Journal of the Brazilian Computer Society, 2025, 31:1, doi: 10.5753/jbcs.2025.4456

A Lightweight Method for Almost Real-Time Recognition of Emotions through Facial Geometry with Normalized Landmarks

Leandro Persona [**University of São Paulo** | ***leandro.persona@alumni.usp.br***]
Fernando Meloni [**University of São Paulo** | ***fernandomeloni@alumni.usp.br***]
Alessandra Alaniz Macedo [**University of São Paulo** | ***ale.alaniz@usp.br***]

Department of Computer Science and Mathematics (DCM), FFCLRP, University of São Paulo (USP), Av. Bandeirantes, 3900 - Monte Alegre, Ribeirão Preto, SP, 14040-901, Brazil.

Received: 26 April 2024 • **Accepted:** 09 April 2025 • **Published:** 19 May 2025

Abstract Recognizing emotion is an intrinsic and fundamental part of human social interactions and is closely tied to behaviors that produce distinct facial patterns. Facial expressions indicate emotional context, so scientific, artistic, medical, and commercial interest in the field has increased, driving the development of computational techniques that can recognize emotion automatically. Although current methods provide satisfactory results, challenges persist, particularly challenges regarding the variable patterns of facial shapes and the response time achieved with low computational resources. For instance, some applications requiring instantaneous emotion recognition with high accuracy and low latency may be limited by the processing power, specially in the case of mobile devices. Here, we present a practical and simple method called REFEL (Recognizing Emotions through Facial Expression and Landmark Normalization), designed to identify facial expressions and human emotions in digital images. This method addresses sequential steps that reduce sample variability such as anatomical, scale factor and geometric variations and performs reductions of color, brightness and others in preprocessing tasks. REFEL normalizes facial fiducial points, commonly referred to as landmarks, and allows fine-tuning of informative aspects delineating facial patterns. Using landmark positions makes the process of recognizing emotions more reliable. REFEL also exploits classifiers explicitly tailored to identify facial emotions accurately. As in the case of related works, we employed Machine Learning algorithms, to achieve average accuracy higher than 90% for emotion recognition, when we applied REFEL before classification. We have experimented REFEL with various datasets including facial images that consider racial, age and gender factors as well as facial rotation. In this study, we also compared emotion classification without grouping emotions and with two emotion groups (Fear-Surprise and Anger-Disgust). Analysis of the ROC curves revealed that grouping emotions led to a slight improvement in the average performance of the REFEL method, with a 3% increase in accuracy. Our method represents an enhanced approach in terms of hit rate and response time, generates resilient outcomes, and relies less on the training set and classifier architecture. Furthermore, REFEL performs well, almost in real time, lowers the processing costs inherent to training, and is particularly suited to devices with limited processing capabilities, like cell phones. Emotion recognition methods usually have almost minimal real-time delay, which enables the system to react fast but not necessarily instantaneously. With REFEL, we hope to improve computational synthesis techniques and resources, and to help robust and motivating assistive technologies to advance. As future efforts, we intend to consider 3D images and videos.

Keywords: Multimedia Processing, Affective Computing, Machine Learning, Multimodal Interaction, Facial Patterns

1 Introduction

Recognizing emotions is a natural process of human cognitive abilities and is essential for navigating social, personal, and professional interactions. From early childhood, individuals begin to learn how to map and to interpret the emotions of others. This skill plays a critical role in effectively communicating and building relationships given that the recognized emotions inform humans of the surrounding context [Darwin, 2013; Hess, 2001; Langner *et al.*, 2010]. Mapping emotions is possible because humans typically translate their emotions into detectable physical movements, particularly facial expressions, which is vital for social interactions. Facial expressions are ubiquitous behaviors that depend little on cultural factors as demonstrated by [Ekman and Friesen, 1971]. Researchers have identified seven emotions across cultures: fear, anger, sadness, happiness, surprise, disgust, along with the neutral emotion [Ekman and Friesen, 1971; Alvarez *et al.*, 2013]. Facial expressions, characterized by their distinct and stable patterns, can be recognized even in unfamiliar individuals. This consistency enables emotion recognition to be systematized and used for automation and allows machines to interpret human emotions. In specific domains, this process can be simplified to focus on fewer than seven universal emotions, such as happiness *versus* unhappiness or interest *versus* disinterest.

Thanks to growing scientific, medical, and commercial interest in the field, facial expression-based methods for recognizing emotions have advanced rapidly [Lahiri *et al.*, 2011; XIE *et al.*, 2015; Ullah and Tian, 2021; He *et al.*, 2023]. The most common approaches for recognizing individuals and facial expressions focus on automatically detecting patterns in digital images. Notably, social media platforms, mobile devices, and digital cameras can now discern whether a person is smiling or not [Chaugule *et al*., 2016]. In addition, assistive technologies have been developed to aid individuals with behavioral syndromes (e.g., autism and mood disorders) [Picard, 2016; Cheng and Ling, 2008; Lahiri *et al*., 2011]. By enabling these individuals to react differently upon perceiving expressed emotions, these technologies contribute to enhancing their social interactions. Therefore, automatically recognizing facial expressions offers new avenues for humans and machines to interact, which is facilitated when both voluntary and involuntary facial movements can be detected.

In general, interpreting information from digital images involves automation through Machine Learning (ML) [Chollet, 2017]. However, manipulating images requires initial preprocessing [Vaillancourt, 2010], so that specific shapes such as a human face can be detected. Next pretrained models scan segments of the image and employ probabilities to confirm patterns [Wu and Ji, 2018]. Some models can detect human faces in images with accuracy above 90% [Wu and Ji, 2018; Viola and Jones, 2001]. Once the face is identified, the Regions of Interest (ROIs), which are specific facial structures including eyes, nose, mouth, and chin, among others, are mapped. The relative positions of these structures are then marked as landmarks, to which bi or tridimensional coordinates are assigned to add a layer of information.

Using landmarks provides a straightforward and objective way to recognize facial expressions because patterns of facial identification (facial recognition) and changes in facial patterns (expression and emotion recognition) can be compared. Although implementations may vary, both approaches consider the coordinates of landmarks as problem variables (reference values) when calculating distances (e.g., Euclidean, cosine, etc.) [Li *et al*., 2012; Garrido and Joshi, 2018; Mehta *et al*., 2018]. In the case of emotions, the muscle movements underlying each facial expression geometrically changes in the relative positions of ROIs [Xie and Jiang, 2017]. These changes alter the coordinates of landmarks and the distances between them. Given that a neutral face and a fearful or smiling face exhibit different relative distances, the set of variations can be employed to identify expressions [Mehta *et al*., 2018]. The generalizability of emotional responses among humans tends to simplify the problem, and allows the distances most affected by each type of facial expression to be identified. After analyzing these variations, ML classifiers can be trained to detect human emotions [Wu and Ji, 2018].

Even though current methods yield satisfactory results for specific contexts, such as photos acquired in controlled environments, recognizing emotions remains challenging. The main issue is that patterns in facial shapes, poses and images obtained under real-world conditions (uncontrolled variables like lighting, brightness, distance from the capturing device, etc.) vary [Mehta *et al*., 2018; Testa *et al*., 2019]. Differences in facial shapes tend to introduce noise into the problem, so that training will depend on factors such as race, age, gender, etc. Aspects like focal length, luminosity, framing, people's poses (rotated face images and expression of sentiment), and hardware configurations affect how pixels are concentrated and arranged, which adds noise to the data. These sources of variability negatively impact the way classifiers perform and make ML more challenging [Chollet, 2017; Mehta *et al*., 2018]. Our hypothesis is that reducing data variability by means of the processing of data can improve computational performance. Nevertheless, pixels represent the information domain in images, so traditional techniques for aligning and normalizing data need to be adapted for this context. Such adaptation demands computationally creative and conceptually elaborate strategies.

According to Oge et al., Gonzales et al. and others, one of the major challenges faced by algorithms in Digital Image Processing (DIP) in uncontrolled environments is the difficulty in achieving good results under varying conditions of luminosity and contrast [Filho, 1999; Gonzales and Woods, 2008; Kim, 2022; Koohestani *et al*., 2024]. These conditions introduce noise and variability into the data, making it difficult for algorithms to perform effectively. Moreover, the relative positioning of objects within a scene adds variability, further hindering algorithmic efficiency.

