



Detection of hypomimia in patients with Parkinson's disease via smile videos

Ge Su¹, Bo Lin¹, Jianwei Yin¹, Wei Luo², Renjun Xu³, Jie Xu², Kexiong Dong⁴

¹College of Computer Science and Technology, Zhejiang University, Hangzhou, China; ²Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China; ³Center for Data Science, Zhejiang University, Hangzhou, China; ⁴Technical Department, Hangzhou Healink Technology Corporation Limited, Hangzhou, China

Contributions: (I) Conception and design: G Su; (II) Administrative support: J Yin, J Xu, K Dong; (III) Provision of study materials or patients: W Luo; (IV) Collection and assembly of data: B Lin; (V) Data analysis and interpretation: G Su, R Xu; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Jie Xu. Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou 310009, China. Email: 2202113@zju.edu.cn.

Background: Parkinson's disease (PD) is a neurodegenerative disease characterized by the impairment of facial expression, known as hypomimia. Hypomimia has serious impacts on patients' ability to communicate, and it is difficult to detect at early stages of the disease. Furthermore, due to bradykinesia or other reasons, it is inconvenient for PD patients to visit the hospital. Therefore, it is appealing to develop an auxiliary diagnostic method that remotely detects hypomimia.

Methods: We proposed an automatic detection system for Parkinson's hypomimia based on facial expressions (DSPH-FE). DSPH-FE provides a convenient remote service for those who potentially suffer from hypomimia and only requires patients to input their facial videos. Specifically, patients can detect hypomimia through two aspects: geometric features and texture features. Geometric features focus on visually representing structures of facial muscles. Facial expression factors (FEFs) are used as the first metric to quantify the current activation state of the facial muscles. Facial expression change factors (FECFs) are subsequently used as the second metric to calculate the moving trajectories of the activation states in the videos. Geometric features primarily concentrate on spatial information, with little involvement of temporal information. Thus, the extended histogram of oriented gradients (HOG) algorithm is introduced. This algorithm can extract texture features within multiple continuous frames and incorporate the temporal information into the features. Finally, these features are applied to four machine learning algorithms to model the relationship between these features and hypomimia.

Results: The DSPH-FE detection system achieved the best performance when concatenating geometric features and texture features, resulting in a F1 score of 0.9997. The best F1 scores achieved with geometric features and texture features were 0.8286 and 0.9446, respectively. This indicated that both geometric features and texture features have an ability to predict hypomimia, and demonstrated that temporal information can boost the model performance. Thus, DSPH-FE is an effective supportive tool in the medical management of PD patients.

Conclusions: Comprehensive experiments demonstrated that proposed features fit well with real-world videos and are beneficial in the clinical diagnosis of hypomimia. In particular, hypomimia had a greater impact on eyes and mouths when patients are smiling.

Keywords: Parkinson's hypomimia (PD); detection system; facial expressions; geometric features; texture features

Submitted Jun 02, 2021. Accepted for publication Aug 20, 2021.

doi: 10.21037/atm-21-3457

View this article at: <https://dx.doi.org/10.21037/atm-21-3457>

Introduction

Parkinson's disease (PD) is the second most common neurodegenerative disease worldwide, characterized by irreversible functional impairment of facial expression, known as hypomimia (1,2). Patients living with hypomimia experience significant difficulties in adjusting their facial muscles to pose emotional expressions (3) so that they usually present a reduced voluntary and spontaneous facial expression, as if they have no interest in the surrounding environment. During daily communications, it even leads to a disconnect between PD patients and healthy people (4,5). The gradual nature of hypomimia onset often results in delayed awareness, prevention and treatment. The mean annual rate of deterioration in motor and disability scores can range from 2.4% to 7.4% (6). Moreover, bradykinesia, the most cardinal symptom of PD, impairs mobility (7,8), in which hospital visits can be a challenge for PD patients. Therefore, it is essential to develop an auxiliary method which can detect hypomimia in Parkinson's patients remotely and online.

To date, neurobiologists have not known exactly the cause of Parkinson's disease and have no direct means of diagnosis as well. Recently, some studies have reported that Parkinson's disease involves progressive degeneration of the nigrostriatal dopaminergic pathway, which further touches off the progressive degeneration of the somatomotor system (1,3). Hypomimia is one of these damaged mechanisms. Specifically, hypomimia results from the death of cells in the substantia nigra, a region of the midbrain, leading to a dopamine deficit. Therefore, levodopa is the most common treatment for Parkinson's disease.

