MIGMA: The Facial Emotion Image Dataset for Human Expression Recognition

Jhennifer Cristine Matias[1](✉), Tobias Rossi Müller[1], Felipe Zago Canal[1], Gustavo Gino Scotton[1], Antonio Reis de Sa Junior[3], Eliane Pozzebon[1,2], and Antonio Carlos Sobieranski[1,2]

[1] Department of Computing (DEC), Federal University of Santa Catarina, Florianópolis, Brazil
j.matias@grad.ufsc.br

[2] Post Graduate Program in Information Technology and Communication (PPGTIC), Federal University of Santa Catarina, Florianópolis, Brazil

[3] Department of Medical Clinic (DCM), Federal University of Santa Catarina, Florianópolis, Brazil

Abstract. Recognition of emotions from facial information is a simple task well-performed by humans, but very complex to be executed computationally. Since many of the computational trials to solve this problem lead to studies for a generic approach, it needs to be comprehensive to provide a solution analytically possible. Several approaches were proposed over the past few years, and apart of the chosen model, a considerable amount of input samples must be used to train the computational approach properly. Over the literature image datasets for facial recognition can be found, and despite the fact that several of them are in public domain available, the presented images are usually very restricted, with slight variations and limited number among participants, low miscegenation mixing and short age ranges, which ends up making the studies and new algorithms very specific. For this purpose, this paper has as main goals (i) to present a newly designed dataset entitled MIGMA for human expression recognition from facial images and (ii) to address essentials features of this dataset such as high-quality spatial resolution images, varied ethnicity, ages and genders, and including non-induced and induced expressions via emoji. The dataset has 323 participants and 15k images on its current first version, taking into account 8 distinct emotions, photographed in an academical environment in a south American country. In contrast, all the participants completed questionnaires for anxiety and depression, allowing to address further studies in this area with facial emotions. The obtained dataset was tested in an experimental environment using a Convolutional Neural Network recognizer for general recognition overall and class separability estimation.

Keywords: MIGMA · Facial emotion dataset · Feature recognition · Computer vision dataset

J. M. R. S. Tavares et al. (Eds.): CIARP 2021, LNCS 12702, pp. 153–162, 2021.
https://doi.org/10.1007/978-3-030-93420-0_15

1 Introduction

Emotion recognition from facial expressions, besides instinctive and natural for humans, is a very complex task to be performed computationally. This fact is due its natural sense, with hard description and explicitness, and as a consequence, it is difficult to be reproduced in analytical terms. The achievement of a good general solution for this problem has several practical implications [1], being applied by medicine education, robotics, driver safety, games and the educational area. However, emotion recognition has been still an open problem in the computational area. Lately, thanks to the recently developed hardware architectures and computational models, approaches were designed to be dynamic in terms of recognition and generalization. This generalization, on the other hand, can be achieved at the cost of requirements for an large input dataset, used to train the recognition model properly.

In fact, the emerging convolutional neural network models (CNN's) are paving the way for very interesting generic solutions, the datasets available are not evolving at the same compass. Larger datasets are required nowadays to properly train the CNN's in order to be effective and provide the expected results. Over the past few years, several image sets for recognizing human emotions in facial expressions have been proposed, as demonstrated in our related works section. However, the found databases are not always able to meet minimal requirements to develop a robust facial emotion recognition system. As a consequence, it is common to see a large amount of projects including their own dataset (but not for public domain) to develop their computational approaches. For instance, JAFFE and Cohn-Kanade both have low miscegenation and a higher number of photos of people of the same gender; JAFFE and MMI have a low number of participants in their bases, being corresponding to 10 and 75, respectively. There are datasets in addition with images in an uncontrolled environment, such as FER-2013, where the images were collected from the WEB. All of these drawbacks result in a problem for the training of classifiers and the new convolutional neural networks, making the computational model restricted to very specific domains.

In the present study, a new dataset for emotion recognition from facial images is proposed. The dataset was developed taking into account the aforementioned limitations and providing higher resolution images, in a well-controlled environment, good number of participant and images. Unlike others, our dataset presents a very high miscegenation of participants, since it was built in an academic environment, including people from the most varied ethnic groups. Additionally, as far as we know, this is the very first time a emotion facial image dataset is provided taking a higher miscegenation from a South American nation for open domain, high-resolution images, and associated with a medical scale of psychiatric symptoms.