In this paper, we present a simple method for recognizing emotion in digital images. The method is called REFEL (Recognizing Emotions through Facial Expression and Landmark normalization) and operates under several conditions of variability, including scale, rotation, racial factors, and image acquisition [1]. REFEL incorporates well-known image manipulation techniques such as histogram of oriented gradient and facial alignment, which helps to harmonize and to normalize the information. Data variability is reduced through a set of normalization steps, to produce an optimized context to identify facial expressions. Next, classifiers suitable for the methodology of the study are employed to assess potential performance gains. Processing and normalization aim to reduce data variability and to simplify the problem. As mentioned, we hypothesized that reducing variability can lead to classifiers that perform better, regardless of the employed ML method. Thus, developing methods that harmonize image information can potentially yield better results.

The remainder of this paper is structured as follows: Section 2 provides a brief review of image defogging and attention mechanism. Section 3 discusses related work. Section 4 discusses the proposed REFEL method. Section 5 presents the results of the experiments. Section 6 provides final remarks and proposes future work.

2 Background

In the early 1970s, Paul Ekman conducted a groundbreaking scientific experiment that revolutionized the field of human emotion recognition [Ekman and Friesen, 1971]. At that time, individuals were believed to use their facial muscles according to a set of social conventions and expressions shaped by societal interactions, much like languages, with each region of the world having its own variations.

Ekman captured numerous images of men and women displaying various facial expressions. He then traveled to Brazil, Argentina, and Japan to conduct his experiments. To his surprise, individuals from different countries obtained the same results when they classified the images. Ekman extended the experiment to the forests of Papua New Guinea in Oceania, reaching the most remote and isolated villages. The results did not differ even among the inhabitants of these

[1] REFEL is based on the master's thesis described in [Persona, 2022]

regions, which led Ekman to conclude that human emotions expressed through facial expressions are universal and do not depend on ethnic or social factors.

Another significant contribution of Ekman's work was that the Facial Action Coding System (FACS) was created. FACS, a system for classifying human emotions, provides a standardized framework for systematically categorizing the physical expression of emotions, which allows any anatomically possible facial expression to be labeled. Figure 1 presents the six primary emotions and the neutral expression, the building blocks of FACS, subdivided into action units. Figure 1 illustrates the main differences between emotions.

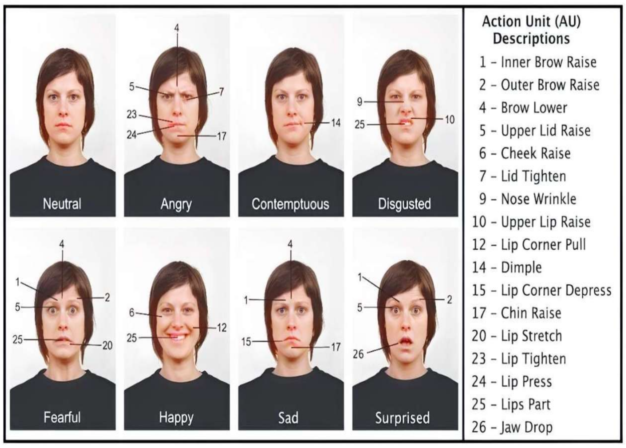


Figure 1. Illustration of FACS (Facial Action Coding System) [Langner *et al.*, 2010].

This section provides general information about the topic of our REFEL method, offering the necessary context. To present this context, we have divided the section into two subsections: (1) *Imaging Process, Face Detection, Landmarks Setting*, and (2) *Classification*. Each subsection focuses on a specific aspect and provides a comprehensive overview of the underlying concepts and techniques related to REFEL.

2.1 Imaging, Face Detection, and Landmark Setting

Interpreting emotions from images in real time includes managing images, that is, normalizing pixels, detecting faces, extracting features, and classifying patterns [Gonzales and Woods, 2008; Chollet, 2017]. During the initial preprocessing of images works out to minimize noise (e.g. pixels acquired outside the standard [Vaillancourt, 2010]) is minimized, which reduces data variance. This step allows more precise classifiers to be trained and performs well even if limited data sets are employed.

The next step analyzes human faces in images and starts by detecting which region of the image contains the target object. Viola and Jones developed a method that detects human faces fast and accurately enough for use in popular cameras [Viola and Jones, 2001]. Specifically for this purpose, today more efficient and reliable solutions exist, including the use of Histograms of Oriented Gradient (HOG) [Monzo and Mossi, 2010; Ghorbani and Dehshibi, 2015]. HOG, a digital image processing technique, is employed in computer vision and multimedia processing. This technique characterizes objects by shape and texture, on the basis of how density gradient or edge directions are distributed [Dalal and Triggs, 2005]. HOG has been used as a reference in the facial detection and recognition of various types of objects [Gonzales and Woods, 2008]. Models trained to detect human faces in images can scan segments through the image and employ probabilities to determine bounds and to confirm the whole pattern [Wu and Ji, 2018]. Certain models can reliably detect faces, even in low-resolution images [Testa *et al.*, 2019; Wu and Ji, 2018; Viola and Jones, 2001].

After face identification is confirmed, the next step is to identify the Regions of Interest (ROIs), like eyes, mouth, nose, chin, or any relevant structure related to facial expressions. Facial landmarks are subsets of the shape prediction problem, which involves locating key points and the overall shape of an object. For instance, the DLib library, available in C, C++, and Python, is commonly used to extract landmarks and provides a feature matrix of the x and y coordinates with 68 facial points, including eyes, nose, mouth, and face bounds. Landmarks may assume bi- or tridimensional coordinates depending on the implementation, and their relative positions in the face carry information from relative distances of facial structures, which can be useful for recognizing faces or analyzing emotions. Regarding facial expressions, muscular movements change in the relative positions of face structures and the relative distances between landmarks will reflect the facial features [Vaillancourt, 2010]. For this reason, relative distances between landmarks are valuable for objectively analyzing facial expressions and emotions or even for recognizing faces. In both cases, the landmark coordinates serve as reference values for calculating distances by using various metrics such as Euclidean, Manhattan, or Minkowski distances [Garrido and Joshi, 2018; Li and Deng, 2018; Mehta *et al.*, 2018]. Nonetheless, implementations may greatly differ.

To make landmarks comparable across all faces, researchers have proposed using fixed common distances, including the distance between the eyes, to normalize facial landmarks in images with different scales [Ismail and Sabri, 2009]. However, this normalization approach cannot effectively reduce variability related to image acquisition conditions, particularly rotation and racial variations. Certain individuals may exhibit prominent facial characteristics that can negatively influence how ML algorithms perform. For instance, the width of the mouth can be critical when identifying the emotion of happiness, which is often manifested through a smile. In such scenarios, a classifier trained on the character Joker, the villain of the Batman superhero series, would have difficulty converging and distinguishing a person's expression of happiness when the Euclidean distance metric is used to normalize landmarks.

Figure 3 illustrates the localization of 68 facial landmarks, which aim to identify various structural components of the human face, including the face contour, eyebrows, eyes, nose, and mouth, among others. With respect to the number of detected landmarks, the range between 60 and 80 coordinates predominates and appears in nearly half of the studies [Testa *et al.*, 2019].

After facial landmarks are extracted, Image Processing is completed. The processed data transitions from the pixel space of the image to the bidimensional (x and y) coordinates representing the positions of the facial landmarks, which reduces the dimensionality of the problem being studied. Mathematically, rotation is a linear transformation that involves spatial coordinates, in this case, in two dimensions, while the magnitude of vector lengths and orientation in the physical space are preserved [Winterle, 2014]. Therefore, using coordinates tends to introduce less variability than using interlandmark distances while avoiding the additional processing time required to compute distance metrics.

For various types of objects, including human faces, frontalization techniques allow frontal views to be artificially synthesized from rotated original sources. Using these techniques can improve how classification and recognition systems relying on images perform. Indeed, frontalization normalizes any variations occurring during the acquisition [Hassner *et al.*, 2015]. Additionally, training and testing employed by ML algorithms can be performed with standardized samples in a consistent position. AS stated by [Vonikakis and Winkler, 2020], image frontalization techniques can be classified into appearance-based frontalization, where the original image is rendered to the frontal view, and coordinate-based frontalization, which uses a model trained with images of the object under study.