Early diagnosis is critical for better treatment outcomes and improved quality of life (9). To detect hypomimia, most research regarding hypomimia has focused on analysis of facial expressions (10). Most investigations use wearable sensors (7,8,11) and statistical analyses (5,12) to compute the similarities and differences between PD patients and healthy participants. In these reports, various tasks have been designed, including expression identification tasks (13,14), emotion discrimination tasks (15-18), and emotion expression tasks (12). Through these tasks, statistical features such as reaction time and accuracy of responses, can be obtained. Professional indicators, such as the Ekman score (5) and the Euclidean distance (12), can be calculated to evaluate emotional expressivity using the foregoing statistical features. In addition, extracted features can be directly fed into machine learning models

(12,19-21) or supervised learning models (22,23), to distinguish patients with the hypomimia symptom from healthy people. However, to the best of our knowledge, the clinical diagnosis of hypomimia is still dependent on in-hospital clinical observations and several criteria (24-27). The present study proposed the use of an automatic detection system for Parkinson's hypomimia based on facial expressions (DSPH-FE). The DSPH-FE provides patients with a convenient and real-time online service for the detection of Parkinson's hypomimia and only requires patients to upload their facial videos.

In this study, smile videos were collected from healthy control (HC) participants and PD patients. After preprocessing the original data, the effective area of facial activity is determined and irrelevant objects, such as background and clothes, are filtered out. Based on the effective area, Parkinson's hypomimia is detected according to two aspects, namely, geometric features and texture features. The geometric features utilize facial keypoints to record the movement trajectory of the facial muscles. The standard expression model (SEM) is constructed as the reference model to represent natural expressions using neutral expressions. Next, facial expression factors (FEFs) are designed to characterize the current activation state of the facial muscles. This then allows the quantification of smile expressions and the SEM. Finally, facial expression change factors (FECFs) are used to measure the changes between smile expressions and the SEM by calculating the moving trajectories of these activation states. The FECFs represent spatial-based changes between the current smile expression and the reference expression, with little consideration for the temporal factors of expression changes (28,29). Thus, the extended histogram of oriented gradients (HOG) feature, HOG-Three Orthogonal Planes (HOG-TOP) (30), was introduced to extract texture features, both spatially and temporally. HOG features are able to characterize dynamic expression changes in a short period of time. The advantages of the above engineered features were then combined with machine learning. Four supervised learning methods (decision trees, random forests, Bayesian, and SVM) were used to explore the relationship between these hand-crafted features and Parkinson's hypomimia. The comprehensive experiments demonstrated that our engineered features fit well with real-world videos and is helpful to the clinical diagnosis of hypomimia.

This report describes the use of geometric features and texture features to quantify facial activities. These quantitative indicators may play a supportive role in the

Table 1 The basic characteristics of the study participants

Participants	Male	Female	Total	Age
HC subjects	21	18	39	56.59±10.08
PD patients	26	21	47	57.62±10.89

HC, healthy control; PD, Parkinson's disease.

medical decision-making process. Furthermore, a non-invasive system was developed to provide patients with a convenient and remote online service for the detection of Parkinson's hypomimia with satisfactory performance. We present the following article in accordance with the STARD reporting checklist (available at <https://dx.doi.org/10.21037/atm-21-3457>).

Methods

HC, healthy control; PD, Parkinson's disease.

Participants and setup

A previous report by Marsili *et al.* demonstrated that PD patients cannot express a natural smile (31). Therefore, in the current study, smile videos of PD patients and HC participants were analyzed to detect PD. A total of 39 HC participants and 47 PD patients were recruited for this study. There were 21 males and 18 females in the HC group, and 26 males and 21 females in the PD group. The average age of the HC participants and PD patients was 56.59±10.08 and 57.62±10.89 years, respectively (Table 1). The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). This study was approved by the Second Affiliated Hospital, Zhejiang University School of Medicine, and the patient informed consent was waived as all data were anonymized.

The smile videos were obtained in a 3 m² space with a solid color background. Such a space minimized irrelevant elements and prevented other unnecessary items from diverting the participant's attention. The light source was natural light, where participants felt comfortable and relaxed. The equipment was consisted of a smartphone (iPhone6sPlus, 1080P/60FPS) and a tripod placed 0.5–0.7 meters in front of the participants.

Data acquisition

Before shooting, the participants were requested to

remove all unrelated facial accessories such as glasses. The participants were seated to maintain a relatively fixed facial position and asked to avoid shaking their heads when the camera was turned on. The facial expressions of all participants were recorded in two parts. The first part is the preparation stage and participants were to remain calm and present a neutral expression with no extra emotion on his or her face. These expressions were suitable for constructing a standard expression model (SEM) salient to the individual and used as a benchmark to measure expression changes. The second part involved the expression generation phase, where participants were asked to smile for 3 to 5 seconds. Feature extraction data were collected during this stage. The videos were 60 frames per second with 1,920×1,080 pixels per frame.

Based on the clinical evaluation by a senior doctor, the experimental data were divided into two categories, HC participants and PD patients. PD patients were considered as positive samples and denoted by '1'. HC participants were considered as negative samples and denoted with '0' labels.

