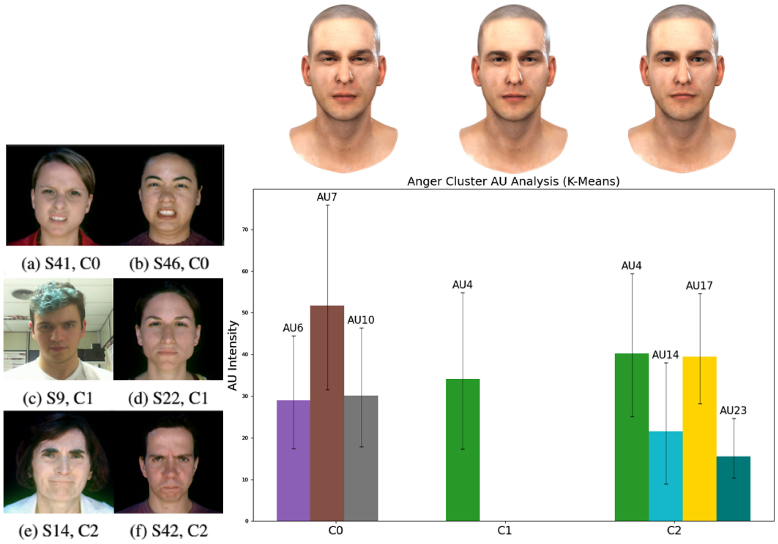
For the Sadness emotion, the number of clusters used was three for K-Means and four for GMM. Tables 17 and 18 in the Supplemental material show the results obtained for each method. When calculating the Jaccard index, the average of similarity was 41%. Even if the clustering methods did not achieve very similar results, we were able to note four different styles of expressing Sadness amongst the subjects. Examples of these four styles can be seen in cases where some subjects were grouped in the same clusters by both methods: {S36, S41, S42} in K-Means cluster C1 and GMM cluster C0, {S2, S3, S5} in K-Means cluster C1 and GMM cluster C1, {S22, S33, S47} in K-Means cluster C1 and GMM cluster C2, and {S21, S25, S45, S46} in K-Means cluster C0 and GMM cluster C3, which were highlighted in bold in Tables 17 and 18. The visual results of the clustering methods, as well as the average intensity of the most frequent AUs in each group, can be seen in Figure 15 (K-Means) and Figure 16 (GMM).

Figure 17 shows a visual representation of the groups mentioned earlier, who also represent the four styles of Sadness found. Regarding these styles, we were able to make some statements: The style found in GMM cluster C0, which is also present in K-Means cluster C1, is characterized by subjects who presented AU17 (Chin Raiser) in combination with AU23 (Lip Tightener), as shown by the graph in Figure 16(ii) and exemplified in Figure 17(i)–(iii). The style that can be



(i) S10 (ii) S25 (iii) S44 (iv) S8 (v) S13 (vi) S38 (vii) S20 (viii) S23 (ix) S26

FIGURE 10. Pictures of some subjects [**S** id—from (i) to (ix)] expressing Disgust, who were grouped together by both the K-Means and GMM methods. These groups also illustrate the three found styles of expressing this emotion.



(i) The two most representative members of each cluster. For each image from (a) to (f), we specify the cluster **C** and subject id **S**.

(ii) Average intensity of the most frequent AUs in each cluster.

FIGURE 11. Results obtained when clustering the emotion Anger with K-Means, using $k = 3$. On the left, we show the two most representative members of each cluster (chosen by proximity to centroid), and on the right, we show a graph of the average AU intensity for the most frequent AUs in each cluster (that is, the ones that were detected in all of the clusters' members), with error bars and each color representing one of the 16 AUs. Values obtained from OpenFace are normalized between 0% and 100% over the whole dataset before calculating the average. Our 3D model expressing these AUs in their corresponding average intensity (Smooth operation) can also be seen.