3 Related Work

Methods for recognizing emotions encompass various approaches aimed at identifying and categorizing human emotions on the basis of observable cues. Unlike REFEL, these methods do not usually explore the set of techniques regarding manipulation of landmarks, frontalization and normalization of coordinates sequentially. Because we hypothesized that reducing variability would lead to classifiers that perform better, regardless of the ML method, our related works have not focused on studies manipulating big data and deep learning techniques, which contrasts with the works of [Agarwal and Susan, 2023; Nhu *et al.*, 2023; Zhang *et al.*, 2023; Gupta *et al.*, 2024]. However, we consider that REFEL can preprocess input data for them.

Testa et al. reviewed literature studies that discovered that facial expressions can be synthesized through landmark extraction [Testa *et al.*, 2019]. The study predominantly detects eyes, eyebrows, mouth, face contour, and nose. The authors observed that ML approaches have been the most used, and that few studies have applied metrics to evaluate the results. Like these studies, our experiments measure accuracy, but it also considers time of response as a fundamental requirement for recognizing emotions in almost real-time.

Álvarez, Luengo, and Lawrence presented a method that uses the Euclidean distance between landmarks in the eye and mouth regions to detect smiles [Alvarez *et al.*, 2013]. The authors employed relative error distribution to validate accuracy. The final average accuracy and standard deviation were respectively 94.53% ± 2.47%, obtained by using just the LabeledFacesInTheWild (LFW) dataset.

Cui developed a method for recognizing smiles in images containing people. The authors used the Euclidean distance as an attribute to train an algorithm called Extreme Learning Machine [Cui *et al.*, 2018], to achieve 93.40% accuracy on the Genki4k image dataset.

Salman, Madani, and Kissi proposed a method for recognizing facial expressions to train a decision tree called Classification and Regression Tree. This tree uses the measurements of the distances between the width and height of the mouth as the feature vector [Salman *et al.*, 2016].

Hassner et al. explored a simpler approach by using an unmodified 3D reference surface, which approximates the shape of all input faces [Hassner *et al.*, 2015]. This allowed a direct, efficient, and easily frontalization method to be developed. Importantly, this method generates aesthetically pleasing frontal views and proved surprisingly effective for recognizing faces and estimating gender. In the LFW dataset, 97.5% of the 1432 images were successfully frontalized.

Jia et al. published a review article that presents spontaneous and posed facial expression databases and various computer vision-based detection methods, including methods specific for detecting smiles [Jia *et al.*, 2021]. These authors highlighted that the generalization ability is important for detecting emotions and specific emotions on the basis of unique facial features.

By considering multimodal input, Bohyet al. developed a deep learning-based multimodal smile and laugh classification system that uses audio and vision-based models as well as a fusion approach [Bohy *et al.*, 2022]

Through tests with linear and nonlinear classifiers, Guyon and Elisseeff demonstrated that selecting features reduces the variability of the studied problem [Guyon and Elisseeff, 2003]. This happens because certain classifiers struggle with duplicate data, so accuracy tends to increase.

Kwon et al. developed an algorithm that considers variations in pose, occlusion, lighting, and skin tones in uncontrolled settings for recognizing facial emotions. These authors used 3D facial landmarks, normalization and ML in a non-cited and unavailable dataset (with 2556 face tessellations), to achieve 73.7% accuracy [Kwon *et al.*, 2023]. In the same context, REFEL achieved 90% accuracy by considering well-known datasets.

Poulose et al. recognized facial emotion by using facial landmarks to classify drivers' emotion. They reached between 95% and 99% accuracy by using a Deep Learning Architecture [Poulose *et al.*, 2021].

Zhang et al. and Nhu et al. also explored a convolutional network to detect emotion [Nhu *et al.*, 2023]). Specifically, Zhang et al. grasped the natural spatial relationships among geometric facial landmarks, but they reached 30-50% accuracy [Zhang *et al.*, 2023].

Hossain et al. used an expanded CK dataset (see above) and employed ML for recognizing facial emotions on the basis of landmark points and blocks [Hossain *et al.*, 2023]. In a challenging context, Agarwal et al. investigated the use of neural networks to recognize emotions from masked faces during the COVID-19 pandemic.

Choosing a method for recognizing emotions depends on several factors, including mainly the data available for processing, application context (mobile, desktop, controlled or uncontrolled environments) or scenario (quality of image, scale, rotation, brightness, luminosity, etc), and desired performance. When we used different datasets under various conditions, the accuracy achieved with REFEL resembled the accuracy obtained with similar methods that considered specific datasets. REFEL showed excellent response time, which is a fundamental requirement for real-time emotion recognition in mobile applications, for instance. This happened because normalized landmarks are generalized two-dimensional coordinates, and allow for low-cost and real-time processing, which is particularly suited to devices with limited processing capabilities, such as cell phones.

4 Recognizing Emotions through Facial Expression and Landmark normalization

REFEL aims to recognize facial expressions and human emotions depicted in digital images. Next, we present collections of digital images (material) and this method.

4.1 Digital Images Datasets

A facial expression dataset is a collection of digital images or video clips featuring different people acting or expressing natural emotions. Its content is essential for training, testing, and validating Machine Learning (ML) algorithms, as well as for developing facial recognition and facial expression recognition methods or systems, encompassing emotions. To experiment REFEL, we selected different datasets aiming to augment the variability of noises in the manipulated images. The theoretical basis of human emotions often guided datasets of digital images for recognizing emotions [Alvarez *et al.*, 2013]. Six types of facial expressions are assumed to exist: happiness, fear, disgust, anger, surprise, and sadness, in addition to the neutral (or indifferent) expression. We selected only datasets with balanced data and this theoretical foundation (six facial expressions and a neutral expression). The following facial expression datasets are freely available and were used to experiment REFEL:

- *The Cohn-Kanade Dataset (CK+)* consists of 593 sequences obtained from 123 subjects. Each sequence captured images from the onset (neutral frame) to the peak expression (final frame) of people of various genders and heritage, aged from 18 to 50 years [Cohn and Kanade, 2010]. The peak frame is reliably FACS coded [Ekman and Friesen, 1971] for facial action.
- *The Japanese Female Facial Expression (JAFFE)* comprises 210 images presented by ten Japanese women [Lyons *et al.*, 2017] in a frontal position.
- *Radboud Faces Database (RafD)* consists of 1407 images of 67 actors (Caucasian men and women, European Caucasian children, and Moroccan men) and comprises three different rotations of the face and seven emotions [Langner *et al.*, 2010]. Only RafD provides an additional emotion labeled as contempt, which was not used.
- *The Karolinska Directed Emotional Faces (KDEF)* has 4,900 images of 70 actors [Goeleven *et al.*, 2008]. Each expression is captured from five angles in Sweden.
- *The Nimstim set of Facial Expressions* contains 672 frontal images of 43 professional actors, including 18 women and 25 men, aged between 21 and 30 years [Tottenham *et al.*, 2009].

The combined images of all the cited datasets form the material on which we tested REFEL. We aimed to gather images depicting emotions while accounting for variations in age, ethnicity, gender, spontaneity *versus* acted expressions, capture equipment settings, focal length, luminosity, framing, people's pose (rotated face image and sentiment expression) and image quality.

4.2 The REFEL Method

REFEL extracts relative positional data of facial structures and calculates a more accurate measure of facial muscle movements, such as the movements produced by facial expressions. Figure 2 shows the steps of the method. Each of the following subsections describe the various steps of the method. The last subsection presents an overall algorithm that represents a systematic and efficient way to accomplish and to visualize the steps of REFEL.

4.2.1 Capture of Images

The two initial steps focus on capturing and preprocessing tasks for detecting faces. First, REFEL establishes communication with the datasets to input the color digital image (see Figure 2, Step 1).

4.2.2 Gray Scale Conversion

Next, REFEL reduces the three color channels to a single channel. This simplification aims to streamline digital image processing, to lower computational demands and to facilitate feature learning (see Figure 2 - Step 2).