Data preprocessing

The video frames were preprocessed successively and some unqualified video clips were filtered out during the preprocessing. The preprocessing includes three parts, face detection, face alignment, and face normalization. MTCNN (multi-task cascaded convolutional neural network) was applied for face detection to remove irrelevant information, such as background, clothing, accessories, and chairs. In some videos, the human faces were rotated and this was not conducive to studying the unilateral illness. Thus, the human faces were rotated to keep the reference points (P_{le} and P_{re}) connected horizontally, where P_{le} and P_{re} refer to the reference points of the left and right eye, respectively. These rotary shafts passed through P_{le} and were perpendicular to the plane parallel to the video. P_{le} and P_{re} are calculated as follows:

$$P_{le}(x) = \frac{P_{12}(x) + P_{13}(x) + P_{15}(x) + P_{16}(x)}{4} \quad [1]$$

$$P_{le}(y) = \frac{P_{12}(y) + P_{13}(y) + P_{15}(y) + P_{16}(y)}{4} \quad [2]$$

$$P_{re}(x) = \frac{P_{18}(x) + P_{19}(x) + P_{21}(x) + P_{22}(x)}{4} \quad [3]$$

$$P_{re}(y) = \frac{P_{18}(y) + P_{19}(y) + P_{21}(y) + P_{22}(y)}{4} \quad [4]$$

where $P_i(x)$ and $P_i(y)$ represent the horizontal and vertical coordinate of the i -th facial keypoint. Points involved in these

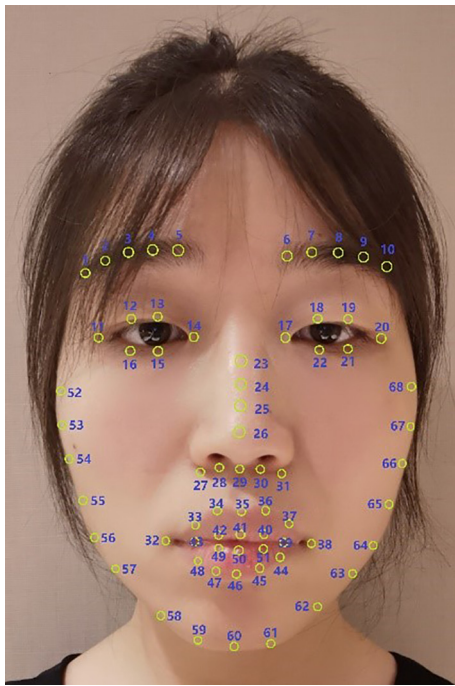


Figure 1 Facial keypoints. These 68 facial keypoints conform to the definition of previous face recognition technologies (32-36). This image is published with the patient/participant's consent.

formulas can be found in *Figure 1*. Lastly, face normalization initialized the face to a fixed size. The image was resized to 128×128 pixels for the convenience of calculations.

Data analysis

Compared to healthy individuals, PD patients have difficulty in the voluntary control of their facial muscles (12). When greeting with them, their families and acquaintances inevitably encounter an 'indifferent' attitude, resulting in a disconnect in daily communication. This is because their capacity for recognizing and analyzing emotion is restricted due to the changes in the internal parts of the brain (4). According to the Movement Disorder Society Unified Parkinson's Disease Rating Scale (MDS-UPDRS) (27), facial muscle bradykinesia is an important clinical indicator for assessing the severity of this disease. Rajnoha *et al.* parameterized facial expressions and used static images to identify facial muscle bradykinesia with an accuracy of 67.33% (37). Therefore, it may be possible to detect PD by recognizing facial muscle bradykinesia. To construct such a Parkinson's detection system, the following three problems need to be resolved:

- ❖ Facial nerves convey instructions from the brain and direct the multitude of facial muscles to stretch and form facial expressions. Therefore, identification of facial muscle bradykinesia requires the quantification of facial expressions.
- ❖ Slight details in facial expressions are difficult to measure. When the human skin is stretched by facial muscles, it tends to produce slight facial textures such as fine lines. Thus, extracting the textural information may provide complimentary information.
- ❖ PD may be unilateral. In clinical diagnosis, facial muscle bradykinesia often can be observed on only one side of the face, and it may affect different areas of the face. Thus, it requires our engineered features to support subregion representation.

To solve these issues, this study used geometric and texture features to evaluate the physiological parameters related to facial expressions of Parkinson's disease patients.

Geometric feature design

It is critical to quantify facial expressions. A total of 68 facial keypoints were extracted from the preprocessed images. These keypoints are denoted as P , as shown in *Figure 1*. Any facial keypoint $p \in P$, needs to satisfy the following four prerequisites. Prerequisite 1: given a facial keypoint $p \in P$, another point is symmetrical to the current point, whose symmetry axis is the bridge of the nose.

If a point $p_i \in P$, and itself are symmetrical, it is on the symmetry axis. Otherwise, there is another symmetrical point $p_j \in P$. Considering the human nose as the symmetry axis, P is divided equally into two parts. One part is located on the left side of human face, recording their current position. The other is on the right side of human face while keeping in memory the current position on the right. This will be beneficial for detecting unilateral onset.