seen in GMM cluster C1, which is also present in K-Means cluster C1, is one where subjects present AU17 (Chin Raiser) in combination with AU15 (Lip Corner Depressor), as can be seen in Figure 16(ii) and exemplified in Figure 17(iv)–(vi). The style represented by subjects in GMM cluster C2, which is also present in K-Means cluster C1, can be considered the least expressive style, where individuals kept a very neutral facial expression while conveying Sadness, presenting only AU2 (Brow Raise) and AU5 (Eye Widen) in low average intensities (avg=34.26% and std=16.33; avg=26.32% and std=13.99), as supported by Figure 16(ii) and exemplified in Figure 17(vii)–(ix). Lastly, the style found in GMM cluster C3 and K-Means cluster C0 can be described as subjects who presented AU17 (Chin Raiser) in combination with AU15 (Lip Corner Depressor) while also tightening their eyelids (AU7), furrowing their brows, and raising their inner brows(AU4 + AU1). This is supported by the graph in Figure 16(ii) and exemplified in Figure 17(x)–(xiii).

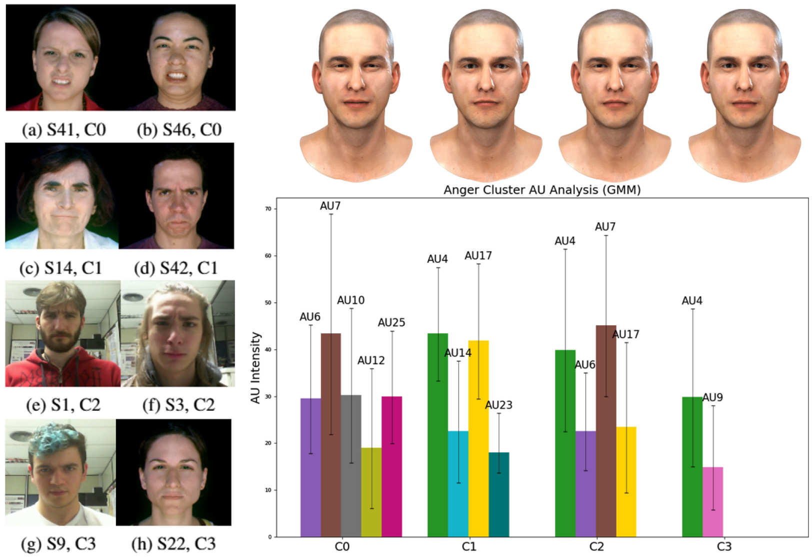
Clustering Evaluation

We proposed to use the Silhouette Score method and the LCS distance between the subjects' facial expressions to evaluate the quality of the clusters obtained. In this section, we present the results obtained with both these evaluations.

Silhouette Score

The average silhouette score obtained by each of our 12 clusterings (six emotions and two clustering methods) can be seen in Table 5. The obtained scores for the generated clusters achieved values > 0 for all emotions, which means that the samples may be close to the decision boundary among the neighbors. In addition, the results also indicate that some emotions may be easier to cluster than others. For example, the highest average scores were obtained with the Surprise emotion (0.1847 for K-Means and 0.1888 for GMM), while the lowest scores were obtained with the Anger emotion (0.0981 for K-Means and 0.0686 for GMM).

Indeed, these values make sense because we are dealing with clusters where the subjects express the same emotion. In the literature, we can find some experiments that support the existence of overlaps when considering independent AUs, even in different emotions. According to Kohler *et al.*,[8] occurrence rates of AUs for facial expressions showed considerable overlap in several studies. For example, they observed that the most frequent AUs in Sadness were also common to other emotions. They also stated that some emotions like Happiness and Sadness seemed to be well recognized with none or only one of the most characteristic AUs being present. This supports the idea that the boundaries of the facial expressions



(i) The two most representative members of each cluster. For each image from (a) to (h), we specify the cluster **C** and subject id **S**.

(ii) Average intensity of the most frequent AUs in each cluster.

FIGURE 12. Results obtained when clustering the emotion Anger with GMM, using $n = 4$. On the left, we show the two most representative members of each cluster (chosen by proximity to centroid), and on the right, we show a graph of the average AU intensity for the most frequent AUs in each cluster (that is, the ones that were detected in all of the clusters' members), with error bars and each color representing one of the 16 AUs. Values obtained from OpenFace are normalized between 0% and 100% over the whole dataset before calculating the average. Our 3D model expressing these AUs in their corresponding average intensity (Smooth operation) can also be seen.

and their intensities can be fuzzy when considering the individual AUs without time coherence and context (for example, photos). Based on this information, we can say that our clustering results seem satisfactory, even if presenting intersecting boundaries.

