2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) | 978-1-6654-4899-4/21/$31.00 ©2021 IEEE | DOI: 10.1109/CVPRW53098.2021.00172

# Improving Parkinson Detection using Dynamic Features from Evoked Expressions in Video

Luis F. Gomez[1], Aythami Morales[1], Juan R. Orozco-Arroyave[2], Roberto Daza[1], Julian Fierrez[1]
[1]Biometrics and Data Pattern Analytics Lab, Universidad Autonoma de Madrid, Madrid, Spain
[2]Universidad de Antioquia, Medellin, Colombia
{luisf.gomez, aythami.morales, roberto.daza, julian.fierrez}@uam.es
rafael.orozco@udea.edu.co

## Abstract

*Hypomimia, also known as "facial masking", is a common symptom of Parkinson's Disease (PD). PD is a neurological disorder characterized by non-motor and motor impairments. Hypomimia is the reduction of facial expressiveness, including the emotion expressions. In this work, we explore the use of static and dynamic features for the analysis of evoked facial gestures in PD patients. The main contributions of this work are: (1) We propose a multimodal PD detection system based on both static and dynamic features obtained from evoked face gestures; (2) we propose a novel set of 17 dynamic features to characterize the facial expressiveness and demonstrate that facial dynamics features can be used to improve PD detection; and (3) we analyze different evoked facial expressions and its performance for PD detection. Different expressions activate different Action Units (AUs) and we analyze to what extent each of these AUs contribute to PD detection. The results show that the use of static features generated by pre-trained deep architectures yield up to 77.36% of accuracy for PD detection and the combination with dynamic features improves PD detection by up to 13.46% (from 75.00% to 88.46%). Our experiments also suggest differences in the performance of evoked face gestures in this PD detection task.*

## 1. Introduction

Parkinson's Disease (PD) is a neurological disorder that affects between 1 and 2 percent of people over 65 years old [6]. PD is characterized by motor deficits including bradykinesia, rigidity, postural instability, tremor, and dysarthria. Non-motor deficits include depression, anxiety, sleep disorders, and slowing of thought [31]. Besides, bradykinesia affects facial muscles, making it difficult for them to express emotions or specific expressions on their faces. Possible signs of such abnormalities include reduced range of facial muscle movement, wider eye-opening, a half-open mouth, and slower blinking. All of these phenomena in their facial expression are grouped in the literature and referred to as hypomimia [4], which is the result of motor alterations at the level of the facial muscles.

The evaluation of the condition of Parkinson's patients is performed at hospitals by neurologist experts, who usually administer the Movement Disorder Society - Unified Parkinson's Disease Rating Scale (MDS-UPDRS) [19]. The items evaluated with the MDS-UPDRS scale range between 0 and 4, where 0 means no perceived impairment and 4 means completely impaired. Section III of the scale includes an item to evaluate hypomimia and consequently facial expressions in patients [19]. The following list indicates the possible values that the hypomimia evaluation item can take:

0. Normal: Normal facial expression.
1. Slight: Minimal masked facies manifested only by decreased frequency of blinking.
2. Mild: In addition to decreased eye-blink frequency, masked facies present in the lower face as well, namely fewer movements around the mouth, such as less spontaneous smiling, but lips not parted.
3. Moderate: Masked facies with lips parted some of the time when the mouth is at rest.
4. Severe: Masked facies with lips parted most of the time when the mouth is at rest.

Because neurological evaluation depends on the expertise of neurologists, this measure is variable and may contain biases of experts at the time of evaluation. The development of automatic systems to assist in the evaluation has increased over the years. These systems are used in

978-1-6654-4899-4/21/$31.00 ©2021 IEEE
DOI 10.1109/CVPRW53098.2021.00172

different aspects of PD, such as speech [29, 28], handwriting [33, 8, 13, 9], gait [17, 10], hand movement [35], and facial expression [2]. Among these physiological modalities, the Facial Expressivity Evaluation (FEE) is a field rarely explored in automatic PD detection, but this modality is very important in the social interaction and the nonverbal communication of PD patients. On the other hand, FEE has become a popular field among engineers and computer scientists, which opens a space for research in different applications related to Affective Computing.

Our work is focused on the study of FEE in PD patients and how facial gesture analysis can be used to improve PD detection. The main aim is to explore dynamic features obtained from videos collected from PD patients and healthy controls and to evaluate their capability to produce specific face gestures during evoked emotions.

### 1.1. Related Works

One of the first studies of FEE in Parkinson's patients was conducted by Simons et al. [34]. The authors used videos with social interactions to evoke expressions in patients. The videos were recorded using facial measurements, self-questionnaires, and subjective measures based on the Facial Action Units (FAUs) presented in [16]. The authors evaluated a total of 44 participants (25 healthy controls and 19 PD patients). The results of the study show that patients have a reduced ability to generate spontaneous facial expressions. In 2006, Bowers et al. [5], based on the bradykinesia in the patients, concluded that the patients would have slower and less amplitude of intentional facial expressions than healthy controls. The patient's videos were evaluated frame-by-frame and the entropy in temporal changes of the frames was calculated [32]. The results showed that healthy controls had a higher entropy than the patients. Additionally, the authors reported that the patients took longer in reaching a peak in facial expression.

In 2016 Almutiry et al. [1] presented a longitudinal analysis on FEE in 4 PD patients who were recorded for 6 weeks. Their experiments considered 5 videos per week per patient (recorded only 1 time per day). Four healthy controls were also recorded for 5 days per week. All the subjects performed specific facial expressions on a daily basis. The authors used 2 feature extractors: 1) Active Appearance Model (AAM) and 2) a Constrained Local Model (CLM), to obtain a total of 27 facial features. The results confirmed the observations made by Bowers et al. [5], showing that patients had less movement capability than controls.

