UIBVFED-Mask: A Dataset for Comparing Facial Expressions with and without Face Masks

Miquel Mascaró-Oliver [ID], Ramon Mas-Sansó [ID], Esperança Amengual-Alcover *[ID] and Maria Francesca Roig-Maimó [ID]

Department of Mathematics and Computer Science, University of the Balearic Islands, Ctra. de Valldemossa, Km. 7.5, 07122 Palma de Mallorca, Spain
* Correspondence: eamengual@uib.es

Abstract: After the COVID-19 pandemic the use of face masks has become a common practice in many situations. Partial occlusion of the face due to the use of masks poses new challenges for facial expression recognition because of the loss of significant facial information. Consequently, the identification and classification of facial expressions can be negatively affected when using neural networks in particular. This paper presents a new dataset of virtual characters, with and without face masks, with identical geometric information and spatial location. This novelty will certainly allow researchers a better refinement on lost information due to the occlusion of the mask.

1. Introduction

Facial expression datasets are widely used in many studies in different areas such as psychology, medicine, art and computer science. Datasets include photographs or videos about real people performing facial expressions. These images are properly labeled according to their use.

Datasets can be classified depending on how the images have been collected. Thus, we can distinguish datasets with images taken in controlled environments, such as the Extended Cohn-Kanade (CK+) [1] or the Karolinska Directed Emotional Faces (KDEF) [2]. The main differences among datasets in controlled environments are the type of focus (frontal or not) and the lighting (controlled or not). Other data sets, for example the Static Facial Expression in the Wild (SFEW) [3] or the AffectNet [4], contain images taken in the wild. In Mollahosseini et al. [4] the main characteristics of the fourteen most relevant datasets used in Facial Expression Recognition (FER) are summarized. None of the data sets mentioned above is composed of synthetic characters.

With the COVID-19 pandemic situation, FER must deal with a new challenge: the ability to accurately categorize an emotion expression [5–7]. As reported in recent literature, face masks change the way faces are perceived and strongly confuse in reading emotions [8,9].

Particularly, Pavlova and Sokolov [10] present an extensive survey about current research on reading covered faces. They conclude that clarification of how masks affect face reading call for further tailored research. A relevant number of works, mainly in the data security field, focus on facial recognition experiments with masks [11].

In Bo et al. [12] the authors recognize that there are no publicly available masked FER datasets and propose a method to add face masks to existing datasets.

We are mainly interested in the synthesis and recognition of facial expressions and in their possible applications. In previous work we presented UIBVFED, a virtual facial expression database [13].

In this work we extend the range of the UIBVFED database to also cope with the face occlusion introduced by the generalized use of masks. We rename this new database

as UIBVFED-MASK. UIBVFED-MASK provides a working environment allowing total control over virtual characters, beyond controlling focus and lighting. The goal is to bring extra control of the precise muscle activity of each facial expression. With the inclusion of face masks on the characters, we can offer the research community two datasets with the same labeled expressions and the effects of mask occlusion as a new tool to remedy the deficiencies mentioned above. To the best of our knowledge, UIBVFED is the facial expression database with the largest number of labeled expressions. The precise description of the expressions is what will allow to expand the study to deal with occlusion problems such as the use of surgical masks.

We propose a dataset with synthetic avatars because the images of the dataset are generated according to the Facial Action Coding System (FACS) and, therefore, we can assure that the labeling of the images is objective. Moreover, the use of synthetic datasets is becoming popular as they provide objective automatic labeling and are free of data privacy issues. They have proved to be a good substitution for real images, as they get recognition rates similar to the real ones [14,15].

2. Data Description

UIBVFED is the first database made up of synthetic avatars that categorizes up to 32 facial expressions. This dataset can be managed interactively by an intuitive and easy to use software application. The dataset is composed of 640 facial images from 20 virtual characters each creating 32 facial expressions. The avatars represent 10 men and 10 women, aged between 20 and 80, from different ethnicities. Figure 1 shows the 32 facial expressions plus the neutral of one of the 20 characters in the database. Expressions are classified based on the six universal emotions (Anger, Disgust, Fear, Joy, Sadness, and Surprise). The 32 facial expressions are codified accordingly with the Facial Action Coding System (FACS) [16] and follow the terminology defined by Faigin [17] whose reference is considered the standard to follow by animators and 3D artists. Information about the landmarks for all characters and expressions is included in the dataset.