4.2.3 Facial Detection

To reduce the search area, REFEL makes a bounding box of the Region of Interest (ROI) of each face using Histogram of Oriented Gradients (HOG). It limits computations to the most relevant areas for recognizing emotions, which enhances processing efficiency (see Figure 2 - Step 3).

4.2.4 Landmark Extraction

After focusing the facial area, REFEL outlines the regions of eyes, nose and mouth as ROI in each image. Pixels must be changed for the two-dimensional coordinates of points of reference called landmarks, as mentioned previously (see Figure 2 - Step 4). In this step, REFEL extracts landmarks to change data dimensionality. These landmarks are the REFEL parameters and they include the coordinates of the eyes, eyebrows, nose, mouth and the face contour. Figure 3 shows the numbering of 68 landmarks of facial images.

After pixels are changed for landmarks, the image domain comprises only 136 variables, namely 68 coordinates for the x-axis and 68 coordinates for the y-axis, which reduces data dimensionality. Furthermore, to minimize the effects of scale variability of the coordinates, REFEL executes the next step.

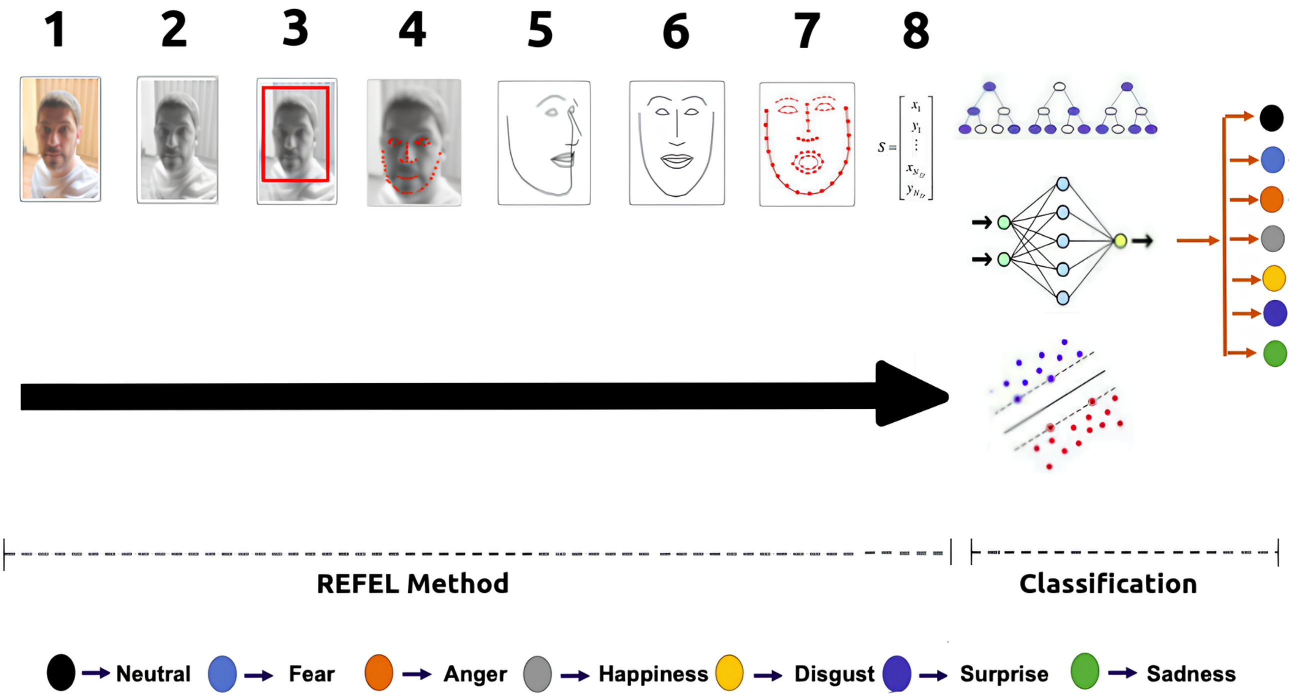


Figure 2. REFEL: Recognizing Emotions through Facial Expression and Landmark normalization. Steps: 1 - Capture of the original image (raw data); 2 - Image processing to gray-scale image; 3 - Pretrained model recognizes the face in the image; 4 - Landmarks are set; 5 - Another pretrained model transforms the 2D landmarks into 3D landmarks; 6 - The 3D landmarks are rotated to the frontal position; 7 - A new round of face recognition and landmarks setting is performed; 8 - Coordinates of the evaluated expression are compared with coordinates of the neutral face of the same person, and the delta divergences are used to create a data vector and to train classifiers.

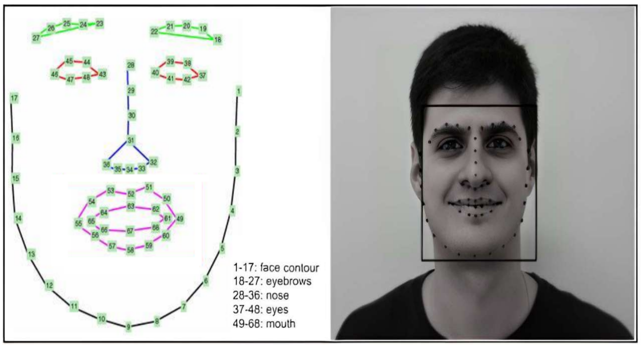


Figure 3. Numbering of the 68 facial landmarks. Facial photo adapted from https://fei.edu.br/ cet/facedatabase.html, accessed on Apr/1st/2025.

.

4.2.5 Min-max Normalization of Coordinates

REFEL applies Min-Max (from 0 to 1) normalization to the coordinates, to reduce scale factors that account for variations in proximity between the person and the capture device. The coordinates are normalized by using the minimum and maximum (min-max) values for each axis, as described in Equation 1 for the horizontal coordinates on the x axis and Equation 2 for the vertical coordinates on the y axis. This normalization process is defined as follows:

$$\widehat{x}_i = \frac{x_i - min(x)}{max(x) - min(x)} \tag{1}$$

$$\widehat{y}_i = \frac{y_i - min(y)}{max(y) - min(y)} \tag{2}$$

Thus, all the x and y coordinates are scaled to values between zero and one,which minimizes the scale effect. x and y coordinates assume values between zero and one, minimizing the scale effect. This transformation does not alter the proportions between the coordinates. Therefore, the rescaled face remains similar to the original one and can be used for facial recognition purposes (see Figure 2 - Step 5). Hence, the method presented here requires prior knowledge of the actor in the image.

This step innovates emotion recognition with a constant O(1) cost and maintains proportionality between coordinates. The actor's neutral image is needed for reference. Before frontalization (next step) is performed, variations in facial expressions must be recognized.

4.2.6 Frontalization

Frontalization aims to reduce variations in facial rotation, both horizontally and vertically, with respect to the image plane, by centering the pose in a frontal manner. The geometric variability caused by various rotations in both the x-axis and y-axis are reduced. The coordinates are adjusted to simulate a frontal position (see Figure 2 - Step 6). Before frontalization, REFEL recognized the variations in facial expressions, as previously mentioned.

4.2.7 Normalization by neutral Actor's Face (Delta standard)

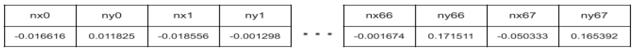
This step eliminates anatomical variations. First, REFEL method first needs to recognize variations in facial expressions because anatomical differences among different actors in the images can be a significant source of noise. Consequently, detecting morpho-geometric patterns of the face can be challenging when the aim is to recognize emotions. Bearing this in mind, we propose evaluating the variation in patterns when a facial expression is performed. In contrast to what is typically done in facial expression detection [Revina and Emmanuel, 2018; Cui *et al.*, 2018; Mohan *et al.*, 2021], our approach requires assessing a neutral facial pattern instead of evaluating static morpho-geometric patterns. Specifically, we propose that a frontal image A of the resting face (neutral facial expression), as reference, be submitted to facial detection, landmark extraction, coordinate normalization, and frontalization, to generate 136 reference coordinates to form a vector $\overrightarrow{A}$.