LCS Distance

The LCS distance[20] was used to measure the similarities between the clusters in regards to AU presence in facial expressions. Using this metric, we can compare how similar two given facial expressions are considering only the AUs detected in them, regardless of intensity. In order to compare the clusters using this metric, we calculated the mean LCS distance between each clusters' subjects, which allowed us to not only measure how similar each of the clusters found for each emotion are, but also how the facial expressions of subjects contained in one of these clusters are similar to each other.

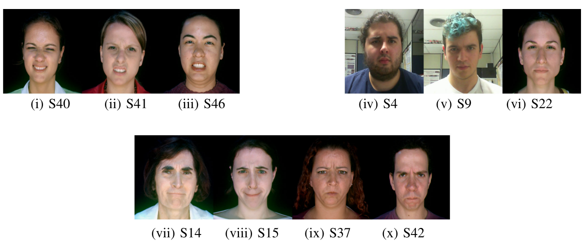
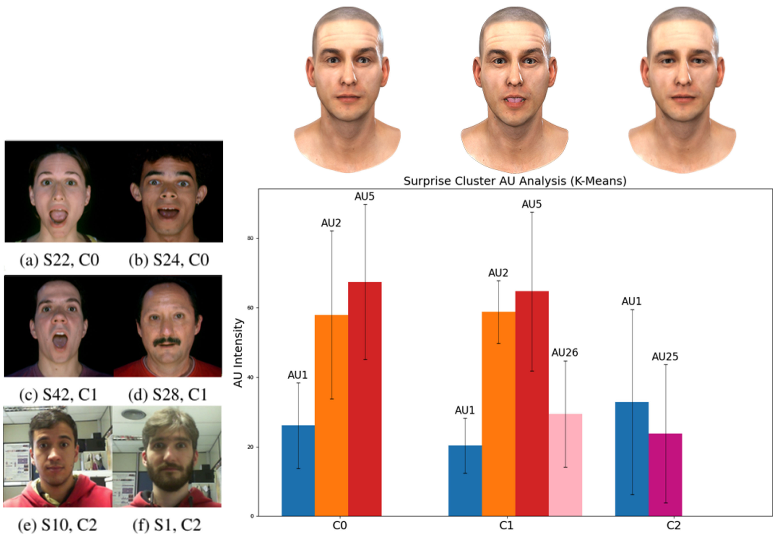


FIGURE 13. Pictures of some subjects [**S** id—from (i) to (x)] expressing Anger, who were grouped together by both the K-Means and GMM methods. These groups also illustrate the three found styles of expressing Anger.

Tables 19–30 in the Supplemental material present the mean LCS distances between the clusters obtained for all emotions and clustering methods. The values are color-mapped by column, meaning that each cluster's lowest distance is highlighted in green, while the highest is in red. We consider the clustering to be acceptable if a cluster's average LCS distance to itself is the lowest when compared to its distance to the other clusters. This was achieved in all clusterings except for K-Means Happiness (see Table 19), since C2's distance to C0 is smaller than C2's distance to itself ($3.5887 < 3.6122$), and for K-Means Sadness (Table 29), since C1's distance to C0 is smaller than C1's distance to itself ($4.8385 < 4.8622$).

Meanwhile, Table 6 shows the average intercluster and intracluster LCS distances for both clustering methods. By observing the results, we can see once again that some emotions have clusterized better than others: Fear and Surprise had one of the highest intercluster distances (5.4252, 5.4837 and 5.1627, 5.1987)



(a) S22, C0 (b) S24, C0
(c) S42, C1 (d) S28, C1
(e) S10, C2 (f) S1, C2

(i) In this example, we show the labels given by the K-Means method. (ii) Average intensity of the most frequent AUs in each cluster. We used the result given by the K-Means method as an example.

FIGURE 14. Pictures of subjects [from (a) to (h)] expressing Surprise, who were grouped in the same cluster when using K-Means and GMM. For means of explanation, we chose to use the results obtained using K-Means as an example. On the left, we show the two most representative members of each cluster (chosen by proximity to centroid). On the right, we show a graph of the average AU intensity for the most frequent AUs in each cluster (that is, the ones that were detected in all of the clusters' members), with error bars and each color representing one of the 16 AUs. Values obtained from OpenFace are normalized between 0% and 100% over the whole dataset before calculating the average. Our 3D model expressing these AUs in their corresponding average intensity (Smooth operation) can also be seen.