The UIBVFED database has been rebuilt in order to include facial masks. With this purpose in mind, a facial geometry has been added to the initial facial geometries. This new facial geometry corresponds to a textured polygonal mesh with 1770 vertices and has been manually added to the geometry of each of the 20 characters in the database. We have made the necessary linear transformations to adjust the mask to the character's face. Figure 2 shows the polygon mesh face mask. The geometry represents a fabric mask that covers part of the face, similar to surgical masks and to those used day-to-day during the COVID-19 pandemic. Character's and mask geometries are implemented in a 3D scene in the Unity environment [18]. We have used the Unity interactive tools to adjust the mask to the geometry of each character.

Figure 3 shows an example of the *enraged shouting* expression in the original UIBVFED dataset and in UIBVFED-MASK.

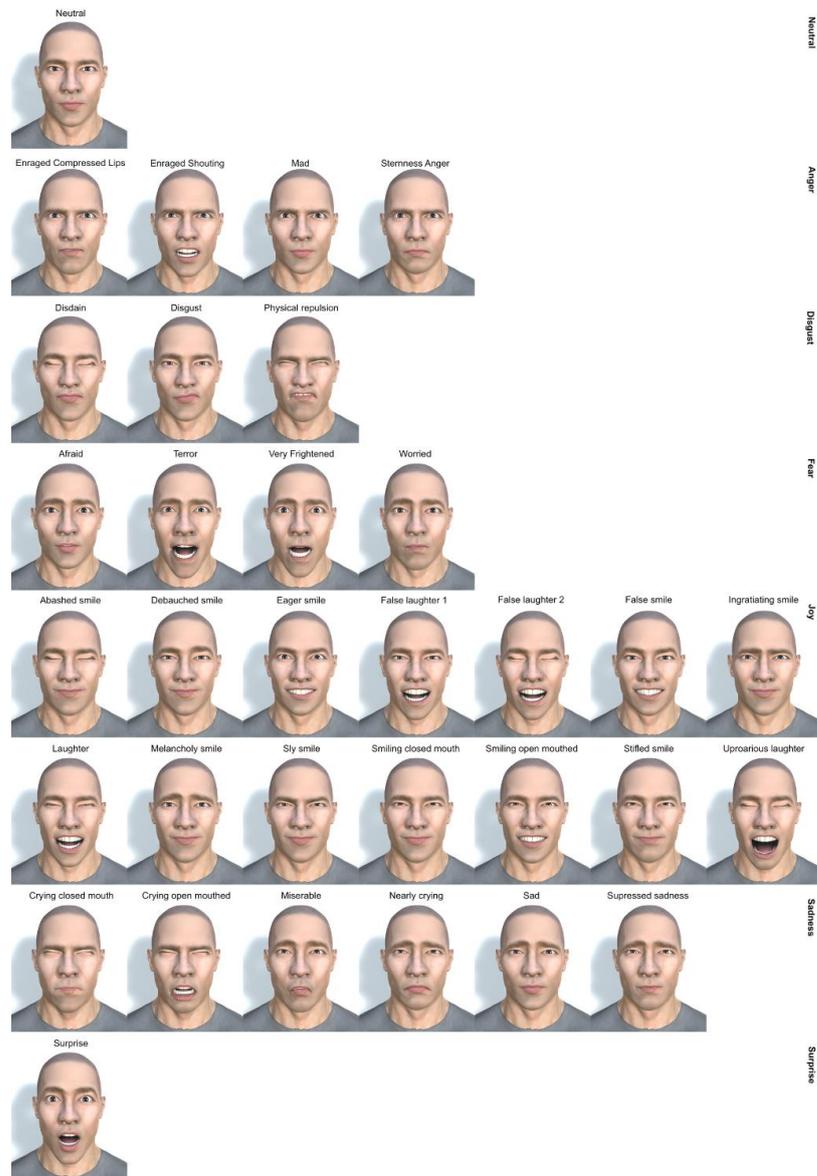Figure 4 shows the complete set of characters in the UIBVFED-MASK dataset.