One year later, Bandini et al. [2] classified a total of 4 acted and imitated expressions (happiness, anger, disgust, and sadness) involving 17 PD patients (13 male) and 17 healthy controls (6 male). Each expression was modeled with 49 landmarks [36, 22] around the face. A linear combination of specific points was performed to extract a total of 20 features. A Support Vector Machine (SVM) was trained to detect each emotion expressed by the participants. The results showed that acted and imitated expressions by the controls yielded a better accuracy than the PD patients.

More recently, Grammatikopoulou et al. [23] analyzed the facial expressions of a total of 34 participants (23 with PD and 11 healthy controls). This study was performed from images captured directly with smartphones. Two geometric feature sets [37] were extracted: one using Google Face API and the other using Microsoft Face API [21]. The stored feature sets were used to estimate the Hypomimia Severity index (HSi) with 2 linear regression models (one per feature set). These indices were also used to perform the classification between healthy controls and PD patients. The authors report sensitivity and specificity values of $0.79$ and $0.82$ for HSi1, and $0.89$ and $0.73$ for HSi2.

### 1.2. Contributions of this Work

As shown in the literature review, there is still much to explore in the field of FEE for modeling hypomimia in Parkinson's Disease (PD) patients. To the best of our knowledge, this is the first work exploring dynamic feature sets to improve PD detection. The main contributions are:

- We propose a multimodal PD detection system based on both static and dynamic features obtained from evoked face gestures.
- We propose a novel set of 17 dynamic features to characterize the facial expressiveness and demonstrate that facial dynamics features improve the PD detection.
- We analyze different evoked facial expressions and their performance for PD detection. Different expressions activate different action units (AUs) and we analyze to what extent each of these AUs contributes to PD detection.

Our experiments are performed on the FacePark-GITA database. This database includes short videos with evoked facial expressions from 30 PD patients and 24 healthy participants. Our results demonstrate that dynamic features improve the performance of PD detection with a classification accuracy of $88.46\%$ for the evoked surprise gesture.

## 2. Parkinson Detection based on Evoked Expressions from Video

Parkinson's patients show difficulties to control the facial gestures mainly caused by bradykinesia. We propose to study different facial gestures aimed to characterize the hypomimia in PD patients. The features included in this study can be divided into: 1) static features obtained from face recognition pre-trained models; and 2) dynamic features extracted from motion sequences based on facial meshes. A

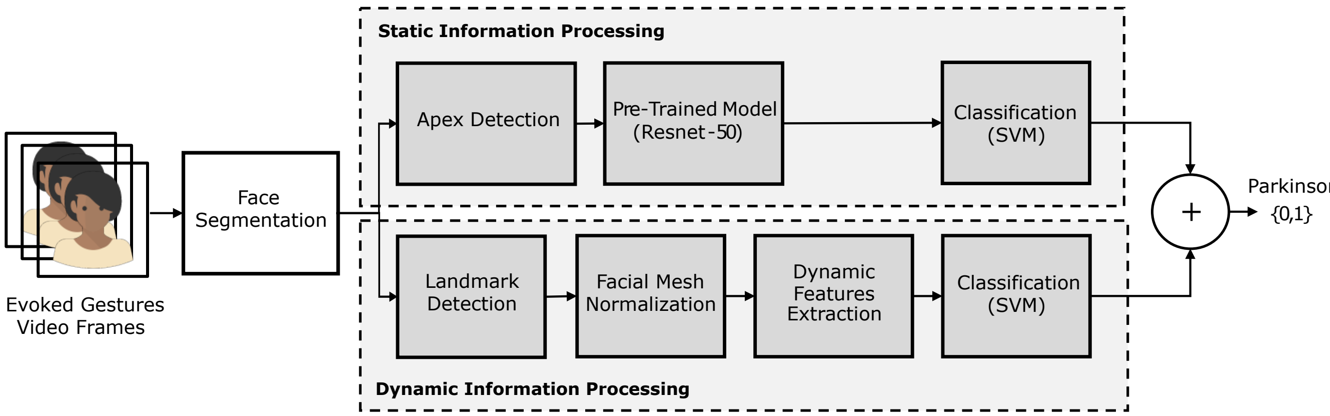


Figure 1: Block diagram of the proposed PD detection system based on static and dynamic features from evoked face gestures.

| Facial Expression | AUs |
|---|---|
| Happy | $6+12$ |
| Anger | $9+15+16$ |
| Surprised | $1+2+5+26$ |
| Wink | 46 |

Table 1: AUs associated to the evoked gestures included in our experiments.

block diagram of the systems can be seen in Figure 1, which shows a multi-classifier approach [14] that simply averages the classification scores of both parallel classifiers [15].

Our experiments include 4 evoked facial gestures (Happy, Angry, Surprise and Wink) determined by different FAUs. Each evoked gesture involves the coordination of several facial muscles in different face regions. We have chosen 3 of the 6 basic emotions [12]. These evoked gestures have been chosen because its simplicity. These three emotions comprise a total number of 9 FAU (see Table 1).

## 2.1. Static Features

In this work we employ a ResNet50 architecture [24], with 50 layers and 25.6M parameters pre-trained for face verification tasks [7]. This model is used to generate an initial face representation. The model is used as feature extractor by removing the final decision layer. For each face image, the model generates a $1 \times 2048$ feature vector. Although the model was initially trained for face verification, the generated embeddings are rich in face features including gestures [20, 30]

For each evoked expression video sequence, we select the frame representing a maximum peak of the facial gesture (Apex). The $1 \times 2048$ feature vector is extracted for the selected frame (only one feature vector per sequence).

## 2.2. Dynamic Features

Searching to characterize the hypomimia in PD patients, we propose to study dynamic features of different facial areas using facial landmark detection and sequence analysis.

The most recent facial landmark detection algorithms are able to get a large number of facial landmarks in different pose and changing illumination conditions. The position of these landmarks provides information about the movement of different face regions. In our approach, we use the landmarks to calculate distances between key face regions and characterize the movements associated to the AUs. Therefore, our dynamic feature set is structured into two blocks: *i*) Facial Landmark Detection Algorithm; *ii*) Dynamic Feature Extraction.