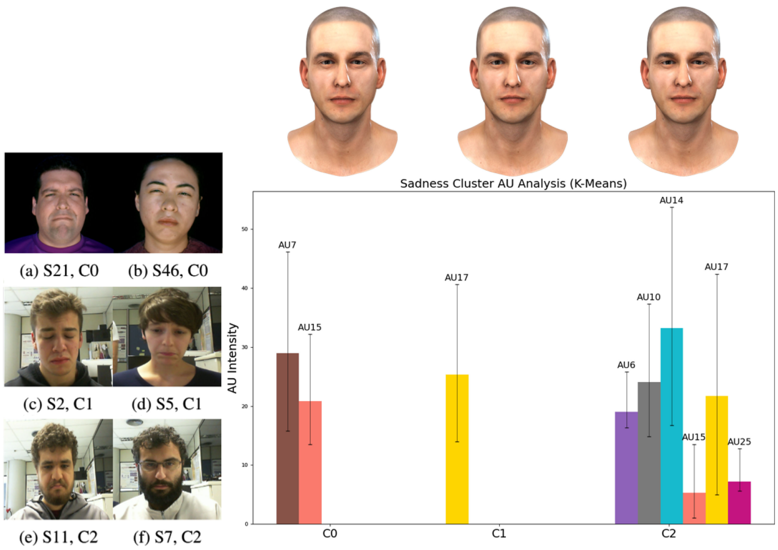
while having one of the lowest intracluster distances (3.8863, 3.7943 and 3.6648, 3.6354). These emotions also had some of the highest silhouette scores (0.1674, 0.1615 and 0.1847, 0.1888). In contrast, Anger and Sadness had the highest intercluster distances (5.2229, 5.2279 and 5.4378, 5.1868) while also having the highest intracluster distances (4.4978, 4.2066 and 4.3352, 4.1391). They also held some of the lowest silhouette scores (0.0981, 0.0686, and 0.0948).

Discussing the Quantitative Results

In this section, we discuss some obtained results concerning the AUs activated by subjects and their intensities, making additional analyses.

First, we analyzed the AU intensities as activated by the subjects. All subjects, on average, present higher intensities when expressing Disgust, while the lower intensities were achieved when expressing Sadness. Figure 19 shows the average values for each emotion, along with the lower and higher values obtained among all subjects. It is interesting to note the biggest variation occurred in Fear, indicating a higher heterogeneity in terms of the subjects' expression intensity. Therefore, S6 is the subject who expressed less AU intensity among all 47 subjects, while S27 is the most expressive (achieved higher intensities). Their six emotions are shown in Figure 18. Indeed, S6 was one of the top five subjects with less expressive intensities in three out of six emotions, while the opposite happens with S27, who has one of the top five highest intensities also in three out of six emotions. In Figure 18, it is interesting to observe the qualitative assessment of those quantitative data, as S6 is visibly less expressive than S27, and it is also easier to recognize the emotions in S27 rather than S6. In addition, S6 is present in the least expressive quartile according to the presence of expected AUs, while S27 is in the group that most activated the expected AUs, according to the literature.[13]

We also investigated whether there exist facial styles related to emotional expressions that can be associated with the specific persons, for instance, if a certain group of subjects in C1 of Happiness GMM stay together in another cluster of other emotions. The conclusion is that no significant correlation between emotional style and subjects in different emotions was found. Let us consider S27 illustrated on the right of Figure 18. S27 is classified as C1 in happiness. If we take all subjects classified as C1 in Happiness GMM (21 subjects), we find an average similarity value of approximately 40% of those subjects in the same clusters as S27, for the remaining emotions. We compute the same analysis for S6 (on the left side of Figure 18) with the subjects on the same cluster (C3 in Happiness) and we found an average similarity



(i) The two most representative members of each cluster. For each image from (a) to (f), we specify the cluster **C** and subject id **S**.

(ii) Average intensity of the most frequent AUs in each cluster.