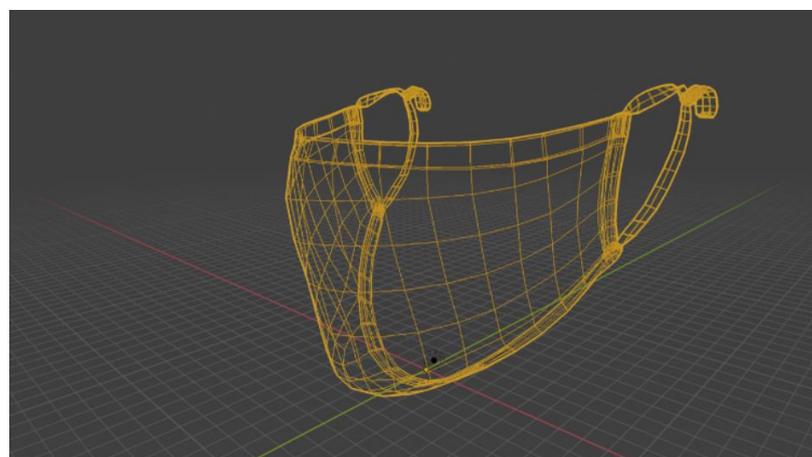**Figure 1.** The 32 expressions plus the neutral one of the UIBVFED database and their associated emotion.

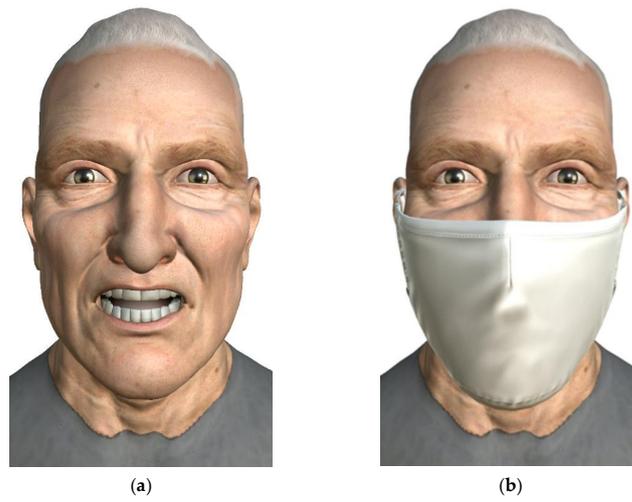**Figure 2.** Polygon mesh face mask.

Figure 3. Example of the *enraged shouting* expression: (**a**) in the UIBVFED dataset without mask; (**b**) in the UIBVFED-MASKS dataset with the facial mask.

Figure 4. UIBVFED-MASK complete set of characters.

3. Methods

The full set of images has been generated with the Unity environment. The basis is a white background scene with the character geometry located at the center of a Cartesian coordinate system.

Characters have been interactively generated with Autodesk Character Generator [19]. The interaction provided by this tool has allowed to select among several options to define 10 men and 10 women from different ethnicities and ages depending on the skin textures applied. Each character is composed by the following geometries:

- A geometry that comprises the character skin (20,951 polygons approximately);
- A geometry that represents the lower denture (4124 polygons approximately);
- Geometries for the upper denture and the eyes (4731 polygons);
- The before mentioned surgical mask geometry (see Figure 2).

For the first two geometries blend shapes have been used to allow animation to reproduce any facial expression. More specifically, the skin and the lower denture have 65 and 31 blend shapes, respectively. Each blend shape has been labeled intuitively with the muscular movement that it represents. For example, *MouthOpen* for the mouth opening or *ReyeClose* for the right eye closure. Blend shapes' activation depends on a numerical value between 0 (no effect) and 100 (maximum effect).