### 2.2.1 Facial Landmark Detection Algorithm

The performance of facial landmarks detectors has been largely improved in the last years [11, 27]. In this work we use the landmark detector from MediaPipe library[1] that estimates 468 3D face landmarks in real-time [27]. The 468 landmarks contain information of different facial areas like cheeks, forehead, mouth, eyes, etc. This detailed face mesh allows to analyze different face regions and the motion of most of the 43 muscles in the face.

The library MediaPipe contains different solutions. In the present study, we use two solutions: *i)* MediaPipe Face Mesh [26], which is a face geometry solution that estimates 468 3D face landmarks; and *ii)* MediaPipe Holistic, which utilizes the previous 468 face landmark model in a newer version that gets better results in the face gestures included in this work. The two face landmark models are based on a custom residual neural network architecture. For all points we used the average between both models.

[1] https://google.github.io/mediapipe/

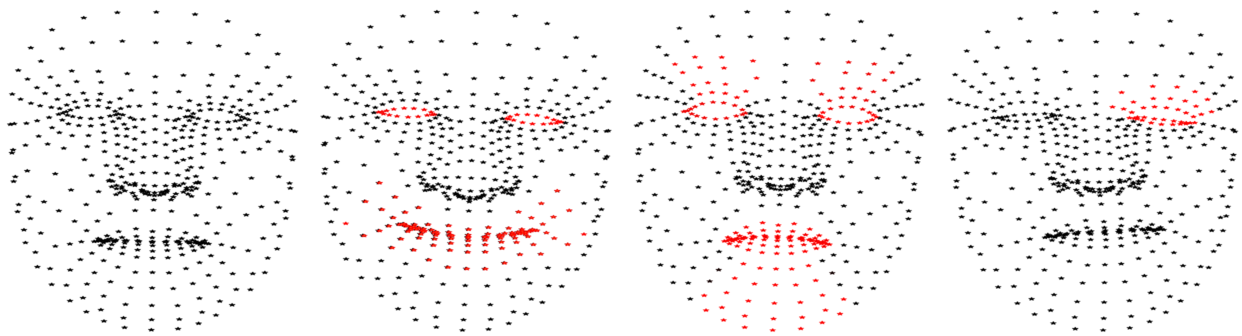


Figure 2: Example of the facial mesh generated by MediaPipe for four different face expressions. Representative FAUs for each expression are highlighted in red.

We analyze each frame of the videos as follows: *i)* The models invoke the face detector based on BlazeFace [3] to select the most centered face. BlazeFace includes Convolutional Neural Networks (CNN) inspired by, but distinct from MobileNetV1/V2 [25]. And *ii)* The models invoke a 3D landmark network whose input is a cropped video frame without additional depth input. The model outputs the positions of the 3D points mesh of 468 landmarks [26].

The models outputs are:

- **x** and **y**: Landmark coordinates normalized to $[0, 1]$ by the image width and height respectively.
- **z**: Represents the landmark depth with the depth at the center of the head being the origin, and the smaller the value, the closer the landmark is to the camera.
- Visibility: A value ranging between $[0, 1]$ indicating the likelihood of the landmark being visible or hidden (occluded by another body part) on a frame.
- Presence: A value that indicates if the point is present or outside a frame.

Figure 2 shows examples of the face meshes of a participant during the evocation of different facial gestures.

#### 2.2.2 Dynamic Feature Extraction

The facial mesh of a single user can be described as a set of landmarks $\boldsymbol{\alpha}_i^j$ with $i = 1, 2, \ldots, 468$, and $j = 1, 2, \ldots, N$:

$$\boldsymbol{\alpha}_i^j = \left[x_i^j, y_i^j, z_i^j\right], \tag{1}$$

where $i$ denotes a particular landmark and $j$ denotes the frame at which the face mesh was captured. The landmark $\boldsymbol{\alpha}_i^j$ is transformed to a common coordinate system by subtracting the center of mass of the face mesh:

$$\mathbf{m}^j = \frac{1}{468}\sum_{i=1}^{468} \boldsymbol{\alpha}_i^j. \tag{2}$$

The normalized landmarks are therefore:

$$\hat{\boldsymbol{\alpha}}_i^j = \boldsymbol{\alpha}_i^j - \mathbf{m}^j. \tag{3}$$

With all the feature points centered on the same coordinate axis we then normalize the scale of the whole facial mesh. This is done by obtaining the InterCanthal Distance (ICD) or the Interocular distance for each frame, then this distance is averaged across frames:

$$ICD = \frac{1}{N}\sum_{j=1}^{N} \left\| \hat{\boldsymbol{\alpha}}_{133}^j - \hat{\boldsymbol{\alpha}}_{362}^j \right\|, \tag{4}$$

where $\|*\|$ is the L2-norm, and the feature points $\hat{\boldsymbol{\alpha}}_{133}^j$ and $\hat{\boldsymbol{\alpha}}_{362}^j$ are the canthal locations of each eye. Finally the normalized landmarks are:

$$\boldsymbol{\beta}_i^j = \frac{\hat{\boldsymbol{\alpha}}_i^j}{ICD}. \tag{5}$$

With the feature points normalized, we calculate a total of 8 distances $D_i^j$ related to the face configuration for each frame, see Table 2 and Figure 3, where the distances

$$d_{m,n}^j = \left\| \boldsymbol{\beta}_m^j - \boldsymbol{\beta}_n^j \right\| \tag{6}$$

are calculated between the landmarks $m$ and $n$.

We propose to model the facial dynamics through 17 features as described in Table 3. This feature set is aimed to characterize the dynamics of each of the 8 proposed distances $D_i^j$ ($i$ denotes each distance) across the $N$ frames ($j = 1, \ldots, N$). The proposed feature set includes measures related to the way each user performs each face gesture: velocity (**v**), acceleration (**a**), and jerk (**j**). For each video sequence (i.e. one video sequence for face gesture), we finally obtain a feature matrix $\mathbf{F} \in \mathbb{R}^8 \times \mathbb{R}^{17}$.