FIGURE 15. Results obtained when clustering the emotion Sadness with K-Means, using $k = 3$. On the left, we show the two most representative members of each cluster (chosen by proximity to centroid), and on the right, we show a graph of the average AU intensity for the most frequent AUs in each cluster (that is, the ones that were detected in all of the clusters' members), with error bars and each color representing one of the 16 AUs. Values obtained from OpenFace are normalized between 0% and 100% over the whole dataset before calculating the average. Our 3D model expressing these AUs in their corresponding average intensity (Smooth operation) can also be seen.

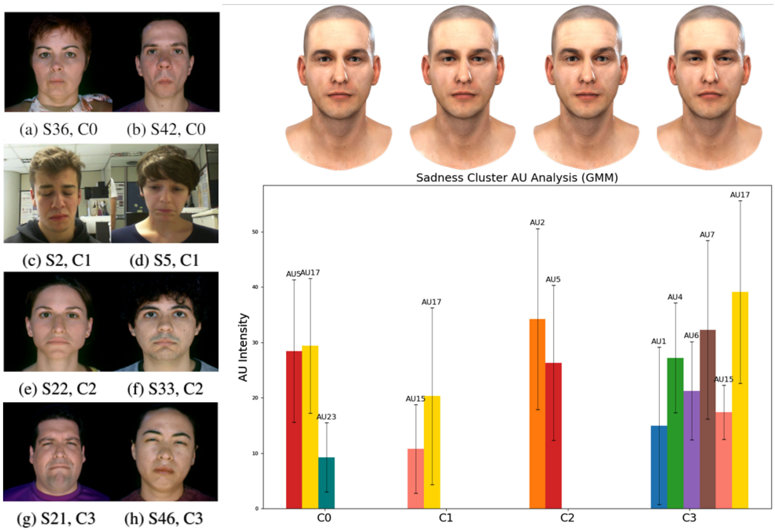
of only 26%. It indicates that a same person can express each emotion differently, e.g., having more AUs' presence in a certain emotion or having a higher intensity in another emotion. For example, S21 in Figure 17(x) is within the less expressive subjects in Happiness, Surprise, and Anger, and is within the more expressive subjects in Fear, Sadness, and Disgust.

Public Survey Results

Our public survey was shared among academic communities in Brazil and received a total of 36 responses. After evaluating which subjects responded to the control questions correctly, 30 valid responses remained. The age of these subjects ranged from 17 to 70 years old, with an average of 27.4 (std=10.83). The gender ratio was 70% male and 30% female. Regarding education, 43.3% of the subjects had completed high school, 30% completed university education, and 26.7% had completed their postgraduate studies. We now present and discuss the obtained results that evaluate whether our method was able to successfully achieve cluster representation, as well as which operation was better suited for this task.

We asked participants to select which facial expression they thought to be most similar to the ones being portrayed by our clustering subjects, as can be seen in Figure 3. In order to process the responses to these questions, we count the number of times each operation (Smooth, Simplify, Exaggerate 50%, and Exaggerate 75%) was chosen by the participants, as well as the number of times that "None" and "I don't know" were selected. Also, for some of these questions, the Exaggerate 50% and 75% of the cluster's style yielded the same result, as both achieved the maximum possible value, and thus, only one of them was presented as an option for the participants. For these questions, we considered that a vote for the Exaggerate option would count as half a vote (0.5) for both 50% and 75%.

The result of this process can be seen in Figure 20, which shows that Exaggerate 50% was the option most voted throughout this section of the questionnaire, being chosen 151.5 times (an average of 7.21 votes per question), followed by "None," which was chosen a total of 149 times (an average of 7.09 votes per question), and Exaggerate 75%, with a score of 137.5 (an average of 6.55 votes per question). The fact that "None" held a high score shows that our method still needs improvement, since this result means that none of the provided options were satisfactory to the participants in a significant amount. Meanwhile, the fact that the Exaggerate 50% method got the most votes while Smooth was the least chosen could indicate that only using the mean AU intensity is not enough to successfully represent an emotion style, but by exaggerating these values, the



(i) The two most representative members of each cluster. For each image from (a) to (h), we specify the cluster **C** and subject id **S**.

(ii) Average intensity of the most frequent AUs in each cluster.