2 Related Works

Given the importance and need for automated emotion recognition systems for facial images, it became necessary to create datasets that allow us to both training and testing the effectiveness for such in developing algorithms. Over the past few years, several datasets have been appeared in different countries with varied purposes and characteristics, such as those presented in the Tables 1 and 2.

The Japanese Female Facial Expression (JAFFE), despite being developed in 1998, it is still widely used since its creation by a group of Japanese researchers [2]. The images in this dataset have a spatial resolution of 256 × 256 pixels and were taken in gray scale [3]. The base has a total of 213 images of 10 Japanese women, in which they express 7 emotions: happiness, sadness, surprise, disgust, fear, anger and neutrality. However, emotions are not considered pure, and were classified according to their predominant emotion.

Another well-known dataset is Cohn-Kanade (CK), and it had its first version developed in 2000, whose project was carried out by [4] and is currently active and managed by the University of Pittsburgh. According to [3] in its first version, it already had 486 image sequences, which, unlike static images, had information about the process of creating emotion. Initially, this dataset had 97 participants of both gender and ages ranging from 18 to 30 years, the majority of participants were Euro-American or African-American, only 3% were Latin or Asian. In 2010, this dataset was updated, and the newest version became known as CK+, having 107 new sequences added and 26 new participants representing the following 6 emotions: anger, contempt, heartbreak, fear, happiness, fear and surprise [5].

MMI is a dataset that started to be developed in 2002, being composed by images and videos of 75 people until the present moment [6]. Participants are men and women, aged from 19 to 62 years old and from different ethnicities. This dataset has static images and sequences of images of the participants in which they express the six basic expressions: happiness, anger, sadness, disgust, fear and surprise. Furthermore, according to [3] this is the first web-based facial expression database, and has as main drawback a controlled environment.

BU-3DFE was created in 2006 by [7], being different from the other ones since its has 3D image sequences. This database is divided into 7 emotions and composed by a total of 2500 sequences with a spatial resolution of 1040 × 1329. As presented by [3], the base has 100 participants, men and women, aged from 18 to 70 years old, of different ethnicities including whites, blacks, Indians, Asians, among others, they performed 6 emotions: sadness, happiness, fear, angry, surprise and disgust.

FER-2013 is a dataset developed for the International Conference on Machine Learning (ICML - 2013), being shared publicly afterwards. The dataset has a total of 35887 images divided into 7 emotions. Unlike other datasets, it is not known for sure about the source participants because the images were obtained through google images API. The resolution of the photos are 48 × 48 and they are in gray scale [8], being already normalized to be used as input layer in training procedures.

In addition to the datasets already mentioned, there are countless others that can be found over the internet. It has been observed that many computational approaches for emotion recognition are closed, since they were developed to provide artifacts for their own computational approaches [9–11], indicating a real research demand and that a new one is still required.

Table 1. Comparison of datasets proposed over time, and the positioning of the presented dataset MIGMA in quantitative terms.

Name	No. of images	Resolution	Color	No. of emotions
JAFFE	213	256 × 256	Gray	7
CK	486	640 × 490	Gray	6
MMI	545	720 × 576	Color	6
BU-3DFE	2500	1040 × 1329	Color	7
CK+	593	640 × 490	Color and Gray	7
FER-2013	35.887	48 × 48	Gray	7
MIGMA	15.071	1920 × 1080	Color	8

Table 2. (Cont.) Comparison of datasets proposed over time, and the positioning of the presented dataset MIGMA in quantitative terms.