Prerequisite 2: any point $p \in P$ is closely related to one of the facial muscles.

Since these points are closely related to facial muscles, each point is used to record a determined position of facial muscles at a certain time. As time goes by, these points are connected in sequence and a series of movement trajectories can be assembled. The facial expressions will then be simulated simultaneously.

Prerequisite 3 (dynamic-flexible): a point $p \in P$ is dynamic-flexible if it can move relatively freely.

Prerequisite 4 (relatively-fixed): a point is relatively-fixed

Table 2 Definitions of facial expression factors around certain facial organs

Subregions	FEFs
Eye factor	
Eye corner	D(11,14), D(17,20), D(14,17)
Eyelid	D(12,16), D(12,16), D(18,22), D(19,21)
Eye angle	$\theta(13,14,15)$, $\theta(18,17,22)$, $\theta(12,11,16)$, $\theta(19,20,21)$, $\theta(14,23,5)$, $\theta(17,23,6)$, $\theta(17,14,5)$, $\theta(14,17,6)$
Eyebrow factor	
Eyebrow	D(5,6), D(5,14), D(6,17), D(1,11), D(10,20)
Eyebrow angle	$\theta(1,3,5)$, $\theta(6,8,10)$
Nose factor	
Nose	D(27,31), D(27,29), D(29,31), D(26,27), D(26,31), D(26,29), D(23,26), D(25,26)
Nose angle	$\theta(27,16,31)$
Lip factor	
Lip	D(32,38), D(32,35), D(32,46), D(38,35), D(38,46)
Outer lip	D(34,47), D(36,45)
Inner lip	D(42,49), D(41,50), D(40,51)
Cross	D(42,47), D(40,45), D(34,49), D(36,51)
Lip corner	$\theta(33,32,48)$, $\theta(37,38,44)$, $\theta(42,32,49)$, $\theta(40,38,51)$, $\theta(34,35,36)$, $\theta(33,34,35)$, $\theta(35,36,37)$, $\theta(47,46,45)$
Chin factor	
Chin	D(56,59), D(61,64), D(59,61), D(60,26), D(56,64)
Chin corner	$\theta(57,59,60)$, $\theta(63,61,60)$, $\theta(59,60,61)$, $\theta(54,56,59)$, $\theta(66,64,61)$

FEFs, facial expression factors.

if it never moves as the expression occurs.

Given a point $p_i \in P$, it is either dynamic-flexible or relatively-fixed. Relatively-fixed points are viewed as baseline points, while dynamic-flexible points are suitable for quantifying the distance from relatively-fixed points. The amplitude of facial muscle movement can then be measured. For instance, p_{25} and p_{26} are relatively-fixed while p_{32} and p_{38} are very flexible. The distance between p_{26} and p_{32} skillfully quantifies the amplitude of the left mouth corner.

With the help of facial keypoints, the trajectory of facial muscles can be measured. If we regard a facial expression as an activation state, it is one of the toughest challenges to characterize the current facial expression. In this paper, the facial expressions factors (FEFs) are used to quantify these activation states.

As mentioned above, unilateral disease requires that the engineered features have to be element-level features which are relatively independent and freely composable. It

implies that each FEF can be calculated individually and only represents a small range of activation region. The simpler the calculation is, the fewer associations are likely to occur, which means stronger independence. Therefore, the human face was divided into different regions from global to local, including eyebrow, eye, nose, lip, and chin. Simple Euclidean distances and some angles were applied to measure the activation state of the current facial organs. *Table 2* lists the specific definitions of the FEFs using facial keypoints in *Figure 1*. Symmetric and discrete variables support subregion representation, relatively independent and freely composable, which solves the third problem. These discrete variables were used to assemble a complete facial expression representation while keeping the local characteristics. For example, $D(13,15)$ can satisfactorily measure the blink. In the table, $D(a,b)$ is the Euclidean distance between two points (point a, point b), and $\theta(a,b,c)$ is the angle between two lines (a line passing through point a and point b, another line passing through point b and

point c).

$$D(a,b) = \sqrt{(P_a(x) - P_b(x))^2 + (P_a(y) - P_b(y))^2} \quad [5]$$

$$\theta(a,b,c) = \arccos \frac{\overline{abc}}{|ba||bc|} = \arccos \frac{\alpha \cdot \beta}{D(b,a) \cdot D(b,c)} \quad [6]$$

$$\alpha = (P_a(x) - P_b(x), P_a(y) - P_b(y)) \quad [7]$$

$$\beta = (P_c(x) - P_b(x), P_c(y) - P_b(y)) \quad [8]$$

$$D(a,b) = D(b,a) \quad [9]$$

$$D(b,c) = \sqrt{(P_c(x) - P_b(x))^2 + (P_c(y) - P_b(y))^2} \quad [10]$$

where α represents the vector from point b to point a. Similarly, β is the vector from point b to point c.