Table 2: Description of the 8 computed distances $D_i^j$ $(i = 1, \ldots, 8)$ measured for each face mesh (one for each frame $j = 1, \ldots, N$). The subscript $d_{xxx,yyy}^j$ indicates the index of the landmarks of the face mesh obtained from MediaPipe.

| | |
|---|---|
| Opening of the left eye: | $D_1^j = \frac{1}{7}(d_{398,382}^j + d_{384,381}^j + d_{385,380}^j + d_{386,374}^j + d_{387,373}^j + d_{388,390}^j + d_{466,249}^j)$ |
| Opening of the right eye: | $D_2^j = \frac{1}{7}(d_{246,7}^j + d_{161,163}^j + d_{160,144}^j + d_{159,145}^j + d_{158,153}^j + d_{157,154}^j + d_{173,155}^j)$ |
| Opening of the jaw: | $D_3^j = \frac{1}{7}(d_{80,170}^j + d_{81,140}^j + d_{82,171}^j + d_{13,175}^j + d_{312,396}^j + d_{311,369}^j + d_{310,395}^j)$ |
| Opening of the mouth: | $D_4^j = \frac{1}{7}(d_{80,88}^j + d_{81,178}^j + d_{82,87}^j + d_{13,14}^j + d_{312,317}^j + d_{311,402}^j + d_{310,318}^j)$ |
| Left eyebrow height: | $D_5^j = \frac{1}{5}(d_{384,336}^j + d_{385,296}^j + d_{386,334}^j + d_{387,293}^j + d_{388,300}^j)$ |
| Right eyebrow height: | $D_6^j = \frac{1}{5}(d_{161,70}^j + d_{160,63}^j + d_{159,105}^j + d_{158,66}^j + d_{157,107}^j)$ |
| Lip Stretch: | $D_7^j = \frac{1}{2}(d_{33,78}^j + d_{263,308}^j)$ |
| Mouth width: | $D_8^j = (d_{78,308}^j)$ |

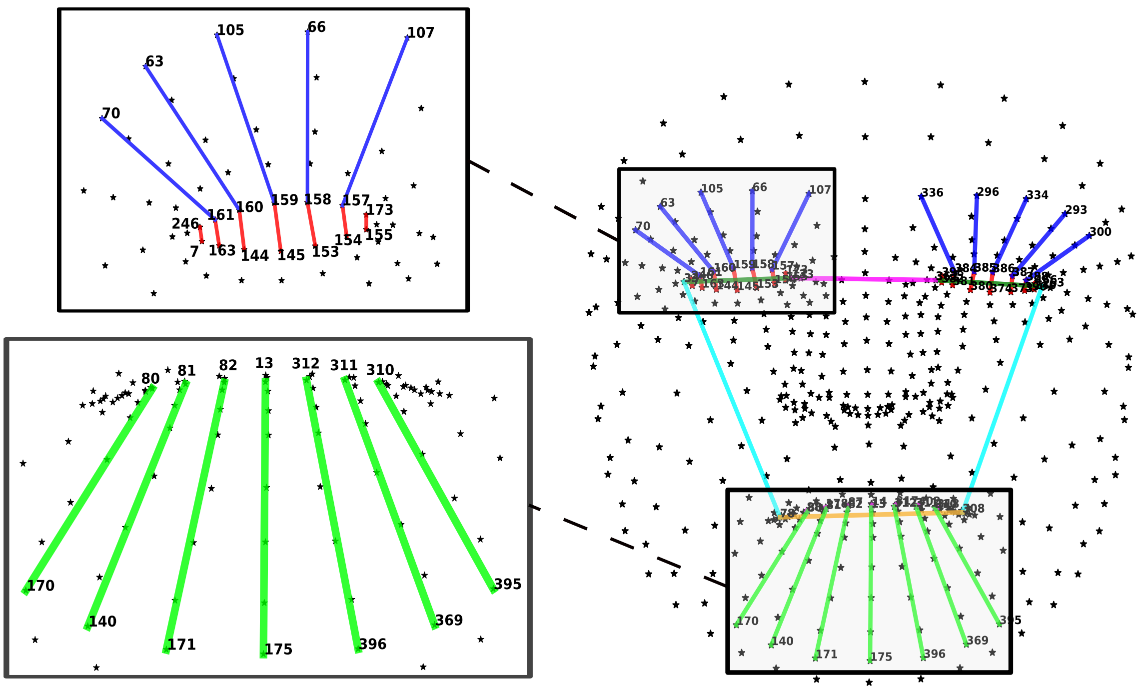


Figure 3: Distances measured to characterize the face dynamics.

### 2.3. SVM Classification

The automatic classification between healthy people and PD patients is performed using Support Vector Machines (SVMs). The SVM classification experiments consider linear and Gaussian kernels. The optimization of hyper-parameters is performed in a search grid of powers of ten with $C \in \{10^{-8}, 10^{-7}, \ldots, 10^2, 10^3\}$ and $\gamma \in \{10^{-8}, 10^{-7}, \ldots, 10^3\}$ for the Gaussian kernel, and for the linear kernel the search considered $C \in \{10^{-8}, 10^{-7}, \ldots, 10^3, 10^4\}$. Optimization and evaluation of the models is performed following a 5-folds cross-validation strategy. Results of the SVM classification are reported in terms of Accuracy (Acc), Precision (Pre), Recall (Rec), and F1-Score (F1).