Name	Images/Images sequence	Year	Environment	No. of participants
JAFFE	Image	1998	Controlled	10
CK	Image sequence	2000	Controlled	97
MMI	Image, Image sequence	2002	Controlled	75
BU-3DFE	Image, 3D models	2006	Controlled	100
CK+	Image sequence	2010	Controlled	123
FER-2013	Image	2013	In-the-wild	–
MIGMA	Image	2020	Controlled	323

3 Methodology: Proposed Dataset Environmental Protocol

Figure 1 illustrates the computational flow adopted in our approach and an example of the computational environment prepared to acquire the images, respectively in (a) and (b). In (a), the protocol flow is demonstrated in five stages: (1) participant information collection, (2) psychological profile gathering, (3) image acquisition in unsupervised manner and (4) image acquisition in supervised manner, and (5) images verification and eventual corrections.

A web-based system was developed in order to conduct the participant information and images acquisition, following the five steps aforementioned in sequence. Concurrent participants of the dataset are allowed in this web system. At the first step (Fig. 1-(a)) the presented system required that the user provides his/her identification such as register, and afterwards, answering a questionnaire elaborated to assess demographic information (age, gender, marital status, ethnicity and religion), as well psychiatry symptoms of the participants.

At the second stage, participants were submitted to a test to track their psychiatric symptoms, where questions about depression, anxiety and stress were

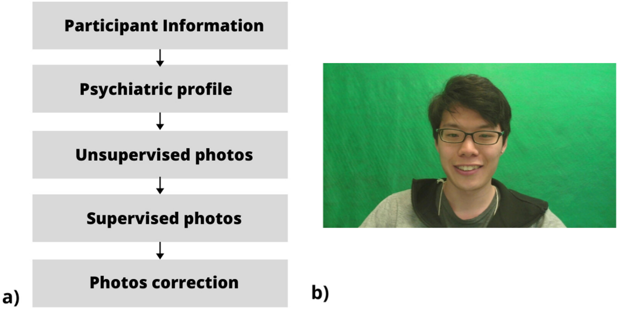

Fig. 1. General overview of the proposed dataset, having its computational flow in (a), and an example of the acquisition process illustrated in (b).

answered. The following instrument for screening were used with the supervision of a psychiatrist: Beck Depression Inventory (BDI) [12]; Beck Anxiety Inventory (BAI) [13]; Adult ADHD Self-Report Scale (ASRS 18) [14]; Obsessive-Compulsive Inventory-Revised (OCI-R) [15] and Self Reporting Questionnaire (SRQ-20) [16,17].

The final step of our protocol is the acquisition itself, in which before the images capture themselves, a script is executed, so that the characteristics of the images such as color and quality are kept. Thus, the 6 capture sessions of 8 images each, are taken by the participant himself through the system. The third stage is removed from the first 3 sessions, which participants alone must imagine each emotion and save their image.

During the fourth stage, 3 more sessions are held with 8 photos each, but different from the previous one, the participant now has an emoticon as support to reproduce the emotion presented. Due to the fact that the participants take their own photos, in the fifth stage a supervisor was chosen to review possible flaws, such as reddish or blurred images, so if an inconsistencies is found, the images can be retaken.

Regarding the capture environment shown in the Fig. 1-b, a controlled environment was chosen for the realization of this data set, so the images contained in it have a standard background: all photographs were taken with the same camera as model being HP hd-4110, with 13 mega-pixels and connected to a conventional computer via USB. During the acquisition procedure, only one face appears at a time for the participant, avoiding induction for the next emotion expressions. Distinct eight emotion expressions from facial images are shown in the Fig. 2, where the expressions sadness, fear, happiness, disgusting, angry, surprised, contempt and neutral are demonstrated from (a) to (h), respectively.

The full dataset can be downloaded in our institutional web site migma.ufsc.br. To download the dataset, a research term must be signed and submitted electronically. No participant information is released, except photos categorized into expressions. Psychiatric symptoms, on the other hand, is protected under ethical terms and can be released only in the form of meta-data,

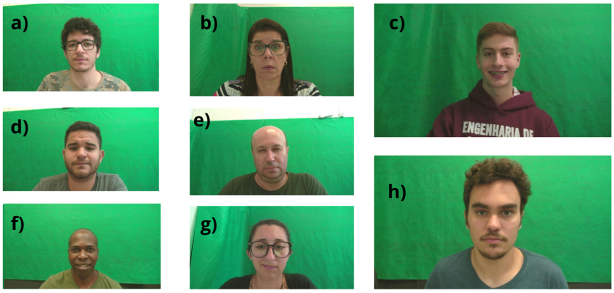

Fig. 2. Distinct expressions considered for the proposed dataset.

excluding names and mentions to images. It can be used for further assays involving computer vision associated with the corresponding medical area, opening a wide range of possible studies whose median faces can be correlated to its medical scales to track possible depressive and anxiety symptoms.