FIGURE 16. Results obtained when clustering the emotion Sadness with GMM, using $n = 4$. On the left, we show the two most representative members of each cluster (chosen by probability of belonging to the cluster), and on the right, we show a graph of the average AU intensity for the most frequent AUs in each cluster (that is, the ones that were detected in all of the clusters' members), with error bars and each color representing one of the 16 AUs. Values obtained from OpenFace are normalized between 0% and 100% over the whole dataset before calculating the average. Our 3D model expressing these AUs in their corresponding average intensity (Smooth operation) can also be seen.

similarity between the 3D model's expression and the cluster's subjects' can become more evident. Also, the fact that Exaggerate 50% was chosen more times than 75% could indicate that, although exaggeration can be more effective to accentuate a cluster's characteristics, too much exaggeration can end up generating a facial expression that is too intense to represent the cluster (i.e., it becomes exaggerated to the point of being unrecognizable). Finally, if we consider the amount of votes that both exaggerations (50% and 75%) attained together, it is easy to see that the exaggeration method was more accepted by the subjects.

FINAL CONSIDERATIONS

This article presents an exploration in search of different styles manifested through the six basic emotions using unsupervised learning techniques. Our results indicate that individual emotion styles do exist, corroborating with Jack *et al.*'s work,[4] and can be captured and analyzed through computational methods, such as clustering algorithms. Regarding the evaluation of these clusters, using LCS distance to measure differences between the subjects' facial expressions, in addition to the average silhouette score of the clusters (which considered only AU intensity data), allowed us to have a richer understanding

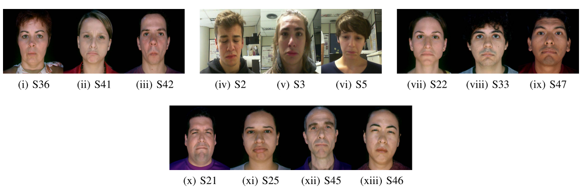


FIGURE 17. Pictures of some subjects [**S** id—from (i) to (xiii)] expressing Sadness, who were grouped together by both the K-Means and GMM methods. These groups also illustrate the four different styles of expressing Sadness found.

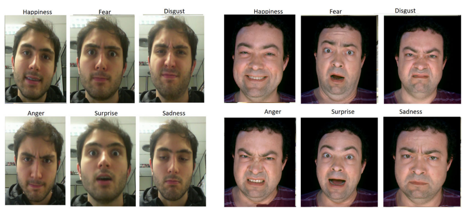


FIGURE 18. Subject on the left is S6 and on the right is S27. S6 is the subject who expressed less AU intensities among all 47 subjects, while S27 is the most expressive (achieved higher intensities).

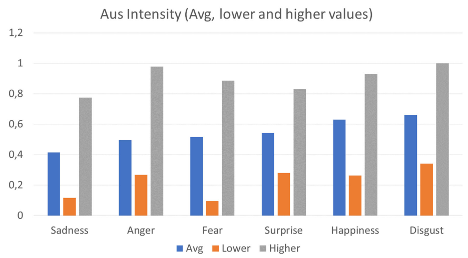


FIGURE 19. Intensity values obtained per emotion. From the left to the right: the blue bar indicates the average intensity per emotion, while the orange and gray indicate lower and higher values of intensity, respectively.

of these clusterings and their quality. Results of both these evaluation methods combined showed that, although we were clearly affected by the intersection between emotions and the subjectivity surrounding human facial expressions, we are still able to obtain clusters that presented visual differences between members of different clusters. The quantitative assessment also supports that these qualitative findings made considering the most frequent AUs in each cluster. Our evaluation also showed that some emotions seem to be more easily clusterable than others. We have discussed how clustering Surprise and Fear yielded better results than Anger and Sadness, for example. Based on the expected and unexpected AUs and their intensities, we evaluate the individual and group facial expressions. Our results indicate that the facial expressions of Happiness are more in agreement with the literature, while Fear is the least in agreement. We also could assess quantitatively and qualitatively the most and least expressive group of people in relation to facial expressions. Finally, we show that people can be very expressive in certain emotions while not expressive in others, showing that, at least in posed pictures, people do not express their emotions homogeneously with regards to different emotions.