## 3. Experiments and Results

### 3.1. Database

The FacePark database was created by GITA Lab[2]. The recording of patients is still ongoing and the most updated version of the corpus contains video recordings of 24 healthy participants and 30 PD patients. The videos were recorded at 15 frames per second in non-controlled environment conditions, i.e., light conditions and the background were not controlled prior the recording and differ among participants. PD patients were diagnosed by an expert neurologist and were evaluated according to the MDS-UPDRS-III scale and the Hoehn and Yahr scale (H&Y) [18]. A summary of the clinical and demographic information is pre-

[2] https://gita.udea.edu.co/

Table 3: Description of the 17 features $F_k^i$ $(k = 1, \ldots, 17)$ extracted from each distance $D_i^j$ across the $N$ frames $(j = 1, \ldots, N)$.

| |
|---|
| $F_1^i$ = number of local maxima $N$ |
| $F_2^i = (\text{average velocity } \tilde{\mathbf{v}}) / \lvert\mathbf{v}\rvert_{\max}$ |
| $F_3^i = (\text{average velocity } \tilde{\mathbf{v}}) / \mathbf{v}_{\max}$ |
| $F_4^i = (\text{velocity rms } \mathbf{v}_{RMS}) / \lvert\mathbf{v}\rvert_{\max}$ |
| $F_5^i = (\text{centripetal acceleration rms } \mathbf{a}_c) / \mathbf{a}_{\max}$ |
| $F_6^i = (\text{tangential acceleration rms } \mathbf{a}_t) / \mathbf{a}_{\max}$ |
| $F_7^i = (\text{acceleration rms } \mathbf{a}_{RMS}) / \mathbf{a}_{\max}$ |
| $F_8^i = (\text{integrated abs. centri. acc. rms } \mathbf{a}_{ic}) / \mathbf{a}_{\max}$ |
| $F_9^i$ = standard deviation of $\mathbf{v}$ |
| $F_{10}^i$ = standard deviation of $\mathbf{a}$ |
| $F_{11}^i$ = average abs. jerk $\lvert\mathbf{j}\rvert$ |
| $F_{12}^i$ = average jerk $\tilde{\mathbf{j}}$ |
| $F_{13}^i = \lvert\mathbf{j}\rvert_{\max}$ |
| $F_{14}^i = \tilde{\mathbf{j}}_{\max}$ |
| $F_{15}^i$ = average rms jerk $\mathbf{j}_{RMS}$ |
| $F_{16}^i = (\sum(\mathbf{v} > 0) / (\sum(\mathbf{v} < 0))$ |
| $F_{17}^i = \max_j D_i^j - \min_j D_i^j$ |

sented in Table 4. All participants gave written informed consent. The study is in accordance with the Declaration of Helsinki and it was approved by the Ethical Research Committee at the University of Antioquia.

The participants of the study were asked to produce different facial expressions while being recorded. A total of five video-task recordings are included: right eye wink, left eye wink, smile, anger, and surprise. The average duration of each video is 6 seconds. Patients have an average age of 69 years old and healthy subjects were chosen with a similar range of age. Possible bias introduced by age or gender were discarded via a chi-square statistical test ($p = 0.44$) and a Welch's t-test ($p = 0.15$), respectively.

### 3.2. Experimental Protocol

Our acquisition protocol contains videos associated with different tasks that evoke facial gestures, including emotional responses (smile, anger and surprise) and coordinated facial gestures (right eye wink, left eye wink). Given the nature of winking, both activities are grouped into a single set of features for the experiments performed in this work. We propose an experimental protocol to explore: 1) the discriminative power of static and dynamic feature sets; and 2) the discriminative power of different face gestures.

**Static features classifier:** We evaluate the use of a pre-trained Face Recognition model and evaluate its capability in the classification of PD patients.

- *Hypothesis (H1):* the facial expressions shown during evoked emotions are different between PD patients and healthy controls.

- *Experiment:* we evaluate the performance of PD detection for different face gestures using a pre-trained Face Recognition model trained with VGGFace2 [7] and an acquisition protocol including 4 evoked gestures.

- *Methodology:* First, one Apex frame per video-task was extracted manually in all the participants. Later, each Apex frame is used as input of the pre-trained ResNet50 model (see Section 2.1) to obtain a set of feature vectors. Finally, these feature vectors are used to determine to what extent they can discriminate between PD patients and healthy controls.

**Dynamic features classifier:** We propose a dynamic feature set for modeling Hypomimia in facial expressions and to improve the PD detection.

- *Hypothesis (H2):* dynamic information in evoked facial gesture allows modeling of hypomimia in PD patients.

- *Experiment:* we propose a feature set to model the dynamics in areas directly related to the FAUs of each facial gesture. We evaluate the performance of this feature set in the detection of PD patients.

- *Methodology:* Using the methodology presented in Section 2.2.2 for dynamic feature extraction, all videos are analyzed and their resulting dynamic feature vectors are used in an SVM classifier to determine the effectiveness in classifying between PD patients and healthy controls.

**Multimodal classifier:** We propose a multimodal system [15] with the combination of static and dynamic features to improve the performance of PD detection.

Table 4: Demographic and clinical information of the participants included in the FacePark-GITA database. Table from [20].

| | **PD patients** | | **Healthy participants** | |
|---|---|---|---|---|
| | Men | Women | Men | Women |
| # of Participants | 18 | 12 | 12 | 12 |
| Age [years] | 70.2 ± 10.4 | 67.4 ± 10.9 | 65.3 ± 8.7 | 65.2 ± 10.1 |
| Age range [years] | 52 – 90 | 53 – 87 | 49 – 83 | 49 – 80 |
| $t$ [years] | 8.7 ± 5.4 | 15.6 ± 17.3 | — | — |
| $t$ range [years] | 2 – 20 | 1 – 45 | — | — |
| MDS-UPDRS-III | 35.4 ± 13.9 | 29.7 ± 12.3 | — | — |
| MDS-UPDRS-III range | 16 – 65 | 15 – 54 | — | — |
| H&Y | 2.3 ± 0.5 | 2.5 ± 0.5 | — | — |
| H&Y range | 2 – 3 | 2 – 3 | — | — |

