Basic Facial Expressions Analysis on a 3D Model: Based on Action Units and the Nose Tip

Meareg A. Hailemariam(✉)

Addis Ababa University, Addis Ababa, Ethiopia
meareg.abreha@aau.edu.et

Abstract. Facial expressions play a significant role in conveying emotions with a widespread use across diverse cultures and societies globally. In particular, the expressions anger, sadness, fear, disgust, surprise, happiness and also neutral are considered universal. 2D and 3D avatar models are used to simulate facial expressions and have different applications in many domains. In this work, we consider a 3D model with facial expressions as a platform to analyze the basic set of expressions. We considered direction weighted intensity values of the FACS Action Units (i.e., also referred here as shape keys) relative to the nose tip, serving as a reference point, to generate direction weighted score for each target expression. The scores also give numerical validations for the repeated correlations indicated between a specific set of expressions (i.e., anger vs. sadness, and fear vs. disgust) in other research works that focus on developing techniques for facial expressions recognition and classification. In addition, the normal distribution of these seven expressions was depicted and gave a close to bell-curve shape which is an indication of a common phenomenon in nature.

Keywords: 3D facial expressions · Facial expression analysis · FACS · Action Units · Blendshapes

1 Introduction

Facial expressions are important in conveying emotions effectively during communication. According to Mehrabian [1] up to 55% of the message during a face-to-face communication is transferred through facial expressions. Even though human facial expressions may vary across individuals, people or cultures there are seven universally accepted basic facial expressions. These are anger, sadness, fear, disgust, surprise, happiness and also the neutral expression [2]. The Facial Action Coding System (FACS) [3] provides description rule for all the visually detectable changes that can be demonstrated by contraction of the facial muscles; which are also widely known as Action Units (AUs). The FACS separates the facial expression into upper and lower expressions based on the set of Action Units evoked during performance.

N. G. Habtu et al. (Eds.): ICAST 2019, LNICST 308, pp. 462–471, 2020.
https://doi.org/10.1007/978-3-030-43690-2_32

In a daily interaction of human beings, the use of common correlation between facial expressions and emotions is a common approach to recognize emotions. However, other approaches, mainly in used in technical settings, such as the use of speech, electrocardiography (ECG), electromyography (EMG) and electroencephalography (EEG) [17–19] can be also be used to recognize emotions. Adolphs [5] gives the facial spatial points positions for the basic facial expressions and their commonly associated emotions. There is a great deal of interest in affect analysis. Ideally, facial expressions analysis consists of three steps [6]. The first is face detection; after localizing the facial expression image or tracking it in case of sequence of images correctly, then facial or face model features extraction follows. Finally, using the extracted features to come up with a categorization mechanism of facial expressions. A work by Sariyanidi et al. [7] surveys briefly the feature extraction techniques used for facial expression recognition. Among them include Principal Component Analysis (PCA), Discrete Cosine Transformation (DCT) and supervised and unsupervised learning approaches.

The use of 3D avatars with realistic facial expressions has a wide application in domains such as entertainment, education, health and others. The advancement in facial expression research is applied when designing facial expressions of 3D avatars; such as the use of AU of the FACS as a basis to realistic facial animation. But similarly, the advancement of 3D avatars based on facial animations can also contribute back to the facial emotion research, since the computer based tools provide a relatively easy and cheaper setting to perform computation and experiments which might be otherwise, if done with human subjects. The use Action Units (AU) of the FACS to perform facial expression analysis is not new. However, much of the literature focuses either on the recognition of AUs [22,23] or on determining the set of AUs useful for facial expression recognition [24,25] using learning algorithms or distance based approaches. On the other, in this study, we aim to extract a new understanding by analyzing the intensity value of actions units (i.e., here, interchangeably, we may refer to them as shape keys; a term associated with the 3D engine called blender which is used to model the 3D avatar we used in this study) of the seven basic facial expressions on a 3D facial animation model. The set of the basic 3D facial expressions used for our analysis were designed according to the AU rule of FACS for facial expressions and their textual description is similar to the description of the basic facial expressions detailed by Pandzic and Forchheimer [4]. Besides adding a new insight into the categorization and distribution of basic facial expression, this study shows the usefulness of 2D or 3D models in research of facial emotions.