4 Results

4.1 Dataset Properties

The target population chosen was college students from different courses and university staffs, reaching a total of 323 participants. All participants completed a structured and self-administered form composed of closed questions about depression, anxiety and attention deficit and hyperactivity symptoms (ADHD). The demographic characteristics of the sample were 200 men (61.91%), 119 women (36.84%) and 3 no reply (0.92%), with an average age of 26.54 years (STD = 8.12; range 18 to 55 years). In additionally, the self-reported skin color showed that most of the sample consisted of self-declared white students (86.06%), followed by mulatto (2.47%), black (5.88%), Asian (1.23%), other mixed colors (0.30%), and 2.78% no declared.

4.2 Dataset Statistical Analysis

Over the obtained dataset, some statistics were obtained in order to validate some premises such as regularity of the expressions and standard deviation. The first analysis performed was the acquisition of the mean faces for each expression, corresponding to a simply summation of the dataset images for each category of emotion, as demonstrated in the Fig. 3. The mean faces need to be registered to a reference planar domain to be representative and avoid deformations and blurring effects. The eyes for each participant were chosen to perform this alignment.

To provide the mentioned statistics, a pre-processing step was applied to extract relevant data features from facial images and correlate them by spatial transformations in an accurate manner. Statistics were extracted from the image

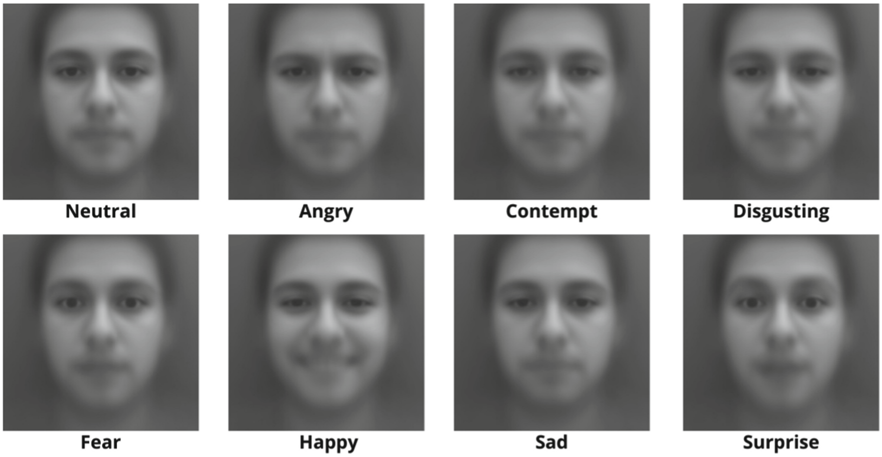

Fig. 3. Images generated summing the faces aligned by eyes for each emotion. The aligner algorithm used is from Dlib package for python 3.7.

domain and converted to a set of 2-dimensional sparse space points, allowing the use of descriptive and inferential statistics metrics, such as central tendency, measures of variance, and permutation tests. For this purpose, the Dlib library was used, being a popular set of algorithms in python package used as feature extractor of facial points. The algorithm was trained from the iBUG 300-W dataset which is used in some machine learning competitions [18]. The data returned by the function provides a cardinality of 68 two-dimensional points of relevant parts of the human face, such as mouth, nose, eyebrows, eyes and the contour of the face.

4.3 Case-Study: Dataset Performance in a Convolutional Neural Network Framework

The second way used to validate this dataset was a case-study, which aims to assess the separability of the classes belonging to the dataset with a trained convolutional neural network.

The processing pipeline performed was initiated with the transformation of the images to gray scale, then the standard facial detection algorithm from the Dlib library was used to perform the alignment. After that, a resizing using cubic interpolation was used to bring the images to 70 × 70 resolution. After the first resizing, a 12-pixel cut for all sides was performed, transforming the images to a 46 × 46 resolution. At that moment, normalization and equalization of the image were carried out, with the intention of reducing the input domain.