MDS-UPDRS: Movement Disorder Society - Unified Parkinson's Disease Rating Scale. H&Y: Hoehn & Yahr scale. $t$: Years since diagnosis

Table 5: Binary classification results (PD vs Healthy) using the 3 proposed strategies. Accuracy (Acc), Precision (Pre), Recall (Rec).

| Facial gestures | Acc[%] | Pre[%] | Rec[%] | F1[%] |
|---|---|---|---|---|
| **Static Features** | | | | |
| Happy | 75.00 | 74.29 | 86.67 | 80.00 |
| Angry | 66.00 | 67.74 | 75.00 | 71.19 |
| Surprise | 75.00 | 72.22 | 89.66 | 80.00 |
| Wink | 77.36 | 75.00 | 90.00 | 81.82 |
| **Dynamic Features** | | | | |
| Happy | 67.31 | 68.57 | 80.00 | 73.85 |
| Angry | 66.00 | 66.67 | 78.57 | 72.13 |
| Surprise | 71.15 | 70.59 | 82.76 | 76.19 |
| Wink | 69.81 | 69.44 | 83.33 | 75.76 |
| **Static + Dynamic (Linear SVM)** | | | | |
| Happy | 67.31 | 69.70 | 76.67 | 73.02 |
| Angry | 78.00 | 75.76 | 89.29 | 81.97 |
| Surprise | 88.46 | 87.10 | 93.10 | 90.00 |
| Wink | 79.25 | 82.76 | 80.00 | 81.36 |
| **Static + Dynamic (RBF SVM)** | | | | |
| Happy | 73.08 | 73.53 | 83.33 | 78.12 |
| Angry | 76.00 | 73.53 | 89.29 | 80.65 |
| Surprise | 86.54 | 86.67 | 89.66 | 88.14 |
| Wink | 81.13 | 85.71 | 80.00 | 82.76 |

- *Hypothesis (H3):* The complementarity between static and dynamic information from the video sequences allows to better classify between PD patients and healthy controls.
- *Experiment:* The decision scores from the static and dynamic feature classifiers are combined for improved PD detection.
- *Methodology:* the classification score of both static and dynamic feature classifiers are combined at score level using Sum Rule fusion [15].

### 3.3. Results

Table 5 summarizes the results of the 3 experiments. The results show that Apex images (Static Features) and pre-trained Face Recognition models obtain PD classification accuracies from 66.00% to 77.36% depending on the face gesture (one decision per video, average duration 6 seconds, see Section 3.1 for the experimental details). These results provide support for the first hypothesis (*H1*), where it is presented the idea that images during maximal evocation of a gesture allow PD detection. Besides, the results obtained for the dynamic feature set show accuracies ranging from 66.00% to 71.15%. This performance suggests that dynamic information allows the modeling of hypomimia and supports the second hypothesis (*H2*).

Regarding the performance obtained for the different face gestures, the best results are achieved using surprise and wink gestures that are directly related to the upper muscles of the face (AU 1, 2, 5, 46, etc.), suggesting that these AUs provide more information to model hypomimia.

Finally, it can be observed that the combination at the score level of static and dynamic information increases the PD detection accuracy for almost all facial gestures (see Figure 4). The improvements of the classifiers are above 3.77% (Wink gesture) and up to 13.46% (Surprise gesture). The new representation space created with the multimodal scores improves the separation between healthy controls and PD patients, demonstrating the third hypothesis (*H3*) presented in this work.

## 4. Conclusions

This work has explored the performance of evoked facial gestures for PD detection. We have studied the use of static and dynamic features for modeling hypomimia in PD patients. Face videos of people while evoking four different facial gestures were considered for this study (i.e., Happy, Angry, Surprise, Wink). We have proposed a novel feature set of 17 features aimed to characterize the expressiveness in evoked facial gestures in video sequences.

The approach based on static features provides up to 77.36% PD detection accuracy in the wink gesture, which is 11.36% higher than using the angry expression. Similarly, it can be observed that the dynamic features provided a performance of 71.15% PD detection accuracy using the surprise gesture. Once gain the angry gesture obtained the lowest PD detection accuracy with 66.15%. These result suggests that: 1) both dynamic and static features can be used for PD detection, and 2) different face gestures (characterized by different AUs) show performance differences with highest PD detection rates obtained by gestures associated to upper muscles of the face.

Finally, the static and dynamic domains were combined following a score level fusion strategy. This approach evidences that the results of both domains are complementary to each other and improved the separation space of healthy controls and PD patients. The results showed improvements in the performance of the system, increasing up to 13.46% in PD detection accuracy (from 75.00% to 88.46%).

The results obtained in this work encourage us to explore new approaches to model hypomimia through static and dynamic features. There is a need for larger databases for PD detection based on facial expressions. With larger datasets, approaches based on data driven methods (e.g. deep learning architectures) should be considered.