From the descriptions of the 32 expressions plus the neutral one considered by Faigin [17], the expressions have been reproduced using muscular movement description. Table 1 summarized the values for each deformation for the *Enraged Shouting* expression (Anger emotion) shown in Figure 3. The first column is the numerical identification of the active blend shape and depends on the file order generated by the Autodesk application. The second column corresponds to the name: h_expressions is for the skin geometry and h_teeh for the lower denture geometry. Finally, the third column is the deformation value.

Table 1. Blend shapes for the *Enraged Shouting* expression.

Id	Name	Value
0	h_expressions.AE_AA_h	25
13	h_expressions.RlipUp_h	50
14	h_expressions.LlipUp_h	50
15	h_expressions.RlipDown_h	50
16	h_expressions.LlipDown_h	50
21	h_expressions.JawCompress_h	25
41	h_expressions.RneckTension_h	100
42	h_expressions.LneckTension_h	100
43	h_expressions.Rnostril_h	100
44	h_expressions.Lnostril_h	100
47	h_expressions.ReyeOpen_h	50
48	h_expressions.LeyeOpen_h	50
51	h_expressions.RlowLid_h	50
52	h_expressions.LlowLid_h	50
61	h_expressions.RRbrowDown_h	100
62	h_expressions.RbrowDown_h	100
63	h_expressions.LbrowDown_h	100
64	h_expressions.LLbrowDown_h	100
12	h_teeth.t_AE_AA_h	25

To provide a more precise description of each expression, the datasets include the correspondence between the deformers and the coding with Action Units (AUs), which is the broad notation used in the Facial Action Coding System (FACS) [16]. For example, the *Enraged Shouting* expression has the following AUs: AU 4 Brow Lowerer, AU 5 Upper Lid Raiser, AU 6 Cheek Raiser, AU 7 Lid Tightener, AU 10 Upper Lip Raiser and AU 11 Nasolabial Deepener.

The 20 characters from each dataset have a similar geometric configuration although they are completely different. This allows to parameterize all 32 expressions with a single configuration of the deformers.

The scene lighting is a single light configuration, Spot Light, positioned in the front left side of the character plus the ambient light, skybox type. The lighting location is shown in Figure 5.
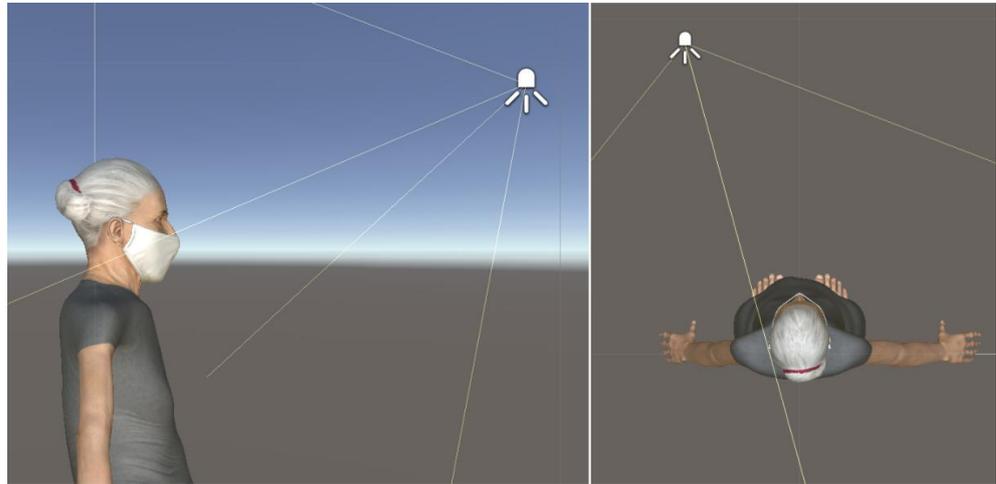

Figure 5. Lighting configuration.

As shown in Figure 5, the result of this configuration is a non-flat lighting but without intention and shadows that could hinder FER tasks.

The databases have been created by running a script that places each character in the center of the scene. The character is portrayed with a front-facing camera and the result is a 1080 × 1920 pixel resolution PNG image, RGB and with no alpha channel. The images have been organized into folders, one for each universal emotion: anger, disgust, fear, joy, sadness and surprise, plus neutral. Each folder contains the files that correspond to all the expressions performed by each one of the 20 characters. The images are labeled with the name of the character and the name of the expression following Faigin's terminology.