A second image B, the object of evaluation, undergoes the same process, to generate 136 coordinates to form vector $\overrightarrow{B}$. This leads the final information vector $\overrightarrow{\Delta}_{AB}$ to be created. $\overrightarrow{\Delta}_{AB}$ consists of 136 variables, where each coordinate i is obtained as $\Delta_{AB}i = Ai - Bi$. Therefore, $\overrightarrow{\Delta}_{AB}$ contains the information about the relative variation of the frontalized coordinates of B with respect to the coordinates of the resting face A (see Figure 2 - Step 7).

Given that the expression of interest in B is always compared to the expression of the same actor in a neutral position, previously labeled and known in A, problems of anatomical variability are minimized. In this way, a static geometric pattern—such as the one characterizing a smile—can be classified, and the deformation resulting from this expression can be quantitatively measured (see Steps 6 and 7 in Figure 2).

4.2.8 Creation of the Coordinate Vector

To concatenate the coordinates into a vector and aiming to facilitate induction of ML algorithms, REFEL creates a final feature vector with all the coordinates for each image (see Figure 2 - Step 8). Thus, REFEL generates feature vectors based on image coordinates without altering the original images. The final dimension of the vector is 1 row with 136 columns (68 columns for the *x* coordinate and 68 columns for the *y* coordinate) for each actor, as presented in Figure . One hundred and thirty-six positions were normalized by coordinates and frontalized, and the neutral face was subtracted for all of them. The vector in Figure corresponds to the emotion fear. The coordinates are presented in sequence, meaning $x0$ and $y0$, so that the frontalization algorithm can be executed.



nx0	ny0	nx1	ny1		nx66	ny66	nx67	ny67
-0.016616	0.011825	-0.018556	-0.001298	. . .	-0.001674	0.171511	-0.050333	0.165392

Figure 4. An example of part of a feature vector corresponding to the emotion fear.

All the steps precede the construction of ML-based algorithms, i.e., the steps pertain to data preprocessing aimed at better standardizing input information in a feature vector. The execution of the machine learning algorithm is detailed in the Experiments Section.

4.2.9 An Overall Algorithm

In summary, Algorithm 1 outlines the technical steps for recognizing facial emotions within REFEL. The input comprises images obtained from facial expression databases, including CK+, RafD, NimStim, KDEF, and Jaffe. The output is a reusable Machine Learning Model (MLM) tailored for recognizing emotions. First, REFEL loads the images and focuses on preprocessing tasks that are essential for detecting faces. These tasks are executed concurrently with image loading. Then, there are three main grouped phases follow: (1) extracting and normalizing reference coordinates, (2) frontalizing and normalizing, and (3) extracting measurements of the relative position of the facial expression in the image relative to the actor's neutral face.

The algorithm starts (line 1) and iterates over all input images G (line 2) and for each image (line 3) and initiates image preprocessing (line 3) and facial detection (line 4). If successful (line 5), the processing for reducing data variability sequentially runs landmark extraction (line 6), normalization of *x* coordinates by using min-max normalization (line 7), normalization of *y* coordinates by using min-max normalization (line 8), frontalization of coordinates (line 9), and normalization by the actor's face (delta pattern) (line 10). Finally, the coordinates are inserted into the resulting feature vector (line 11). After all the images are processed, the final vector undergoes the classification process and the creation of the MLM (line 14). Several types of ML classifiers exist, and each has their own structures, rules, processes, and reasoning. REFEL exploits (i) the connectionist Perceptron that identifies nonlinear relationships between input and output data by considering a single input and output layer (without hidden layers); (ii) the instance-based K-Nearest Neighbor (KNN), which assigns labels on the basis of the closest k-nearest neighbors, often by using Euclidean distance; (iii) the Decision Tree (D3), which creates classification rules by splitting attributes to maximize performance gain; (iv) the ensemble Random Forest (RFC), which builds multiple decision trees and aggregates their predictions, to improve generalization and to reduce overfitting; (v) the neural network Multilayer Perceptron (MLP) with an input layer, one or more hidden layers, and an output layer (MLP can solve nonlinear problems through activation functions); and (vi) the supervised learning Support Vector Machine (SVM), which is used for classification, regression, and detection of outliers. The configuration of our overall algorithm is presented next.

5 Experiments and Results

This section presents the experiments we conducted by considering the presented datasets, the obtained results, and the discussions regarding REFEL processing.

5.1 Training and Evaluation

We evaluated the proposed REFEL method by using the mentioned datasets (see Section 4.1). In its sequence of steps and after performing all normalization steps, REFEL created a feature vector comprising a set of 136 coordinates, which served as input for the SVM, MLP, RFC, Extra Trees, and D3 ML algorithms. These algorithms assess the quality of the extracted features and the classification of emotions. D3, RFC, Extra Trees MLP and SVM classified the resulting REFEL

Algorithm 1 - REFEL

```
  Input....: Datasets (G)
  Output...: ML Models (MLM)
BEGIN
WHILE G has images
    Load Image G(I) and preprocessing
    face ← face detection G(I)
    IF face $\neq \varnothing$
        coord ← 68 $Landmarks$ Extraction
        coord ← Norm. coordinates x
        coord ← Norm. coordinates y
        coord ← Coord Frontalization
        coord ← Normalization ($\Delta$)
        coordArray ← coord
    END-IF
END-WHILE
MLM ← Classifier(coordArray)
RETURN MLM
END
```

vector by considering the following configuration: (i) D3, RFC and Extra Trees set the maximum depth of the nodes used to 4, while RFC and Extra Trees considered this depth value along with a tree count parameter set to 1000; (ii) MLP algorithm had the number of iterations set to 3000 epochs, a learning rate of 0.001, and the Adam algorithm to activate the neural network; and (iii) SVM employed a radial basis kernel to separate categories, with a penalty parameter of $C = 5$ to penalize incorrect classifications.

To present the results, we employed graphs depicting the percentages of accuracy and error were employed. Figure 5 showcases the results regarding emotion recognition through the sequential execution of the REFEL steps. The INPUT consisted of the raw data, such as each unprocessed image obtained from the dataset, while the OUTPUT wss the emotion classified after each REFEL step. The succession of coordinate normalization techniques played a crucial role in improving the accuracy of the results, irrespective of the algorithm employed for classification. Thus, we assume that ML algorithms took longer to detect patterns in data when normalization was not applied, particularly if the scale factor (such as the proximity to the image acquisition equipment) remained unnormalized.

According to Tables 1 and 2, we achieved the best OUTPUT performance of REFEL by using SVM, to obtain 87.1% accuracy and processing time of only 4.33 seconds (s). The experiments used 10-fold cross-validation during training. Emotions are suggested to be also universally recognizable by machines, and their recognition can be effectively performed artificially.

Table 1. Accuracy of each ML algorithm used to recognize emotions by the REFEL method.

Algorithm	INPUT	Norm.	Front.	OUTPUT
SVM	0.596	0.675	0.826	**0.871**
MLP	0.446	0.781	0.837	0.861
ExtraTrees	0.636	0.697	0.818	0.860
RFC	0.652	0.707	0.819	0.860
D3	0.463	0.578	0.743	0.794

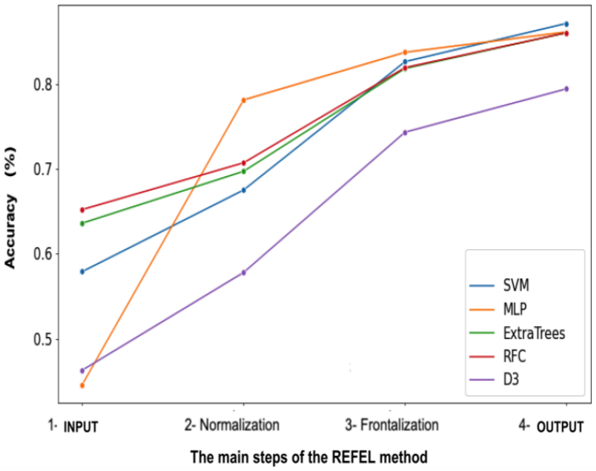


Figure 5. Accuracy evolution for recognizing emotions by using the REFEL method.