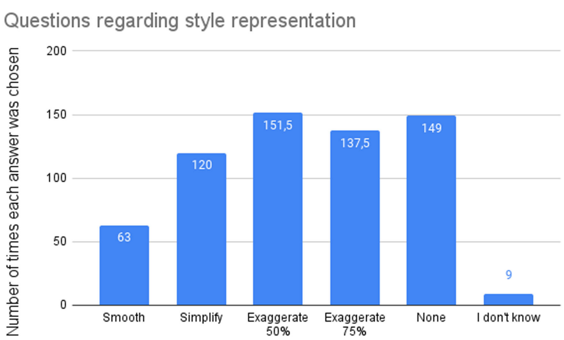


FIGURE 20. Amount of times each operation was chosen by the participants in all 21 questions regarding emotion style representation.

We also propose methods for generating a visually coherent facial expression in virtual humans using generalized cluster data, such as the Smooth operation. As our public survey indicates, there are still more improvements to be made on the coherency between the synthesized facial expressions and the original clusters that they mimic. Although the rigging of the 3D model is complete enough to be able to represent the facial expressions seen in this study, since we want this synthesizing process to be automatic, the stage of translating processed AU data to the model is what needs to be improved. One idea could be, for example, to detect when a smile is meant to be open or closed with the presence of AU25, so that the virtual human can mimic accordingly.

For future work, we would like to further investigate how to distinguish different emotion styles from different emotion intensities. We believe that style encompasses intensity, and thus, we would be able to find different styles within different intensities. In order to do this, we could change our data collection procedure by asking participants to express emotions in varying intensities (e.g., "strong Anger," and "mild Surprise") and then cluster these intensities separately to find styles within them. In addition, we intend to investigate emotion styles in more facial expressions datasets, with respect to different age groups and cultures, applying our method in spontaneous datasets. We could also explore the possibility of using different methods to measure the subjects' AU activity, since although facial recognition software such as the one used in this work can provide fast results, they also are more likely to commit errors when attempting to label an individual's facial expressions. For this, we could have AUs in the footage be labeled by trained professionals, as other work in this area have done.[10] Still regarding our clustering process, we could also investigate explainability techniques in order to better understand which aspects of the facial expressions presented by the subjects were more relevant when forming the clusters.

ACKNOWLEDGMENTS

This work was supported in part by CAPES, in part by CNPQ, and in part by PUCRS. This article has supplementary downloadable material at http://ieeexplore.ieee.org, provided by the authors.

REFERENCES

1. T. Weise, S. Bouaziz, H. Li, and M. Pauly, "Realtime performance-based facial animation," *ACM Trans. Graphics*, vol. 30, no. 4, Jul. 2011, Art. no. 77, , doi: 10.1145/2010324.1964972.