While the FEFs can defined the current activation state, the relative change between two states must be considered. Changes in expressions mean that human faces leap from one activation state to another. This paper adopted a strategy similar to Bandini *et al.* (4). A standard facial model, showing a calm and natural expression, was constructed as a reference model for each person. An average activation state was calculated using Platts analysis. We iterate it over a series of static images showing calm expressions until a termination trigger is encountered. Notably, the closer the image is to the termination trigger, the bigger its contribution to the SEM. Coincidentally, the starting point of the expression generation was close to the trigger. This suggested that the SEM was also close to the initial expression, and this was beneficial for measuring the changes between the current expression and the initial expression. The algorithm is summarized in Algorithm 1.

Algorithm 1: Standard Expression Model Construction Algorithm

Input: set of frames F, Number of iterations N;

Output: SEM

```

1 First_F ← Read first frame
2 Base_img ← Preprocess (First_F)
3 While N > 0 do
4   New_F ← Read next frame
5   New_img ← Preprocess (New_F)
6   Base_img ← Average (Base_img, New_img)
7   N--
8 SEM ← Base_img
9 Return SEM

```

During the preparation stage, participants expressed a neutral expression with no extra emotion on his or her face. These expressions were suitable for constructing an SEM as a reference model. Through Algorithm 1, a SEM model was generated with an FEF set of Γ^S and DIS^S .

$$\gamma_i^S = \theta(a_i^S, b_i^S, c_i^S) = \arccos \frac{\alpha_i^S \cdot \beta_i^S}{D(b_i^S, a_i^S) \cdot D(b_i^S, c_i^S)} \quad [11]$$

$$dis_i^S = D(a_i^S, b_i^S) \quad [12]$$

$$\Gamma^S = \{\gamma_1^S, \gamma_2^S, \dots, \gamma_m^S\} \quad [13]$$

$$DIS^S = \{dis_1^S, dis_2^S, \dots, dis_n^S\} \quad [14]$$

$$SEM = \Gamma^S \cup DIS^S = \{\gamma_1^S, \gamma_2^S, \dots, \gamma_m^S, dis_1^S, dis_2^S, \dots, dis_n^S\} \quad [15]$$

where Γ^S is a set of FEF variables with respect to the SEM angles, γ_i^S is the *i*th angle variables on SEM; DIS is a set of FEF variables with respect to the SEM distances, dis_i^S is the *i*th distance variables on SEM; SEM consists of Γ^S and DIS^S , *m* and *n* are respectively the number of angle variables and distance variables.

When a person smiles, the FEFs $\{\Gamma^C \cup DIS^C\}$ quantify his or her expression while the SEM $\{\Gamma^S \cup DIS^S\}$ is created as a true portrayal of their neutral expressions. Naturally, the distance between a set of FEFs and its SEM is considered the change in the activation states. This distance is called the facial expression change factor (FECF).

$$\begin{aligned} FECF &= \{\Gamma^C \cup DIS^C\} - \{\Gamma^S \cup DIS^S\} \\ &= \{\gamma_1^C - \gamma_1^S, \gamma_2^C - \gamma_2^S, \dots, \gamma_m^C - \gamma_m^S\} \\ &\quad \cup \{dis_1^C - dis_1^S, dis_2^C - dis_2^S, \dots, dis_n^C - dis_n^S\} \end{aligned} \quad [16]$$

where γ_i^C are the FEF variables of the current activation state with respect to angles, and dis_i^S are the FEF variables of the current activation state with respect to the distances. FECF was effective at quantifying the changes between different facial expressions, thus solving the first problem above.

Texture feature design

While the above geometric features capture some facial muscles, other facial muscles are not represented. Thus, geometric features ignore a large quantity of details. It is noted that the change in expressions is a continuous dynamic process. Although the FECF measures the change between the current facial expression and the reference model, it does not take into account temporal factors (37),