Along with UIBVFED-MASK, the UIBVFED database follows the same structure [13]. Moreover, both datasets also contain information about the landmarks. This information consists of 51 points in 3D space for the identification of facial features. Landmarks information is provided with the points overlapping the image. This information is additionally provided in textual form with a file that contains the numerical values that define the spatial coordinates of the 51 points for all characters and for all the expressions.

4. An Experience Using the UIBVFED-Mask Dataset

We used a simple previously trained Convolutional Neural Network (CNN) with the UIBVFED dataset (the facial expression dataset without facial masks) [20] to recognize the six universal emotions plus the neutral emotion. Using this CNN, we obtained an overall accuracy of 0.79.

With this CNN model (trained with the images without facial masks), we tried to classify a random subset of the 20% of the images of the UIBVFED-Mask (the facial expression dataset of avatars wearing facial masks), and we got an overall accuracy of 0.38. That represents a dramatic decrease of almost 52% on the overall accuracy.

However, when we trained (and tested) the model with facial masks (UIBVFED-Mask) we obtained an overall accuracy of 0.65, which improved 71% the initial overall accuracy.

Table 2 shows a summary of the accuracy results obtained for each of the three training/testing datasets scenarios evaluated with the same CNN model.

Table 2. Overall accuracy obtained in each of the training/testing scenarios.

Training Dataset	Testing Dataset	Accuracy
UIBVFED	UIBVFED	0.79
UIBVFED	UIBVFED-Mask	0.38
UIBVFED-Mask	UIBVFED-Mask	0.65

5. Conclusions

In this paper we present a new dataset of virtual characters with face masks, the UIBVFED-Mask, with identical geometric information and spatial location than the already existing UIBVFED dataset. We think that this novelty will certainly allow researchers a better refinement on lost information due to the occlusion of the mask. To show how the proposed database can improve the results of the recognition of emotions while wearing face masks, we introduce an experience using the new database. In this experience, we have numerically observed that the loss of facial information produced by the partial occlusion of the face negatively affects the recognition rate of our neural network. In this situation, our experiments show a noticeable decrease in accuracy when the network is trained using non-masked images. With no occlusion, the network accuracy was up to a 0.79 value. However, when trying to recognize the expressions from the images wearing a mask, we went down to a 0.38 accuracy value. The use of the proposed database UIBVFED-MASK for training and testing purposes led us to an increase of a 70% in the recognition rate (up to a 0.65 value).

The results of this work show that a database of facial expressions of people wearing a mask can be a relevant contribution in the emotion and facial expression recognition field.

Author Contributions: Conceptualization, M.M.-O. and E.A.-A.; methodology, M.M.-O., M.F.R.-M., R.M.-S. and E.A.-A.; software, M.M.-O.; validation, M.M.-O., M.F.R.-M., R.M.-S. and E.A.-A.; formal analysis, M.M.-O., M.F.R.-M., R.M.-S. and E.A.-A.; investigation, M.M.-O., M.F.R.-M., R.M.-S. and E.A.-A.; resources, M.M.-O.; data curation, M.M.-O., M.F.R.-M., R.M.-S. and E.A.-A.; writing—original draft preparation, M.M.-O. and E.A.-A.; writing—review and editing, M.M.-O., M.F.R.-M., R.M.-S. and E.A.-A.; visualization, M.M.-O.; supervision, M.M.-O., M.F.R.-M., R.M.-S. and E.A.-A.; project administration, M.M.-O., M.F.R.-M., R.M.-S. and E.A.-A.; funding acquisition, M.F.R.-M. and R.M.-S. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lucey, P.; Cohn, J.F.; Kanade, T.; Saragih, J.; Ambadar, Z.; Matthews, I. The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), San Francisco, CA, USA, 13–18 June 2010; pp. 94–101.
2. Lundqvist, D.; Flykt, A.; Öhman, A. Karolinska directed emotional faces. *Cogn. Emot.* **1998**. [CrossRef]
3. Dhall, A.; Goecke, R.; Lucey, S.; Gedeon, T. Static facial expression analysis in tough conditions: Data, evaluation protocol and benchmark. In Proceedings of the 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), Barcelona, Spain, 6–13 November 2011; pp. 2106–2112.
4. Mollahosseini, A.; Hasani, B.; Mahoor, M.H. AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild. *IEEE Trans. Affect. Comput.* **2019**, *10*, 18–31. [CrossRef]
5. Freud, E.; Stajduhar, A.; Rosenbaum, R.S.; Avidan, G.; Ganel, T. The COVID-19 pandemic masks the way people perceive faces. *Sci. Rep.* **2020**, *10*, 22344. [CrossRef] [PubMed]