2. C. Darwin and P. Prodger, *The Expression of the Emotions in Man and Animals*. New York, NY, USA: Oxford Univ. Press, 1998.
3. P. Ekman, “An argument for basic emotions,” *Cogn. Emotion*, vol. 6, no. 3/4, pp. 169–200, 1992.
4. R. Jack, O. Garrod, H. Yu, R. Caldara, and P. Schyns, “Facial expressions of emotion are not culturally universal,” *Proc. Nat. Acad. Sci. United States Amer.*, vol. 109, pp. 7241–7244, 2012.
5. H. A. Elfenbein , M. Beaupré, M. Lévesque, and U. Hess, “Toward a dialect theory: Cultural differences in the expression and recognition of posed facial expressions,” *Emotion*, vol. 7, no. 1, pp. 131–146, 2007.
6. P. Ekman and W. Friesen, *Facial Action Coding System: Manual*. Palo Alto, CA, USA: Consulting Psychol. Press, 1978.
7. A. Elgammal and C.-S. Lee, “Separating style and content on a nonlinear manifold,” in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* , 2004, vol. 1, pp. 1–8.
8. C. G. Kohler *et al.*, “Differences in facial expressions of four universal emotions,” *Psychiatry Res.*, vol. 128, no. 3, pp. 235–244, 2004.
9. J. P. Mena-Chalco, R. M. C. Junior, and L. Velho, “Banco de dados de faces 3D-Relatório Técnico,” 2011. [Online]. Available: http://www.vision.ime.usp.br/ jmena/publications/pdf/banco-dados-faces-3d.pdf
10. M. Kunz and S. Lautenbacher, “The faces of pain: A cluster analysis of individual differences in facial activity patterns of pain,” *Eur. J. Pain*, vol. 18, no. 6, pp. 813–823, 2014.
11. X. Ma, B. H. Le, and Z. Deng, “Style learning and transferring for facial animation editing,” in *Proc. ACM SIGGRAPH/Eurographics Symp. Comput. Animation*, 2009, pp. 123–132.
12. J. K. Melgare, S. R. Musse, N. R. Schneider, and R. B. Queiroz, “Investigating emotion style in human faces and avatars,” in *Proc. 18th Braz. Symp. Comput. Games Digit. Entertainment*, 2019, pp. 115–124.
13. W. V. Friesen *et al.*, “EMFACS-7: Emotional facial action coding system,” *Univ. California San Francisco*, vol. 2, no. 36 , 1983, Art. no. 1.
14. A. Likas, N. Vlassis, and J. J. Verbeek, “The global k-means clustering algorithm,” *Pattern Recognit.*, vol. 36, no. 2, pp. 451–461, 2003.
15. D. Reynolds, “Gaussian mixture models,” in *Handbook of Statistics*. Boston, MA, USA: Springer, 2015, pp. 827–832, doi: 10.1007/978-1-4899-7488-4_196.
16. Y. Sakamoto, M. Ishiguro, and G. Kitagawa, *Akaike Information Criterion Statistics*, vol. 81, D. Reidel Ed., Dordrecht, The Netherlands: KTK Sci. Publishers , 1986.
17. G. Schwarz *et al.*, “Estimating the dimension of a model,” *Ann. Statist.*, vol. 6, no. 2, pp. 461–464, 1978.
18. J. J. Dziak, D. L. Coffman, S. T. Lanza, R. Li, and L. S. Jermiin, “Sensitivity and specificity of information criteria,” *Brief. Bioinf.*, vol. 21, pp. 553–565, 2020.
19. P. J. Rousseeuw, “Silhouettes: A graphical aid to the interpretation and validation of cluster analysis,” *J. Comput. Appl. Math.*, vol. 20, pp. 53–65, 1987.
20. S. Velusamy, H. Kannan, B. Anand, A. Sharma, and B. Navathe, “A method to infer emotions from facial action units,” in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2011, pp. 2028–2031.
21. K. S. Shea and R. B. McMaster, “Cartographic generalization in a digital environment: When and how to generalize,” in *Proc. 9th Int. Symp. Comput.-Assist. Cartogr.*, 1989.

JULIA KUBIAK MELGARE is currently an undergraduate computer science student at the Pontifical Catholic University of Rio Grande do Sul, Porto Alegre, Brazil, working on her undergraduate thesis under the supervision of Soraia Musse. Her research interests include perception and detection of human emotions and expressions, computer graphics, and data science. She is the corresponding author of this article. Contact her at julia.melgare@acad.pucrs.br.

ROSSANA BAPTISTA QUEIROZ is currently a Lecturer of computer science and game development at Unisinos, São Leopoldo, Brazil. Her research interests include facial animation, studies about micro- and macro-expressions in virtual characters and user perception. She received the Ph.D. degree in computer science from the Pontifical Catholic University of Rio Grande do Sul, Porto Alegre, Brazil. Contact her at rossanaquerioz@unisinos.br.

SORAIA RAUPP MUSSE is currently an Associate Professor at the School of Technology, Pontifical Catholic University of Rio Grande do Sul, Porto Alegre, Brazil, where she created and still coordinates the Virtual Human Laboratory. Recently, she spent a year at UPENN (USA) working as a visiting scholar with Prof. Norman Badler. She has supervised more than 40 graduate students and postdocs, and authored or coauthored more than 160 papers. She also authored or coauthored three books on crowd simulation with Springer-Verlag. Her research interests include crowd simulation and analysis, facial animation, and integration of computer graphics, pattern recognition and computer vision. She received the Ph.D. degree in computer science from the Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, where she was supervised by Prof. Daniel Thalmann. Contact her at soraia.musse@pucrs.br.