2 Related Work

As facial expressions are basically combinations of movements of different set of facial action units, measuring these distances and intensities of action units is important in quantifying facial expressions. In particular, images of posed facial expressions have been instrumental in the research of facial expressions analysis. Bartlett et al. [8] explore three different techniques of image cues detection for classifying six facial actions that deal with the brows and the eye area. They used a database of over 1,100 sequences containing over 150 distinct actions or action combination images of facial actions. They applied feed-forward network based spatial analysis which also includes PCA generated coefficients as their input. The second is feature based measurement of facial wrinkles and eye openings which is a sum squared of differences of pixel intensities along the chosen segments. Then measurement results were fed to a neural network for target AU classification. Additionally, optic flow fields that estimate direction of motion gradient were also used for action unit classification based on template matching procedure. The three methods gave accuracy results of 88.6%, 57.1% and 84.5% respectively; while their hybrid improved the result to 90.9% which is close to the FAC human expert based classification accuracy of 91.8%.

Local features extraction, via Independent Component Analysis and also by measuring the ratio based geometrical relationship of different parts of the face during the basic facial expressions, was used for facial expression classification on image based dataset [11]. They give as input 9 geometrical ratio features which deal with various eye and mouth parts to a K-NN algorithm; while ANN used 5 local feature vectors of eye and mouth parts that are extracted by ICA. The hybrid of these techniques achieved an accuracy level above 90%.

Another work uses geometric positions and Gabor Wavelets coefficients extracted from facial expression image datasets at fiducial points for expression recognition [12]. Each image is represented by 68 vectors of geometric positions and 612 vectors as Gabor Wavelet coefficients and these extracted features are fed to Multi-layer Perceptron networks. The geometric based recognition gave 73%, the Gabor Wavelet 92.2% and their hybrid 92.3% level of accuracies.

Jaffar and Al Eisa [13] extracted features from facial image datasets using Discrete Cosine Transform (DCT), Haar wavelet transform and Gabor wavelets. The fusion of these features were trained on an SVM and gave a classification of accuracy of 76.11% on MMI, 77.40 on MUG and 95.69 on JAFFE facial expression datasets for the seven basic facial expressions.

A facial expression recognition system [14] based on six 3D distance-vectors that are calculated from 11 feature points located around the eyes, lips and close to ear areas of the face. A Neural Network was trained using the 3D distance vectors as inputs and gave an average accuracy 91.3% for the seven basic facial expressions.

3 Action Unit Intensities and Their Direction Based Analysis of Posed Facial Expressions

The Facial Action Coding System (FACS) is a widely used technique of quantifying facial movement which are the basis of facial expressions. Similarly, here, we rely on the use of Action Units of the FACS for analyzing target facial expressions but with a novel way of computation. Our goal is to get a new insight on understanding the distribution and correlations between the basic facial expressions. Basic facial expressions engage the most prominent AUs which cause detectable facial deformations allowing easy mapping to a subset of the 84 MPEG-4 feature point sets [10]. This relative advantage and their important role among of the list of expressions makes them a target of interest for analysis.

3.1 Dataset and Facial Feature Extraction

We used a 3D model [26] that has all the target 3D facial expressions. The Expressions (i.e., in this text it is also referred to as blendshapes, interchangeably) were encoded following the AU FACS rule for human facial expressions. In this work, since the target of analysis are the posed forms of the basic facial expressions, we captured the peak state (maximum level of intensity) for each target expression. The head pose information is not considered in our case; since we focus on frontal on peak level posed expressions. Each pose of an expression has 45 action units (i.e., which, interchangeably, in this text referred to as shape keys). Most of these shape keys deal with the brows, eyes and lips area of the face; and have the range of their intensity values set to between [0, 1]. The extracted intensity values of all the action units of each captured expression during its peak state are used as a feature vector for each target expression respectively. Obviously, intensity values of all the shape keys for the neutral expression are set to zeros.