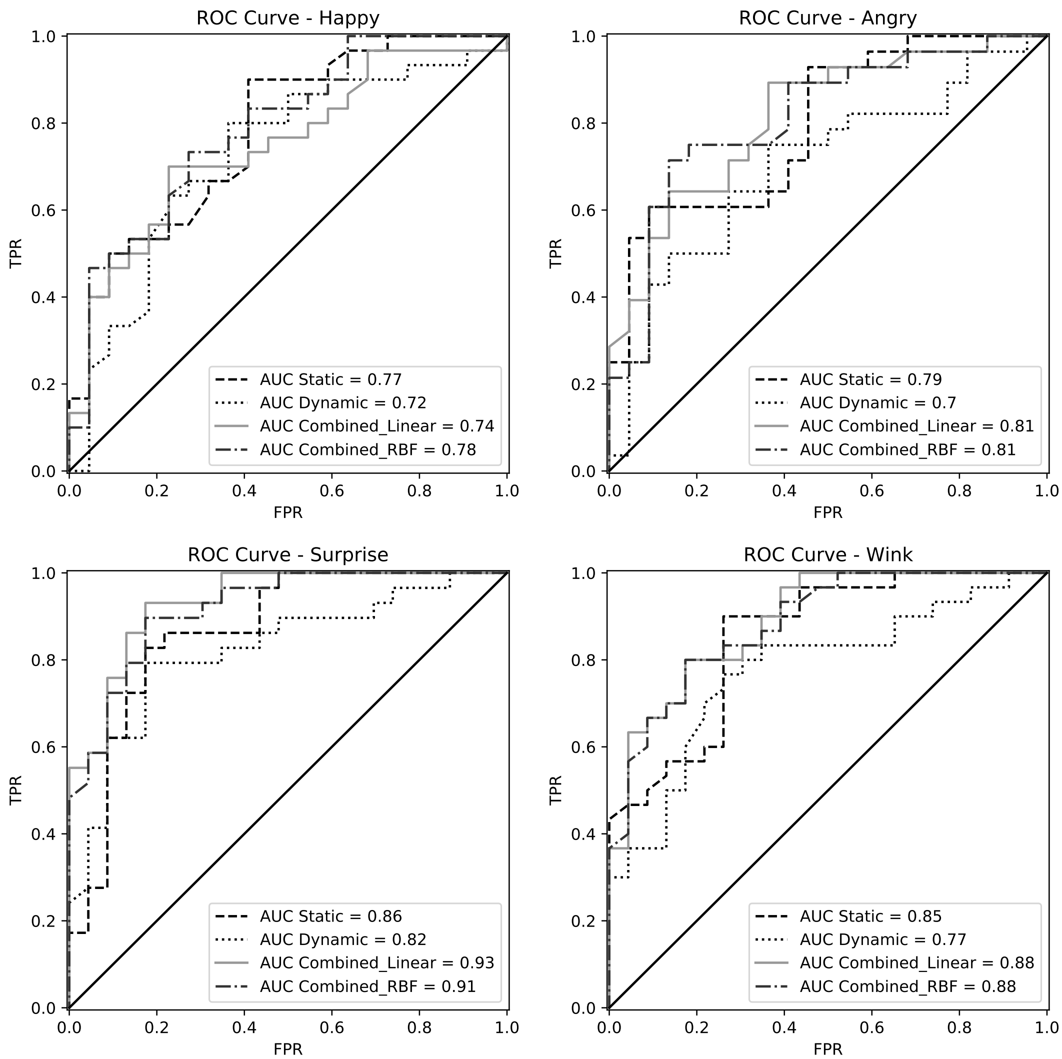


Figure 4: ROC curves for the different feature sets and facial gestures.

## 5. Acknowledgments

UAM is supported by projects BIBECA (RTI2018-101248-B-I00 MINECO/FEDER), TRESPASS-ETN (MSCA-ITN-2019-860813), and IDEA-FAST (H2020-IMI2-2018-15-853981). R. Daza is supported by a research fellowship from Spanish MINECO. Support also by CODI at Universidad de Antioquia, Grant # PRG2017-15530.

## References

[1] R. Almutiry, S. Couth, E. Poliakoff, S. Kotz, M. Silverdale, and T. Cootes. Facial behaviour analysis in Parkinson's disease. *Springer LNCS*, 9805:329–339, 2016. 2

[2] A. Bandini, S. Orlandi, H. J. Escalante, F. Giovannelli, et al. Analysis of facial expressions in Parkinson's disease through video-based automatic methods. *Journal of Neuroscience Methods*, 281:7–20, 2017. 2

[3] V. Bazarevsky, Y. Kartynnik, A. Vakunov, K. Raveendran, and M. Grundmann. Blazeface: Sub-millisecond neural face detection on mobile GPUs. *arXiv:1907.05047*, 2019. 4

[4] M. Bologna, G. Fabbrini, L. Marsili, G. Defazio, P. D. Thompson, and A. Berardelli. Facial bradykinesia. *J Neurol Neurosurg Psychiatry*, 84(6):681–685, 2013. 1

[5] D. Bowers, K. Miller, W. Bosch, D. Gokcay, O. Pedraza, U. Springer, and M. Okun. Faces of emotion in Parkinson's disease: Micro-expressivity and bradykinesia during voluntary facial expressions. *Journal of the International Neuropsychological Society*, 12:765–773, 2006. 2

[6] R. Cacabelos. Parkinson's disease: from pathogenesis to pharmacogenomics. *International Journal of Molecular Sciences*, 18(3):551, 2017. 1

[7] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman. VGGFace2: A dataset for recognising faces across pose and age. In *IEEE Intl. Conf. on FG*, pages 67–74, 2018. 3, 6

[8] R. Castrillon, A. Acien, J.R. Orozco-Arroyave, A. Morales, J.F. Vargas, R. Vera-Rodriguez, J. Fierrez, J. Ortega-Garcia, and A. Villegas. Characterization of the handwriting skills as a biomarker for Parkinson disease. In *IEEE Intl. Conf. on Automatic Face and Gesture Recognition (FG)*, May 2019. 2

[9] C. De Stefano, F. Fontanella, D. Impedovo, G. Pirlo, and A. S. di Freca. Handwriting analysis to support neurodegenerative diseases diagnosis: A review. *Pattern Recognition Letters*, 121:37–45, 2019. 2

[10] V. Dentamaro, D. Impedovo, and G. Pirlo. Gait analysis for early neurodegenerative diseases classification through the kinematic theory of rapid human movements. *IEEE Access*, 8:193966–193980, 2020. 2

[11] X. Dong, S. Yu, X. Weng, S. Wei, Y. Yang, and Y. Sheikh. Supervision-by-registration: An unsupervised approach to improve the precision of facial landmark detectors. In *Proc. CVPR*, pages 360–368, 2018. 3

[12] P. Ekman. Strong evidence for universals in facial expressions: a reply to Russell's mistaken critique. *Psychological Bulletin*, 115 2:268–87, 1994. 3