Another important aspect to analyze regarding REFEL is the processing time of the algorithms in relation to the chaining of normalization steps. According to Figure 6, the data were consistent and demonstrated that the REFEL steps decreased the processing time of all the experimented algorithms. On the basis of Table 2, SVM and D3 provided the best constant performance. Therefore, the reduced sample variability in the coordinates, provided by REFEL with min-max normalization, frontalization, and normalization by actor's face, contributes to the good performance of ML algorithms responsible for recognizing human emotion as well as optimizes the processing time required in each step. These results proved our hypothesis.

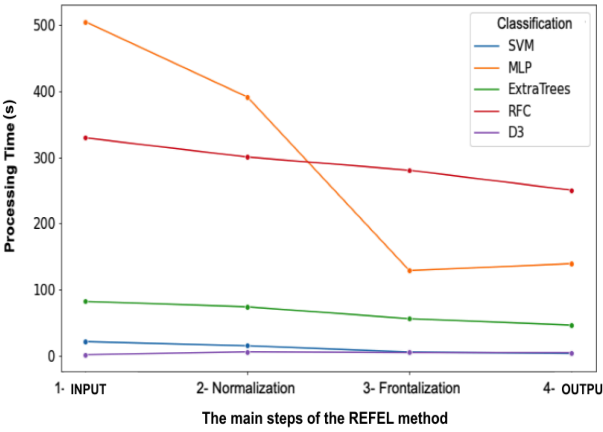


Figure 6. Processing time evolution for recognizing emotions by using the REFEL Method.

Table 2. Processing time (seconds) of each ML algorithm used to recognize emotions by the REFEL method.

Algorithm	INPUT	Norm.	Front.	OUTPUT
SVM	22.11	15.82	6.37	**4.33**
MLP	504.11	390.61	129.04	139.72
ExtraTrees	82.58	74.53	56.64	46.96
RFC	329.19	300.29	280.39	250.33
D3	6.79	5.66	5.58	2.46

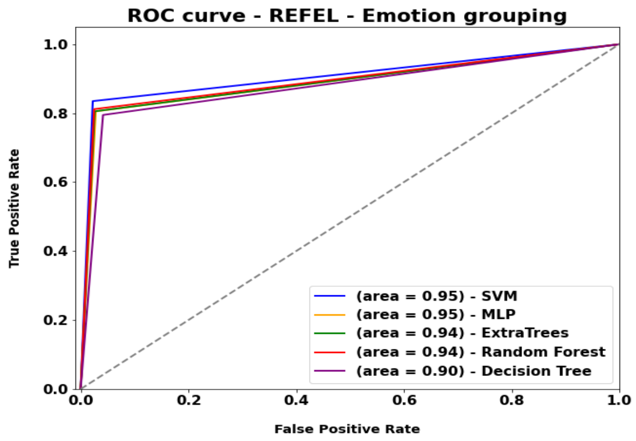


Figure 7. ROC curve for recognizing emotions by using the REFEL and considering surprise with fear and disgust with anger.

According to Ekman's work [Ekman and Friesen, 1971], brightness between the emotions surprise and fear is strong, even when the latter is an extension of the former. By considering this and applications that demand fewer than seven universal emotions, we evaluated REFEL by grouping surprise and fear and disgust with anger.

The ROC (Receiver Operating Characteristic) curves in Figure 7 were created on the basis of the true positive (TP) and false positive (FP) rate so that the discriminative capability of the classifiers, i.e., the ability of the classifier to classify specific emotions correctly, would be represented. In the ROC graphs, points closer to (0,1) indicate higher TP and lower FP and represent a more consistent and generalizable classification. Thus, Figure 7 presents a ROC graph depicting comparative analysis of the performance of the various ML algorithms we used to recognize emotions after implementing all the variability reduction steps of REFEL and grouping the emotions. The dashed diagonal line represents how a random classifier performs and serves as a reference for evaluating the others. Points above the diagonal in the ROC space represent better classification than points below the diagonal. By analyzing Figure 7, we can see that the ROC curves of the SVM and MLP yielded more accurate results than the others. In conclusion, the average performance of SVM was the best (over 93%) for the evaluated measures when we grouped fear with surprise and anger with disgust.

With SVM selected as the classifier for REFEL, Table 3 shows how it performed in terms of accuracy, sensitivity, and F-measure for the universal emotions studied herein. REFEL easily identified happiness with accuracy rates above 95%, even in experiments that used the raw coordinates from the initial input image. Therefore, REFEL can potentially detect smile automatically in digital images, which confirms previous experiments published in [Persona *et al.*, 2023]. Conversely, REFEL provided the lowest accuracy and performance rates for sadness regardless of the employed techniques and after the REFEL normalization steps were carried out. The table also shows the result for SVM concerning the grouped emotion. The two groups had performance rates above 92%.

Without grouping emotions, Figure 8 shows graph plotted for SVM and MLP produced more accurate results. The performance of the classifiers varied from 0.87 to 0.92 in TP (True Positives) and 0.1 in FP (False Positives), which demonstrated good discrimination. Finally, the average performance of SVM was the best for all the measures evaluated in the different experiments (over 90%).

Table 3. The REFEL method with the SVM classifier.

Emotion	Acc	Sensit.	F-measure
Neutral	0.996	0.999	0.998
Happiness	0.954	0.964	0.959
Sadness	0.861	0.835	0.848
Fear-Surprise	0.932	0.921	0.926
Anger-Disgust	0.927	0.945	0.936

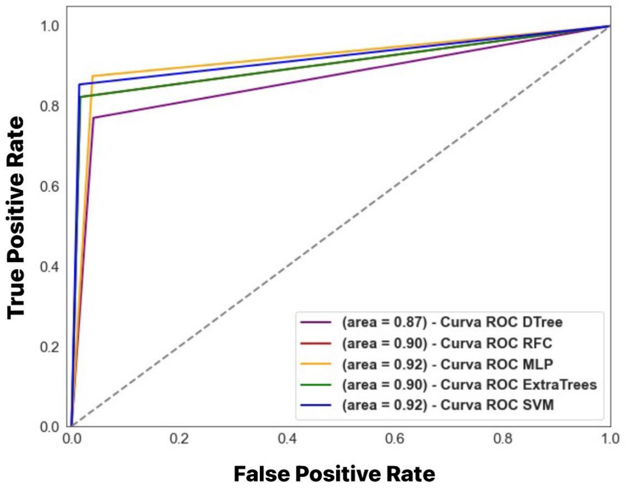


Figure 8. ROC curve for recognizing emotions by using the REFEL method without grouping emotions.

We have not compared REFEL against our previous work described in [Persona, 2022], which only manipulated smiles. However, we compared our classification of emotions without grouping emotions to [Macedo *et al.*, 2024a] and by considering two groups of emotions: fear with surprise and anger with disgust. Comparisons of the two ROC curves in Figures 7 and 8 showed that the grouping emotions improved the average performance of REFEL. In the literature, we did not find a baseline as well to assess the results mainly considering response time, resulting in a lack of comparison to other state-of-the-art systems. Remarkably, REFEL exhibits a real-time performance accuracy exceeding 90%.

6 Final Remarks

Emotions provide our first means of nonverbal communication, which is developed throughout our lives. Through emotions, humans can interact with each other and the environment where they live. This interaction is possible because humans almost always translate their emotions into detectable physical movements, such as facial expressions, which is essential for social interactions.

Although recognizing emotions is trivial for human beings, it is a challenging task for machines and computers.

Thus, we have developed an artificial emotion recognition method called REFEL to extract and to analyze the morphometric characteristics of the facial region. REFEL combines various techniques of digital image processing and statistical methods to reduce variability in the employed images. Among these techniques, normalizing coordinates by using min-max is noteworthy because it can optimize the effects of scale factor, frontalization (which reduces the effects caused by facial rotation), and delta normalization (which uses the actor's neutral face to identify other emotions), to minimize the effects of anatomical and racial variations.