The network architecture was implemented using *Tensorflow 2*, with a pattern of *Conv2D* folowed by Batch Normalization and *Max Pooling*. This pattern was used 3 times followed by flatten, a MLP with 2 layers and a softmax for 8 classes. The training parameters were the Adam optimizer, batch size of 128,

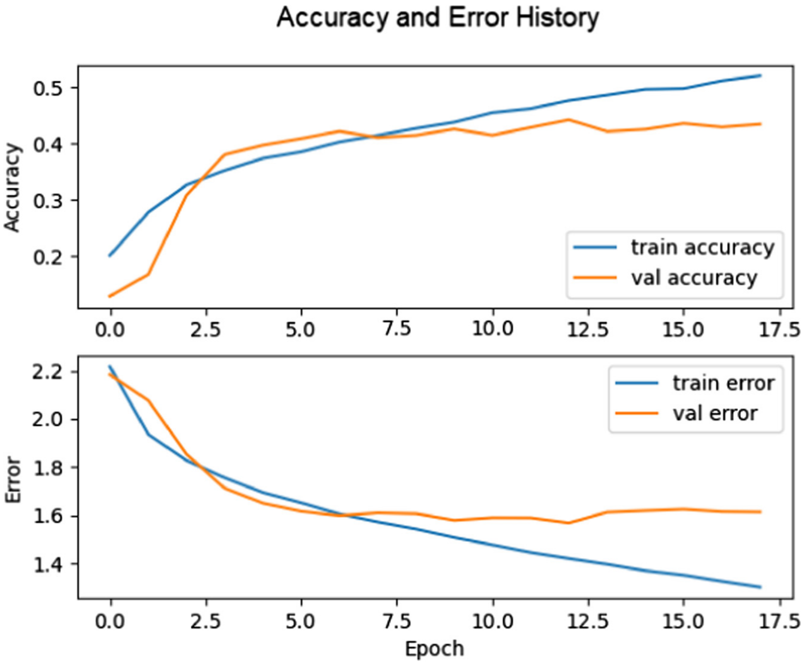

Fig. 4. History of the model training plotted with Matplotlib, a very popular python package.

learning rate of 0.0001. Two callback functions were used in the training of the network, *ModelCheckpoint* for save the model with best accuracy on validation set, and *EarlyStopping* with patience 5.

Sets of training, testing and validation were separated with size respectively 70%, 15% and 15% of the original. The samples of each participant are present in only one of the three sets, making the analysis more robust. The training history can be seen in the Fig. 4. The best model saved has 44% of accuracy on validation set and 41% in the test set.

5 Conclusion and Discussions

In present study a new dataset for emotion recognition from facial images containing more than 15k images and 323 participants was presented. The dataset entitled ******* was designed taking into account higher spatial resolution in a frontal well-controlled environment aiming to cover the most common and widely-used expressions from facial images. Differently from the existing datasets, here we focus in a high-miscegenated source, presenting facial images from the most distinct and rich types, ages and genders, since it was developed in a south American country in an academical environment. Although, this dataset includes induced and non-induced expressions associated with a psychiatric profile for each participant, performed by a physician in the area.

The preliminary statistics we obtained reveals interesting aspects in terms of regularity of the dataset when distinct emotions are compared, indicating those expressions that are consensus and ambiguity among participants. Also, we presented the mean faces obtained in our dataset, which can be used as a general pattern distribution for those expressions in classifiers or decision-making systems.

There are some important points that can limit the aforementioned classifier to achieve a higher performance when compared to other models available over the literature. A first aspect is the larger number of classes presented in our dataset, increasing the number of cluster combinations in the feature-space and the complexity. Moreover, one can observe the inclusion of some properties that may increase the dataset variance, such as the high miscegenation and larger number of participants. Nonetheless, the proposed dataset was designed taking into account a significant increment of the spatial resolution used during the acquisition, which infers directly in terms of execution time for classifiers. An example of a classifier presented a high accuracy is presented in [19], in this article the test is made with dataset JAFFE and the accuracy is 100% however, it has 213 images and 10 participants, which ends up affecting the generalization of the solution.