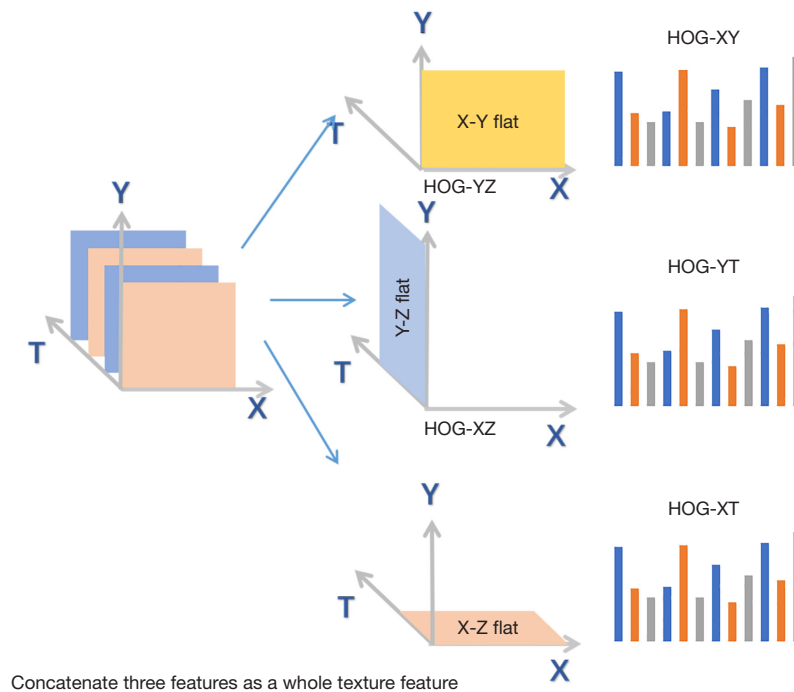


Figure 2 The extended HOG algorithm. The video information is regarded as a three-dimensional information space, which is represented by a space made up of X, Y, and T axis. The XY plane represents spatial information, and the XT plane or YT plane mainly represents temporal information and linear spatial information.

also claimed that the image sequence could further boost the performance. Therefore, we introduce the extended HOG algorithm (30) which combines temporal dimensions and spatial dimensions. As shown in *Figure 2*, the video consists of a sequence of images, denoted as the X, Y, and T axis. The X and Y axes jointly constitute a two-dimensional space while the T axis represents the time dimension. Conventional HOG algorithm is only performed in the XY plane, that is, HOG-XY features are formed to capture spatial-based texture information. The extended HOG algorithm is performed in the YT direction and the XT direction to construct HOG-YT features and HOG-XT features. These three features were combined into a complete HOG feature, both spatially and temporally, thereby enriching abundant time and detailed information into our system, and thus, solving the second problem. The specific algorithm is shown in Algorithm 2.

```

1   for n = 1 : N do
2       read n-th video → Video
3       while getFrame(Video) == True do
4           face detection
5           face alignment
6           crop face picture and resize to 128*128
7       get three-dimensional arrays for each video
8       compute HOG-XY, HOG-XT, HOG-YT from XY, XT
          and YT plane
9       concatenate three features
10      get feature vectors for each video
    
```

$$\begin{aligned}
 HOG &= HOG - XY \cup HOG - XT \cup HOG - YT \\
 &= \{HOG - XY_i | i \in 1, 2, \dots, k\} \\
 &\quad \cup \{HOG - XT_i | i \in 1, 2, \dots, q\} \\
 &\quad \cup \{HOG - YT_i | i \in 1, 2, \dots, t\}
 \end{aligned}
 \tag{17}$$

where k, q, t are respectively the number of HOG features

Algorithm 2: Texture Feature Extraction Algorithm

Input: videos from HC or PD, number of videos N,
Output: texture feature vectors for each video