3.2 Feature Analysis

Having gathered all the feature vectors, the next step would be the analysis. We consider the nose tip as the reference point on which we base our feature vector transformation. The main reason we chose the nose tip is, it is the part on the face which stays constant (without significant movement) for different set of expressions while other parts may change. It has also a special symmetrical property in respect to the whole face which makes it an important feature on the face [27]. Its symmetry makes it suitable for our direction based analysis of action units on the blendshapes of the basic expressions. In addition, the use of the nose tip in areas such as pose estimation, face alignment has given good results [15]. In our case, considering the nose tip as a reference point, we observe the intensities of shape keys causing facial muscles or bones to move either towards it or away from it. Therefore, we applied the 3D euclidean distance (on x, y, z dimensions) to determine and compare the distance between a target basic expression blendshape's particular shape key's position distance to the nose tip and a similar shape key's distance to the nose tip during the neutral expression. As shown below, we applied the euclidean distance based technique to decide the coefficient sign for a given shape key's intensity from a particular basic expression, to be applied when calculating the its direction weighted score later.

$$(dist\_neutral_{shapekey\_i})^2 = (X_{neutral\_nose\_tip} - X_{neutral\_shapekey\_i})^2 \\ + (Y_{neutral\_nose\_tip} - Y_{neutral\_shapekey\_i})^2 + (Z_{neutral\_nose\_tip} - Z_{neutral\_shapekey\_i})^2$$

$$(dist\_exp\_k_{shapekey_i})^2 = (X_{neutral\_nose\_tip} - X_{exp\_k\_shapekey\_i})^2 + (Y_{neutral\_nose\_tip} - Y_{exp\_k\_shapekey\_i})^2 + (Z_{neutral\_nose\_tip} - Z_{exp\_k\_shapekey\_i})^2$$

We compare these two distances of a given shape key 'i' which were calculated from the neutral expression and the peak state in a target expression's respectively. If a shape key's distance w.r.t to the nose tip during neutral expression is bigger than the corresponding shape key's distance with w.r.t the nose tip during a target expression, then a negative coefficient unit (as it is direction weighted) will negate the intensity score value of that shape key for the target expression else intensity stays positive.

This whole process will be done for each action unit of each target basic expression. The list (represented by 'd') will contain the ± 1s coefficient units vector for the corresponding shape keys' vector of the given basic expression as shown in Eq. 1.

In the cases, where the intensity of a given action unit during a target expression and neutral expression stays the same, the value of the corresponding 'd' will not matter; as equality of value with a neutral expression's shape key implies the zero (default) state.

Finally, we will have a set of feature vectors of intensity values of shape keys of each of the basic expression determined according to their movement direction w.r.t to the reference point. Then, we can further use these transformed features of each target expression to generate a single numeric value that can represent it on a linear line. Similar to techniques in image based difference calculated via difference which approximates derivatives, we apply direction based difference estimator in reference to the origin point.

Equation 1 calculates the weighted score of each target expression using the intensity values of its shape keys and their corresponding multiplier signs.

$$Expr_{weighted} = \frac{1}{N}\sum_{i=1}^{N} Shapekey_i Intensity * d \tag{1}$$

In Eq. 1, the intensity of each shapekey 'i' will be multiplied by its corresponding direction coefficient 'd' which can be either + or -1. The total number of shape keys is represented by N, which in this case are 45. The final iterated summation of all the shape keys, averaged by the total number of shape keys will generate a single direction-weighted numeric score with a value between $[-1, 1]$ for each target expression,

Table 1 shows a list of the average direction weighted intensities of the target expressions calculated using Eq. 1. The expression anger scored the highest based on total average direction weighted intensity value of its AUs heading towards the reference/origin point, while the expression fear on the other end scored highest for its direction weighted average intensity of its AUs' moving away from the origin point.

Table 1. List of average direction weighted intensities of action units of the basic expressions on the 3D model

Expression name	Average direction weighted expression value
Anger	−0.1026
Sadness	−0.0415
Neutral	0.0000
Happiness	0.0304
Surprise	0.0590
Disgust	0.0733
Fear	0.1212

Therefore, according to our results, the new sequence in reference to the origin point, starting from the expression with the highest direction weighted average heading towards the nose tip to the expression with the highest weighted average set of AUs moving away from it, would be anger, sadness, neutral, happiness, surprise, disgust and fear.

3.3 Probability Density Estimation of the Basic Facial Expressions

Measuring the normal probability density [20] of these basic set of expressions can be useful in understanding the distribution patten of the basic facial expressions and offer further insight on human facial expressions in general. Figure 1 depicts the probability density of the normal distribution for the basic facial expressions, including the neutral expression, based on the numeric value results from Table 1.