Acknowledgment. To the CNPq/PIBIC program and UFSC for the grant of the Scientific Initiation scholarship for the research project. This study was approved by the Ethics Committee of Plataforma Brasil, under number 13300919.6.0000.0121. All participants signed an informed consent form before participating in data collection.

References

1. Happy, S., Routray, A.: Automatic facial expression recognition using features of salient facial patches. IEEE Trans. Affect. Comput. **6**(1), 1–12 (2014)
2. Lyons, M.J., Akamatsu, S., Kamachi, M., Gyoba, J., Budynek, J.: The Japanese female facial expression (JAFFE) database. In: Proceedings of Third International Conference on Automatic Face and Gesture Recognition, pp. 14–16 (1998)
3. Anitha, C., Venkatesha, M., Adiga, B.S.: A survey on facial expression databases. Int. J. Eng. Sci. Technol. **2**(10), 5158–5174 (2010)
4. Kanade, T., Cohn, J.F., Tian, Y.: Comprehensive database for facial expression analysis. In: Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580), pp. 46–53. IEEE (2000)
5. 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: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops, pp. 94–101. IEEE (2010)
6. Pantic, M., Valstar, M., Rademaker, R., Maat, L.: Web-based database for facial expression analysis. In: 2005 IEEE International Conference on Multimedia and Expo, pp. 5-pp. IEEE (2005)

7. Yin, L., Wei, X., Sun, Y., Wang, J., Rosato, M.J.: A 3d facial expression database for facial behavior research. In: 7th International Conference on Automatic Face and Gesture Recognition (FGR06), pp. 211–216. IEEE (2006)
8. Goodfellow, I., et al.: Challenges in representation learning: a report on three machine learning contests (2013). http://arxiv.org/abs/1307.0414
9. Jazouli, M., Majda, A., Zarghili, A.: A $ p recognizer for automatic facial emotion recognition using kinect sensor. in: Intelligent Systems and Computer Vision (ISCV), pp. 1–5. IEEE (2017)
10. Zhang, Y., Ji, Q.: Active and dynamic information fusion for facial expression understanding from image sequences. IEEE Trans. Pattern Anal. Mach. Intell. **27**(5), 699–714 (2005)
11. Tarnowski, P., Kolodziej, M., Majkowski, A., Rak, R.J.: Emotion recognition using facial expressions. In: ICCS, pp. 1175–1184 (2017)
12. Beck, A.T., Steer, R.A., Carbin, M.G.: Psychometric properties of the beck depression inventory: twenty-five years of evaluation. Clin. Psychol. Rev. **8**(1), 77–100 (1988)
13. Beck, A.T., Epstein, N., Brown, G., Steer, R.A.: An inventory for measuring clinical anxiety: psychometric properties. J. Consult. Clin. Psychol. **56**(6), 893 (1988)
14. Kessler, R.C., et al.: The world health organization adult ADHD self-report scale (ASRS): a short screening scale for use in the general population. Psychol. Med. **35**(2), 245 (2005)
15. Foa, E.B., et al.: The obsessive-compulsive inventory: development and validation of a short version. Psychol. Assess. **14**(4), 485 (2002)
16. Harding, T.W., et al.: Mental disorders in primary health care: a study of their frequency and diagnosis in four developing countries. Psychol. Med. **10**(2), 231–241 (1980)
17. de Jesus Mari, J., Williams, P.: A validity study of a psychiatric screening questionnaire (SRQ-20) in primary care in the city of Sao Paulo. Br. J. Psychiatry **148**(1), 23–26 (1986)
18. Sagonas, C., Tzimiropoulos, G., Zafeiriou, S., Pantic, M.: 300 faces in-the-wild challenge: the first facial landmark localization challenge. In: Proceedings of the IEEE International Conference on Computer Vision Workshops, pp. 397–403 (2013)
19. Chen, T., et al.: Emotion recognition using empirical mode decomposition and approximation entropy. Comput. Electr. Eng. **72**, 383–392 (2018)