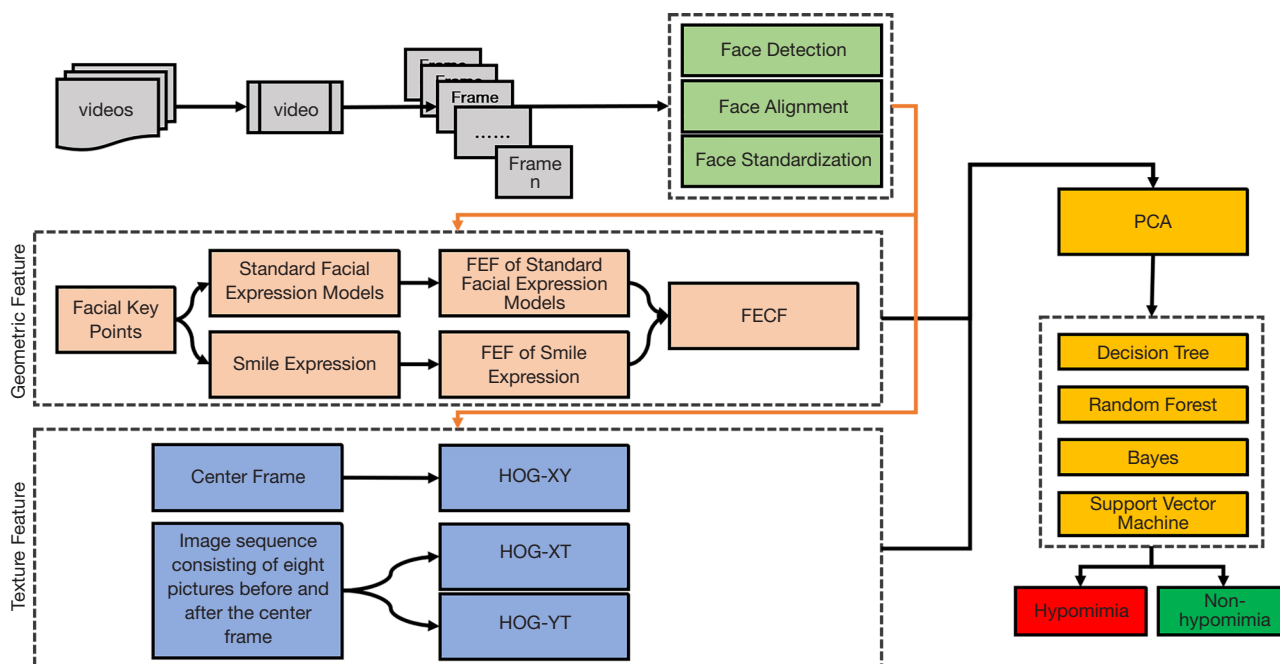


Figure 3 The overall framework of the DSPH-FE. The geometric features process a single image while the texture features process the image sequence. DSPH-FE, Parkinson's hypomimia based on facial expressions.

from the XY, XT, YT plane.

Model implementation

In this section, the geometric features and texture features were constructed. Specially, the SEM was constructed iteratively using three hundred images at the preparation stage while 63 FEFs ($m=24, n=39, m+n=63$) were calculated for each activation state. The distance from the SEM was computed as expression changes. As for texture features, the cropped images (128×128 pixels) were further cut into 16×16 cells. Each cell, which contained 8×8 pixels, was characterized into a nine-dimensional vector. A 2304-dimensional vector ($k=16 \times 16 \times 9=2,304$) was then assembled in the XY plane. Similarly, the other two planes were constructed. Notably, there were some differences between the XT (YT) plane and the XY plane. The XT (or YT) plane spanned eight continuous images on T-axis and was scanned from top to bottom (from left to right). Hence, each layer of the XT plane consisted of 16×1 cells whose resolution was 8×8 pixels. Scanning each layer from top to bottom, a 2,304-dimensional vector was assembled ($q/t=16 \times 1 \times 16 \times 9=2,304$). In the end, these vectors were sequentially concatenated to a 6912-dimensional vector ($k+q+t=2,304 \times 3$).

Principal component analysis (PCA) was used to address the issue of dimensionality. The variables with significant correlations were retained while maintaining 95% of the original information. Finally, the training set accounted for 80% of the samples, and four machine learning methods were used to model the relationship between facial expression features and Parkinson's bradykinesia. In the comprehensive experiments, the dimensionality of geometric features (GFs), texture features (TFs), and fusion features (FFs) were reduced to 25, 2,559, and 2,574, respectively. The overall framework of the DSPH-FE is shown in *Figure 3*.

Statistical Analysis

In this paper, the proposed features were evaluated through the F1 score.

The evaluation indicators are defined as follows:

$$\text{Precision} = \frac{TP}{TP + FP} \quad [18]$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad [19]$$

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad [20]$$

Table 3 The performance of different features

Methods	Precision	Recall	F1
GF + Bayesian	79.39%	61.83%	0.6952±0.14
GF + Decision Tree	77.17%	67.32%	0.7191±0.07
GF + SVM	77.42%	81.31%	0.7931±0.03
GF + Random Forest	77.06%	89.61%	0.8286±0.06
TF + Bayesian	94.16%	66.95%	0.7826±0.14
TF + Decision Tree	84.63%	86.47%	0.8554±0.02
TF + Random Forest	88.51%	95.79%	0.9200±0.00
TF + SVM	96.11%	92.87%	0.9446±0.01
FF + Bayesian	99.95%	62.32%	0.7677±0.14
FF + Decision Tree	99.79%	81.07%	0.8946±0.01
FF + Random Forest	99.82%	100.00%	0.9991±0.02
FF + SVM	99.94%	100.00%	0.9997±0.00

GF, geometric features; SVM, support vector machine; TF, texture features; FF; fusion features.

where TP and TN are the true prediction for patients and normal people, FP represents the false prediction for patients.

For the statistical analysis of features, we utilize GBDT (gradient boosted decision trees) algorithm to evaluate the importance of engineered features. On the other hand, the receiver operating characteristic (ROC) curves present the feature adaptation on the GBDT model.

Results

Five cross-validation experiments were conducted and the average results are reported in *Table 3*. Fusion features combine geometric features and texture features.

The fusion feature achieved the best results in three classifiers, and thus may greatly support medical decision-making in the clinical setting. However, geometric features were unsatisfactory, with F1 scores between 0.69 and 0.83. In comparison, the F1 scores for the texture features were 0.78–0.95. This suggested that texture features have a stronger ability to predict PD than geometric features. This also verified that dynamic information can boost the performance of Parkinson's detection. In addition, although the geometric features did not achieve excellent performance in terms of accuracy, the visualization results provide interesting interpretation that will be elaborated in detail in the next section.

Compared to individual geometric features and texture

features, the fusion features had higher accuracy and stability with a F1 score of 0.9997. That is because fusion features incorporate the advantages of geometric features and texture features. Interestingly, the Bayesian classifier showed poor performance. This is because Bayesian classifiers are suitable for small-scale datasets, without a number of attributes. With the increase of attributes, it is difficult to satisfy the conditional independence assumption.