The low number of the total target expressions (which is seven) we considered literally make it impossible to make a proper comparison of the distribution result generated with the 68%–95%–99.7% rule of a perfectly normal distribution. However, from the result depicted, it is still possible evaluate the symmetry; and also quantitatively show that almost the overall probability mass falls within the 3 units of deviation only with anger's score lying beyond that range with an excess value around 0.02 which can be considered negligible. The mean value calculated is 0.011 with a deviation of 0.29.

The generated distribution is symmetric; it has a close to bell curve shape which indicates a distribution pattern for the basic set of facial expressions that is common in nature. This shows the significant role of the basic set of facial expressions in giving, potentially, a deeper insight into understanding the human emotions and their distribution patterns. The probability densities of the normal distributions for the six basic facial expressions based on the intensity distribution of their shape keys are shown in Figs. 2, 3 and 4 (except for the neutral expression; as all of its AU's intensity values stay at zero). These visualizations

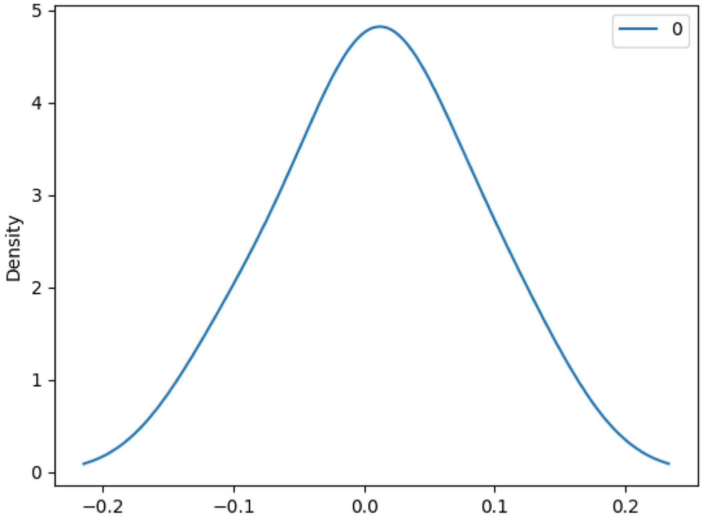

Fig. 1. Normal probability density visualization of the basic facial expressions.

depict the distributions of intensity values of all the shape keys for each of the basic expression. While, for the expressions angry and sad, it can be observed that there is a slight tendency for majority of the shape keys to be less than zero; it is on the contrary for the rest of the expressions, in particular, more visible for the disgust and fear expressions (Fig. 4).

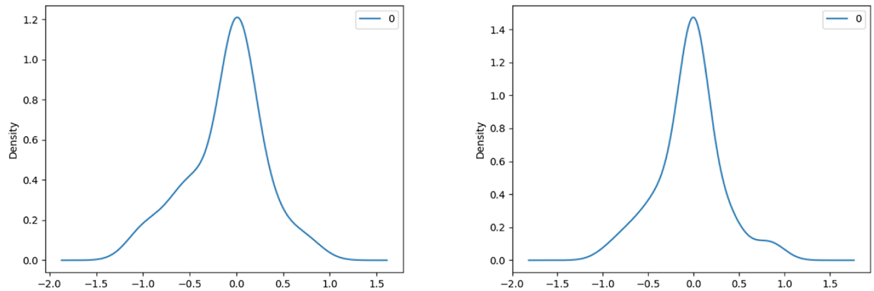

Fig. 2. Density distributions of shape key intensity values for angry and sad expressions depicted on the left and right side respectively.

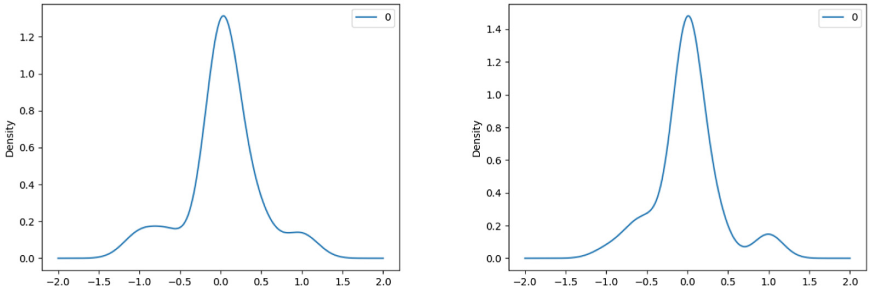

Fig. 3. Density distributions of shape key intensity values for happy and surprise expressions depicted on the left and right side respectively.