Our results demonstrated that REFEL effectively classifies and artificially recognizes human emotions, achieving over 93% accuracy for grouped emotions and 90% accuracy for non-grouped emotions, both across various facial expressions from diverse databases. Regarding smile detection only, we achieved a final accuracy close to 95%, which is superior to accuracy reported in the literature. REFEL also surpassed the time processing of methods evaluating it. Here, we have not exploited quantitative aspects of movements because our focus was classification. However, REFEL can measure the intensity and speed of movements. We have used it in the SofiaFala Project[2] [Macedo *et al.*, 2024b].

The main contribution of this work is our technique for normalizing facial landmark coordinates - the technique reduces the scale factor effect on images. The main limitation of REFEL is associated with the fact that REFEL is tailored for contexts which the identity of the actor is assured. However, this limitation does not pose a problem for applications such as routine use in detecting driver's emotion or assistive technology. The acquisition of landmarks can also be a limitation, but it works effectively up to a rotation of 45 degrees (left and right). Experiments considering uncontrolled environments would demand coupling a routine of face recognition. Moreover, this work will be continued in terms of:

- analyzing the performance of REFEL with a greater number of facial landmark coordinates and a fewer number of landmarks. For example, face contour coordinates are exploited by REFEL, but recent tests revealed that removing these coordinates produced nearly identical results. Although, these parameters are computationally lightweight, future experiments should explore the potential of excluding the face contour coordinates from processing to further simplify the method without compromising accuracy.
- analyzing REFEL with three-dimensional coordinates of facial landmarks.
- using other databases to evaluate the performance of REFEL, preferably databases with greater representation of Black and Asian actors. Facial geometry plays a critical role in extracting landmarks effectively, making it important to ensure a wide variety of facial structures in future datasets.
- assessing how REFEL performs in video-based applications.
- investigating facial recognition with coordinates normalized by min-max and frontalization.
- using the quantitative responses of REFEL, especially considering fear and surprise.

Declarations

Acknowledgements

This research was developed with the help of the HPC resources provided by the Information Technology Superintendence of the University of São Paulo.

We would like to thank the Brazilian Postal and Telegraph Company and the assistance of HPC resources provided by the Information Technology Superintendence of the University of São Paulo.

Funding

The presented method was discussed in the SofiaFala project supported by CNPq (Brazil), through the process 442533/2016-0, which supports the development of an application for mobile devices to assist children with Down Syndrome, to support training speech therapy for speech development.

Authors' Contributions

LP contributed to conceptualization, data curation, formal analysis, investigation, methodology, software and writing – original draft. FM contributed to conceptualization, validation, writing, review. AAM contributed to conceptualization, validation, writing, review & editing. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

The datasets and/or software generated and/or analyzed during the current study will be made upon request.

References

Agarwal, A. and Susan, S. (2023). Emotion recognition from masked faces using inception-v3. In *2023 5th International Conference on Recent Advances in Information Technology (RAIT)*, pages 1–6. DOI: 10.1109/RAIT57693.2023.10126777.

Alvarez, M., Luengo, D., and Lawrence, N. (2013). Linear latent force models using gaussian processes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(11):2693–2705. DOI: 10.1109/TPAMI.2013.86.

Bohy, H., El Haddad, K., and Dutoit, T. (2022). A new perspective on smiling and laughter detection: Intensity levels matter. In *2022 10th International Conference on Affective Computing and Intelligent Interaction (ACII)*, pages 1–8. DOI: 10.1109/ACII55700.2022.9953896.

Chaugule, V., Abhishek, D., Vijayakumar, A., Ramteke, P. B., and Koolagudi, S. G. (2016). Product review based on optimized facial expression detection. In *2016 Ninth International Conference on Contemporary*

[2] https://dcm.ffclrp.usp.br/sofiafala/ or https://sites.usp.br/sofiafala/

Computing (IC3), pages 1–6, Noida, India. IEEE. DOI: 10.1109/IC3.2016.7880213.

Cheng, Y. and Ling, S. (2008). 3d animated facial expression and autism in taiwan. In *IEEE International Conference on Advanced Learning Technologies (ICALT 2008)*, pages 17–19, Los Alamitos, CA, USA. IEEE Computer Society. DOI: 10.1109/ICALT.2008.220.

Chollet, F. (2017). *Deep Learning with Python*. Manning Publications Co., Greenwich, CT, USA, 1st edition. Book.

Cohn, J. and Kanade, T. (2010). The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. pages 94–101. IEEE Computer Society Conference on Computer Vision and Pattern Recognition. DOI: 10.1109/CVPRW.2010.5543262.

Cui, D., Huang, G.-B., and Liu, T. (2018). Elm based smile detection using distance vector. *Pattern Recognition*, 79:356–369. DOI: 10.1016/j.patcog.2018.02.019.

Dalal, D. and Triggs, B. (2005). Histograms of oriented gradients for human detection. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 1, pages 886–893 vol. 1, San Diego, CA, USA. IEEE. DOI: 10.1109/CVPR.2005.177.

Darwin, C. (2013). *The Expression of the Emotions in Man and Animals*. Cambridge Library Collection - Darwin, Evolution and Genetics. Cambridge University Press, England. DOI: 10.1017/CBO9781139833813.

Ekman, P. and Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, 17(2):124–129. DOI: 10.1037/h0030377.

Filho, Oge Marques; Vieira Neto, H. (1999). *Processamento Digital de Imagens*. Brasport, Brasil. Available at:`http://projetoaprendizagemgrupo4.pbworks.com/w/file/fetch/96395952/Processamento%20Digital%20de%20Imagens.pdf`.

Garrido, G. and Joshi, P. (2018). *OpenCV 3.X with Python By Example: Make the most of OpenCV and Python to build applications for object recognition and augmented reality*. Packt Publishing, US, 2nd edition. Book.

Ghorbani, G; Targhi, A. T. and Dehshibi, M. (2015). Hog and lbp: Towards a robust face recognition system. In *2015 Tenth International Conference on Digital Information Management (ICDIM)*, pages 138–141, Jeju, South Korea. IEEE. DOI: 10.1109/ICDIM.2015.7381860.

Goeleven, E., Raedt, R. D., Leyman, L., and Verschuere, B. (2008). The karolinska directed emotional faces: A validation study. *Cognition and Emotion*, 22(6):1094–1118. DOI: 10.1080/02699930701626582.

Gonzales, R. C. and Woods, R. E. (2008). *Digital Image Processing*. Pearson, New Jersey, US, 3rd edition. Book.

Gupta, B. B., Gaurav, A., Chui, K. T., and Arya, V. (2024). Deep learning-based facial emotion detection in the metaverse. In *2024 IEEE International Conference on Consumer Electronics (ICCE)*, pages 1–6. DOI: 10.1109/ICCE59016.2024.10444217.

Guyon, I. and Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3:1157–1182. Available at:`https://www.jmlr.org/papers/v3/guyon03a.html`.

Hassner, T., Harel, S.and Paz, E., and Enbar, R. (2015). Effective face frontalization in unconstrained images. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4295–4304. DOI: 10.1109/CVPR.2015.7299058.

He, J., Wen, X., and Zhou, J. (2023). Advances and application of facial expression and learning emotion recognition in classroom. In *Proceedings of the 2023 6th International Conference on Image and Graphics Processing*, ICIGP '23, page 23–30, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3582649.3582670.

Hess, U. (2001). The communication of emotion. In *Emotions, Qualia and Consciousness*, pages 397–409. Singapore. DOI: 10.1142/9789812810687$_0$031.

Hossain, M. A., Osman, M. H., Hamdan, A. A., Abdelhag, M. E., and Kechadi, M. T. (2023). Ferlp: Facial emotion recognition based on landmark points using artificial intelligence and machine learning. In *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, pages 1–6. DOI: 10.1109/ICCCNT56998.2023.10308392.

Ismail, N. and Sabri, M. I. M. (2009). Review of existing algorithms for face detection and recognition. In *Proceedings of the 8th WSEAS International Conference on Computational Intelligence, Man-Machine Systems and Cybernetics*, page 30–39, Stevens Point, Wisconsin, USA. World Scientific and Engineering Academy and Society (WSEAS). Available at`https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=066aefa23a69c8066c4337153ab97a0b48c58296`.