[13] M. Faundez-Zanuy, J. Fierrez, M. A. Ferrer, M. Diaz, R. Tolosana, and R. Plamondon. Handwriting biometrics: Applications and future trends in e-security and e-health. *Cognitive Computation*, August 2020. 2

[14] J. Fierrez. *Adapted Fusion Schemes for Multimodal Biometric Authentication*. PhD Thesis, Univ. Politecnica de Madrid, Spain, May 2006. 3

[15] J. Fierrez, A. Morales, R. Vera-Rodriguez, and D. Camacho. Multiple classifiers in biometrics. part 1: Fundamentals and review. *Information Fusion*, 44:57–64, Nov. 2018. 3, 6, 7

[16] E. Friesen and P. Ekman. *Facial Action Coding System: A technique for the measurement of facial movement*. CA: Consulting Psychologist Press, 1978. 2

[17] H. Gaßner, S. Steib, S. Klamroth, C.F. Pasluosta, W. Adler, B.M. Eskofier, K. Pfeifer, J. Winkler, and J. Klucken. Perturbation treadmill training improves clinical characteristics of gait and balance in Parkinson's disease. *Journal of Parkinson's Disease*, 9(2):413–426, 2019. 2

[18] C. G. Goetz et al. Movement disorder society task force report on the Hoehn and Yahr staging scale: status and recommendations the movement disorder society task force on rating scales for Parkinson's disease. *Movement Disorders*, 19(9):1020–1028, 2004. 5

[19] C. G. Goetz et al. Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results. *Movement Disorders*, 23(15):2129–2170, 2008. 1

[20] L. F. Gomez-Gomez, A. Morales, J. Fierrez, and J. R. Orozco-Arroyave. Exploring facial expressions and affective domains for Parkinson detection. *arXiv preprint arXiv:2012.06563*, 2020. 3, 6

[21] E. Gonzalez-Sosa, J. Fierrez, R. Vera-Rodriguez, and F. Alonso-Fernandez. Facial soft biometrics for recognition in the wild: Recent works, annotation and COTS evaluation. *IEEE Trans. on Information Forensics and Security*, 13(8):2001–2014, August 2018. 2

[22] E. Gonzalez-Sosa, R. Vera-Rodriguez, J. Fierrez, and J. Ortega-Garcia. Exploring facial regions in unconstrained scenarios: Experience on ICB-RW. *IEEE Intelligent Systems*, 33(3):60–63, May 2018. 2

[23] A. Grammatikopoulou, N. Grammalidis, S. Bostantjopoulou, and Z. Katsarou. Detecting hypomimia symptoms by selfie photo analysis: for early Parkinson disease detection. In *Proc. ACM Intl. Conf. on Pervasive Technologies Related to Assistive Environments*, pages 517–522, 2019. 2

[24] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proc. CVPR*, 2016. 3

[25] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam. MobileNets: Efficient convolutional neural networks for mobile vision applications. *arXiv:1704.04861*, 2017. 4

[26] Y. Kartynnik, A. Ablavatski, I. Grishchenko, and M. Grundmann. Real-time facial surface geometry from monocular video on mobile GPUs. *arXiv:1907.06724*, 2019. 3, 4

[27] C. Lugaresi et al. MediaPipe: A framework for building perception pipelines. *arXiv:1906.08172*, 2019. 3

[28] L. Moro-Velazquez, J. A Gomez-Garcia, J. I. Godino-Llorente, F. Grandas-Perez, S. Shattuck-Hufnagel, V. Yagüe-Jimenez, and N. Dehak. Phonetic relevance and phonemic grouping of speech in the automatic detection of Parkinson's disease. *Scientific Reports*, 9(1):1–16, 2019. 2

[29] J. R. Orozco-Arroyave, J. C. Vásquez-Correa, J. F. Vargas-Bonilla, R. Arora, N. Dehak, P. S. Nidadavolu, et al. Neurospeech: An open-source software for Parkinson's speech analysis. *Digital Signal Processing*, 77:207–221, 2018. 2

[30] A. Pena, J. Fierrez, A. Lapedriza, and A. Morales. Learning emotional-blinded face representations. In *IAPR Intl. Conf. on Pattern Recognition (ICPR)*, January 2021. 3

[31] W. Poewe and G. Wenning. The differential diagnosis of Parkinson's disease. *European journal of neurology*, 9:23–30, 2002. 1

[32] C. Richardson, D. Bowers, R. Bauer, K. M. Heilman, and C. M Leonard. Digitizing the moving face during dynamic expressions of emotion. *Neuropsychologia*, 38:1026–1037, 2000. 2

[33] C. D. Rios-Urrego, J. C. Vásquez-Correa, J. F. Vargas-Bonilla, E. Nöth, F. Lopera, and J. R. Orozco-Arroyave. Analysis and evaluation of handwriting in patients with Parkinsonnson's disease using kinematic, geometrical, and non-linear features. *Computer Methods and Programs in Biomedicine*, 173:43–52, 2019. 2

[34] G. Simons, M. C. S. Pasqualini, V. Reddy, and J. Wood. Emotional and nonemotional facial expressions in people with Parkinson's disease. *Journal of the International Neuropsychological Society*, 10(4):521–535, 2004. 2

[35] S. Spasojević et al. Quantitative assessment of the arm/hand movements in Parkinson's disease using a wireless armband device. *Frontiers in Neurology*, 8(388):1–15, 2017. 2

[36] P. Tome, J. Fierrez, R. Vera-Rodriguez, and D. Ramos. Identification using face regions: Application and assessment in forensic scenarios. *Forensic Science International*, (233):75–83, 2013. 2

[37] P. Tome, R. Vera-Rodriguez, J. Fierrez, and J. Ortega-Garcia. Facial soft biometric features for forensic face recognition. *Forensic Science International*, 257:171–284, December 2015. 2