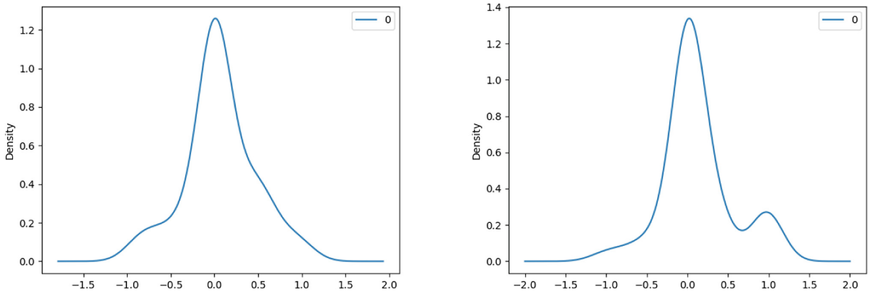

Fig. 4. Density distributions of shape key intensity values for disgust and fear expressions depicted on the left and right side respectively.

3.4 Result Discussion

Our analysis on shapes' intensity scores presents a new way to categorize the basic expressions using distance based scores. The result generated shows consistency with the pattern shown in other research work results; especially in terms of explaining the high confusion matrix error between a specific set of expressions. For instance, there is a high confusion error between the expressions anger and sadness [9,14,21]. Similarly, shown in Table 1, the distance based categorization puts both the expressions anger and sadness on the same side. They both have a weighted average intensities of AUs to the left side of zero. On the other hand expressions surprise, happiness, disgust and fear show a significant overlap in [18,21]. In Table 1, the direction weighted score puts the expressions surprise and happiness next to each other and also followed by disgust and fear in their order. In general, four of them are categorized on the right side from zero (all with positive numeric values).

According to [16] anger and fear are in the same dimension in terms of arousal/activity. Both are claimed to have a high activity level. Similarly in Table 1, magnitudes of the weighted numeric scores of anger and fear are found in the extreme ends. Having the highest results from the rest of the expressions can give a quantitative validation for the previous claim of a high activity group. In general, our analysis of the basic facial expressions generates a quantitative order and categorization along a plausible validation to the confusion error seen between a specific set of expressions.

4 Conclusion and Future Work

This research showed analysis of FACS AU (shape keys) in a 3D model generates a sensible way of categorization and quantification of distances between the basic human facial expressions. The results gained here confirms the usefulness of a 3D model platforms for analysis of facial expressions. Expanding this experiment to the other facial expressions on a 3D platform other than the basic ones, and also doing similar analysis on other different 3D models with FACS based AU, would be useful to further evaluate our method. Another interesting future task would be extracting the facial muscles' movement intensity measurements of the basic facial expression directly from human subjects or determining intensity values from images and then apply the same computation technique as done here. This would possibly help validate the insight gained here and also to show the co-relation between different settings; facial expressions analysis on virtual models and images or humans.

References

1. Mehrabian, A.: Communication without words. Psychol. Today **2**(4), 53–56 (1968)
2. Ekman, P., Friesen, W.: Pictures of Facial Affect. Consulting Psychologist, Palo Alto (1976)
3. Ekman, P., Friesen, W.: Facial Action Coding System: A Technique for the Measurement of Facial Movement. Consulting Psychologists Press, Palo Alto (1978)
4. Pandzic, I., Forchheimer, R. (eds.): MPEG-4 Facial Animation: the Standard, Implementationand Applications. Wiley, Hoboken (2002)
5. Adolphs, R.: Recognizing emotion from facial expressions: psychological and neurological mechanisms. Behav. Cognitive Neurosci. Rev. **1**(1), 21–62 (2002)
6. Pantic, M.: Automatic analysis of facial expressions: the state of the art. IEEE Trans. Pattern Anal. Mach. Intell. **22**(12), 1424–1445 (2000)
7. Sariyanidi, E., Gunes, H., Cavallaro, A.: Automatic analysis of facial affect: a survey of registration, representation, and recognition. IEEE Trans. Pattern Anal. Mach. Intell. **37**(62015), 1113–1133 (2015)
8. Bartlett, M.S., Hager, J.C., Ekman, P., Sejnowski, T.J.: Measuring facial expressions by computer image analysis. Psychophysiology **36**, 253–263 (1999)
9. Zhong, L., Liu, Q., Yang, P., Liu, B., Huang, J., Metaxas, D.N.: Learning active facial patches for expression analysis. In: CVPR (2012)
10. ISO/IEC JTC1/SC29/WG11 MPEG96/N1365: MPEG4 SNHC: Face and body definition and animation parameters (1996)
11. Hai, T.S., Thai, L.H., Thuy, N.T.: Facial expression classification using artificial neural network and k-nearest neighbor. Int. J. Inf. Technol. Comput. Sci. **7**, 27–32 (2015)