Jia, S., Wang, S., Hu, C., Webster, P., and Li, X. (2021). Detection of genuine and posed facial expressions of emotion: Databases and methods. *Front. Psychol. - Sec. Perception Science*, 11. DOI: 10.3389/fpsyg.2020.580287.

Kim, W. (2022). Low-light image enhancement: A comparative review and prospects. *IEEE Access*, 10:84535–84557. DOI: 10.1109/ACCESS.2022.3197629.

Koohestani, F., Karimi, N., and Samavi, S. (2024). Revealing shadows: Low-light image enhancement using self-calibrated illumination. In *2024 32nd International Conference on Electrical Engineering (ICEE)*, pages 1–7. DOI: 10.1109/ICEE63041.2024.10667748.

Kwon, J., Oh, K. T., Kim, J., Kwon, O., Kang, H. C., and Yoo, S. K. (2023). Facial emotion recognition using landmark coordinate features. In *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 4916–4918. DOI: 10.1109/BIBM58861.2023.10385536.

Lahiri, U., Bekele, E., Dohrmann, E., Warren, Z., and Sarkar, N. (2011). Design of a virtual reality based adaptive response technology for children with autism spectrum disorder. In *Affective Computing and Intelligent Interaction,Springer*, pages 165–174. DOI: 10.1007/978-3-642-24600-5$_2$0.

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., and van Knippenberg, A. (2010). Presentation and validation of the radboud faces database. *Cognition and Emotion*, 24(8):1377–1388. DOI: 10.1080/02699930903485076.

Li, K., Xu, F., Wang, J., Dai, Q., and Liu, Y. (2012). A data-driven approach for facial expression synthesis in video. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 57–64, Providence, RI, US. IEEE. DOI: 10.1109/CVPR.2012.6247658.

Li, S. and Deng, W. (2018). Deep facial expression recognition: A survey. *Computing Research Repository (CoRR)*, abs/1804.08348. DOI: 10.1109/TAFFC.2020.2981446.

Lyons, M., Kamachi, M., and Gyoba, J. (2017). Japanese Female Facial Expression (JAFFE) Database. DOI: 10.6084/m9.figshare.5245003.v2.

Macedo, A., Persona, L., and Meloni, F. (2024a). Recognition of emotions through facial geometry with normalized landmarks. In *Proceedings of the 30th Brazilian Symposium on Multimedia and the Web*, pages 257–266, Porto Alegre, RS, Brasil. SBC. DOI: 10.5753/webmedia.2024.243252.

Macedo, A. A., de Souza Gonçalves, V., Mandrá, P. P., Motti, V., Bulcão-Neto, R. F., and da Hora Rodrigues, K. R. (2024b). A mobile application and system architecture for online speech training in portuguese: design, development, and evaluation of sofiafala. DOI: https://doi.org/10.1007/s11042-024-19980-5.

Mehta, D., Siddiqui, M. F. H., and Javaid, A. Y. (2018). Facial emotion recognition: A survey and real-world user experiences in mixed reality. *Sensors*, 18(2). DOI: 10.3390/s18020416.

Mohan, K., Seal, A., Krejcar, O., and Yazidi, A. (2021). Facial expression recognition using local gravitational force descriptor-based deep convolution neural networks. *IEEE Transactions on Instrumentation and Measurement*, 70:1–12. DOI: 10.1109/TIM.2020.3031835.

Monzo, D; Albiol, A. and Mossi, M. J. (2010). A comparative study of facial landmark localization methods for face recognition using hog descriptors. In *2010 20th International Conference on Pattern Recognition*, pages 1330–1333, Istanbul, Turkey. IEEE. DOI: 10.1109/ICPR.2010.1145.

Nhu, H. L., Dang, H. V., and Xuan, H. H. (2023). Facial emotion recognition by combining deep learning and averaged weight of face-regions. In *2023 15th International Conference on Knowledge and Systems Engineering (KSE)*, pages 1–4. DOI: 10.1109/KSE59128.2023.10299482.

Persona, L. (2022). Reconhecimento de emoções por meio da geometria facial com coordenadas normalizadas dos landmarks (em inglês, recognition of emotions through facial geometry using normalized landmark coordinates). Master thesis. DOI: 10.11606/D.59.2022.tde-21112023-114806.

Persona, L., Meloni, F., and Macedo, A. A. (2023). An accurate real-time method to detect the smile facial expression. In *Proceedings of the 29th Brazilian Symposium on Multimedia and the Web*, WebMedia '23, page 46–55, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3617023.3617031.

Picard, R. W. (2016). Automating the recognition of stress and emotion: From lab to real-world impact. *IEEE MultiMedia*, 23(3):3–7. DOI: 10.1109/MMUL.2016.38.

Poulose, A., Kim, J. H., and Han, D. S. (2021). Feature vector extraction technique for facial emotion recognition using facial landmarks. In *2021 International Conference on Information and Communication Technology Convergence (ICTC)*, pages 1072–1076. DOI: 10.1109/ICTC52510.2021.9620798.

Revina, I. and Emmanuel, W. S. (2018). A survey on human face expression recognition techniques. *Journal of King Saud University - Computer and Information Sciences*. DOI: 10.1016/j.jksuci.2018.09.002.

Salman, F. Z., Madani, A., and Kissi, M. (2016). Facial expression recognition using decision trees. In *2016 13th International Conference on Computer Graphics, Imaging and Visualization (CGiV)*, pages 125–130, Beni Mellal, Morocco. IEEE. DOI: 10.1109/CGiV.2016.33.

Testa, R. L., Corrêa, C. G., Machado-Lima, A., and Nunes, F. L. S. (2019). Synthesis of facial expressions in photographs: Characteristics, approaches, and challenges. *ACM Comput. Surv.*, 51(6):124:1–124:35. DOI: 10.1145/3292652.

Tottenham, N., Tanaka, J., Leon, A., Mccarry, T., Nurse, M., Hare, T., Marcus, D., Westerlund, A., Casey, B., and Nelson, C. (2009). The nimstim set of facial expressions: Judgments from untrained research participants. *Psychiatry research*, 168:242–9. DOI: 10.1016/j.psychres.2008.05.006.

Ullah, S. and Tian, W. (2021). A systematic literature review of recognition of compound facial expression of emotions. In *Proceedings of the 2020 4th International Conference on Video and Image Processing*, page 116–121. DOI: 10.1145/3447450.3447469.

Vaillancourt, J. (2010). Statistical methods for data mining and knowledge discovery. In *Proceedings of the 8th International Conference on Formal Concept Analysis*, ICFCA'10, pages 51–60, Berlin, Heidelberg. Springer-Verlag. DOI: 10.1007/978-3-642-11928-6_4.

Viola, P. and Jones, M. J. (2001). Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154. DOI: 10.1023/B:VISI.0000013087.49260.fb.

Vonikakis, V. and Winkler, S. (2020). Identity-invariant facial landmark frontalization for facial expression analysis. In *International Conference on Image Processing (ICIP)*, pages 2281–2285. DOI: 10.1109/ICIP40778.2020.9190989.

Winterle, P. (2014). *Vetores e Geometria Analítica*. MAKRON. Book.

Wu, Y. and Ji, Q. (2018). Facial landmark detection: A literature survey. *International Journal of Computer Vision*, 2:115–142. DOI: 10.1007/s11263-018-1097-z.

Xie, W.; Shen, L. and Jiang, J. (2017). A novel transient wrinkle detection algorithm and its application for expression synthesis. *IEEE Transactions on Multimedia*, 19(2):279–292. DOI: 10.1109/TMM.2016.2614429.

XIE, W., SHEB, L., YANG, M., and HOU, Q. (2015). Lighting difference based wrinkle mapping for expression synthesis. In *2015 8th International Congress on Image and Signal Processing (CISP)*, pages 636–641, Shenyang, China. IEEE. DOI: 10.1109/CISP.2015.7407956.

Zhang, Q., Wang, Z., Liu, Y., Qin, Z., Zhang, K., and Gedeon, T. (2023). Geometric-aware facial landmark emotion recognition. In *2023 6th International Conference on Software En-*

gineering and Computer Science (CSECS), pages 1–6. DOI: 10.1109/CSECS60003.2023.10428424.