6. Grundmann, F.; Epstude, K.; Scheibe, S. Face masks reduce emotion-recognition accuracy and perceived closeness. *PLoS ONE* **2021**, *16*, e0249792. [CrossRef] [PubMed]
7. Marini, M.; Ansani, A.; Paglieri, F.; Caruana, F.; Viola, M. The impact of facemasks on emotion recognition, trust attribution, and re-identification. *Sci. Rep.* **2021**, *11*, 5577. [CrossRef] [PubMed]
8. Carbon, C.-C. Wearing Face Masks Strongly Confuses Counterparts in Reading Emotions. *Front. Psychol.* **2020**, *11*, 566886. [CrossRef] [PubMed]
9. Barros, P.; Sciutti, A. I Only Have Eyes for You: The Impact of Masks on Convolutional-Based Facial Expression Recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Nashville, TN, USA, 19–25 June 2021; pp. 1226–1231.
10. Pavlova, M.A.; Sokolov, A.A. Reading Covered Faces. *Cerebral Cortex* **2021**, *32*, 249–265. [CrossRef] [PubMed]
11. Golwalkar, R.; Mehendale, N. *Masked Face Recognition Using Deep Metric Learning and FaceMaskNet-21*; Social Science Research Network: Rochester, NY, USA, 2020.
12. Bo, Y.; Wu, J.; Hattori, G. Facial Expression Recognition with the advent of Face Masks. In Proceedings of the 19th International Conference on Mobile and Ubiquitous Multimedia, Essen, Germany, 22 November 2020.
13. Oliver, M.M.; Alcover, E.A. UIBVFED: Virtual facial expression dataset. *PLoS ONE* **2020**, *15*, e0231266. [CrossRef] [PubMed]
14. Colbois, L.; de Freitas Pereira, T.; Marcel, S. On the use of automatically generated synthetic image datasets for benchmarking face recognition. In Proceedings of the 2021 IEEE International Joint Conference on Biometrics (IJCB), Shenzhen, China, 4–7 August 2021; pp. 1–8.
15. del Aguila, J.; González-Gualda, L.M.; Játiva, M.A.; Fernández-Sotos, P.; Fernández-Caballero, A.; García, A.S. How Interpersonal Distance Between Avatar and Human Influences Facial Affect Recognition in Immersive Virtual Reality. *Front. Psychol.* **2021**, *12*, 675515. [CrossRef] [PubMed]
16. Ekman, P.; Friesen, W.V. *Facial Action Coding System: A Technique for the Measurement of Facial Movement*; Consulting Psychologists Press: Palo Alto, CA, USA, 1978.
17. Faigin, G. *The Artist's Complete Guide to Facial Expression*; Watson-Guptill: New York, NY, USA, 1990; ISBN 0-307-78646-3.
18. Unity Technologies. Unity Real-Time Development Platform | 3D, 2D VR & AR Engine. Available online: https://unity.com/ (accessed on 10 October 2022).
19. Autodesk Character Generator. Available online: https://charactergenerator.autodesk.com/ (accessed on 25 June 2018).
20. Carreto Picón, G.; Roig-Maimó, M.F.; Mascaró Oliver, M.; Amengual Alcover, E.; Mas-Sansó, R. Do Machines Better Understand Synthetic Facial Expressions than People? In Proceedings of the XXII International Conference on Human Computer Interaction, Teruel, Spain, 7–9 September 2022; pp. 1–5.