12. Zhang, Z., Lyons, M.J., Schuster, M., Akamatsu, S.: Comparison between geometry-based and gabor-wavelets-based facial expression recognition using multilayer perceptron. In: 3rd Automatic Face & Gesture Recognition. IEEE (1998)
13. Jaffar, M.A., Al Eisa, E.: Classification of facial expression using transformed features. Int. J. Inf. Electron. Eng. **4**(4), 269–273 (2014)
14. Soyel, H., Demire, H.: Facial expression recognition using 3D facial feature distances. In: Affective Computing, Focus on Emotion Expression, Synthesis and Recognition (2008)
15. de Bittencourt Zavan, F.H., Nascimento, A.C.P., e Silva, L.P., Bellon, O.R.P., Silva, L.: 3D face alignment in the wild: a landmark-free, nose-based approach. In: Hua, G., Jégou, H. (eds.) ECCV 2016. LNCS, vol. 9914, pp. 581–589. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-48881-3_40
16. Scherer, K.R.: What are emotions? And how can they be measured? Soc. Sci. Inf. **44**(4), 695–729 (2005)
17. Takahashi, K.: Remarks on emotion recognition from bio-potential signals. In: Proceedings of the 2nd International Conference on Autonomous Robots and Agents. IEEE (2004)
18. Matlovic, T., Gaspar, P., Moro, R., Simko, J., Bielikova, M.: Emotions detection using facial expressions recognition and EEG. In: International Workshop on Semantic and Social Media Adaptation and Personalization. IEEE (2016)
19. Hamedi, M., Salleh, S.-H., Swee, T.T., Kamarulafizam: Surface electromyography-based facial expression recognition in bi-polar configuration. J. Comput. Sci. **7**(9), 1407–1415 (2011)
20. DasGupta, A.: Probability for Statistics and Machine Learning: Fundamentals and Advanced Topics. Texts in Statistics. Springer, New York (2011). https://doi.org/10.1007/978-1-4419-9634-3
21. Zhang, F., Zhang, T., Mao, Q., Xu, C.: Joint pose and expression modeling for facial expression recognition. In: CVPR (2018)
22. Ying, A.C.C., Ujir, H., Hipiny, I.: 3D facial expression intensity measurement analysis. In: Zulikha, J.N.H., Zakaria (eds.) Proceedings of the 6th International Conference of Computing and Informatics, pp. 43–48. School of Computing, Sintok (2017)
23. Pantic, M., Rothkrantz, L.: An expert system for recognition of facial actions and their intensity. American Association for Artificial Intelligence (2000)
24. Hussain, N., Ujir, H., Hipiny, I., Minoi, J.-L.: Facial action units recognition for emotional expression. Advanced Science Letters (ICCSE 2017), vol. 24 (2017)
25. Sun, Y., Reale, M., Yin, L.: Recognizing partial facial action units based on 3D dynamic range data for facial expression recognition. In: 8th IEEE International Conference on Automatic Face & Gesture Recognition, pp. 1–8 (2008)
26. https://github.com/hansonrobotics/HEAD/tree/master/src/blender_api. Accessed 11 Nov 2019
27. Mahmood, S.A., Ghani, R.F., Kerim, A.A.: Nose tip detection using shape index and energy effective for 3D face recognition. Int. J. Mod. Eng. Res. (IJMER) **3**(5), 3086–3090 (2013)