Discussion

Importance analysis

To evaluate the importance of engineered features, GBDT algorithms were performed to calculate the weights of the GFs, TFs, and FFs. Analyses were carried out three times for each set of features. As shown in *Table 4*, when only using GFs, Parkinson's hypomimia was primarily identified using variables based on the human eyes and mouth, such as D(19,21), D(13,15), and $\theta(47,46,45)$. When using TFs, areas of hypomimia-interest were located at the chin and the eyes. The encoded areas 2241, 603, 759, and 71 determined at least 81 percent of the decision-making process. After combining TFs and GFs, the above remained true for FFs. In particular, TFs dominated the direction of the classification, which was consistent with the above performance analysis. Several temporal features additionally assisted the detection of hypomimia, including the encoded areas 2857 and 6376. *Figure 4* shows the position of the top ten most significant GFs and TFs depicted in red and blue, respectively.

Feature adaptation on model

Figure 5 shows three groups of ROC curves with respect to the nine experiments in *Table 4*. The fused features achieved the best adaptation on the GBDT model. The area under ROC curve accounts for more than 94.3 percent of the whole areas. Among these graphs, though GF has not shown stable performance, TF combined with GF improved by nearly 4 points (*Figure 5*). This suggested that TF has better adaptation on the GBDT model, and GF assisted in improving the performance of TF.

Physical meaning of geometric features

This paper analyzed the FEFCF changes [0(33,32,48) and $\theta(37,38,44)$] in HC participant number 0838 (0838HC)

Table 4 The importance weights of the engineered features

Features	Accuracy	1	2	3	4	5	6	7	8	9	10
GF	0.7450	D(19,21)	D(13,15)	D(11,14)	$\theta(47,46,45)$	D(18,22)	D(12,16)	$\theta(40,38,51)$	D(14,17)	D(17,21)	D(27,29)
		0.1071	0.1042	0.1036	0.1036	0.0691	0.0649	0.0567	0.0563	0.0539	0.0425
	0.7207	D(19,21)	D(13,15)	$\theta(47,46,45)$	D(14,17)	D(12,16)	D(27,31)	$\theta(40,38,51)$	D(17,20)	D(60,26)	D(11,14)
		0.3191	0.1219	0.0755	0.0724	0.0568	0.0469	0.0460	0.0387	0.0272	0.0223
	0.7082	D(19,21)	D(13,15)	D(14,17)	$\theta(47,46,45)$	D(32,38)	$\theta(40,38,51)$	D(27,29)	D(27,31)	D(18,22)	D(17,20)
		0.2814	0.1237	0.0848	0.0655	0.0448	0.0438	0.0359	0.0352	0.0330	0.0314
TF	0.9460	71	2,241	2,156	759	603	278	1,313	2,857	998	4,918
		0.5581	0.1903	0.1172	0.0423	0.0345	0.0252	0.0103	0.0037	0.0032	0.0029
	0.9621	2,241	603	759	71	2,156	998	2,857	5,656	6,376	4,918
		0.5969	0.1576	0.1341	0.0438	0.0170	0.0150	0.0069	0.0043	0.0038	0.0030
	0.9646	2,241	603	759	71	2,156	2,857	6,376	3,190	278	639
		0.4955	0.2714	0.1125	0.0413	0.0300	0.0092	0.0065	0.0045	0.0039	0.0038
FF	0.9869	2,241	71	759	2,156	603	998	639	413	4918	2,857
		0.4549	0.3035	0.1044	0.0696	0.0153	0.0105	0.0101	0.0059	0.0056	0.0031
		D(17,20)	D(13,15)	D(11,14)	D(60,26)	D(12,16)	$\theta(40,38,51)$	D(32,38)	D(27,31)	D(27,29)	D(14,17)
		8.356E-04	1.086E-04	4.087E-05	3.386E-05	2.394E-05	1.481E-05	1.446E-05	4.148E-06	1.905E-06	6.953E-07
	0.9822	2,241	71	603	759	2,156	278	6,376	998	2,857	1,124
		0.3773	0.2860	0.1199	0.0986	0.0706	0.0104	0.0084	0.0077	0.0042	0.0031
		$\theta(47,46,45)$	D(11,14)	D(17,20)	D(18,22)	$\theta(40,38,51)$	D(12,16)	D(13,15)	D(27,29)	D(14,17)	D(27,31)
		4.454E-04	1.849E-04	1.721E-04	1.077E-04	1.028E-04	4.199E-05	3.917E-05	2.360E-05	2.193E-05	1.923E-05
	0.9899	603	2,241	71	759	2,156	2,857	998	6,376	639	5,345
		0.5134	0.2825	0.0684	0.0573	0.0443	0.0047	0.0046	0.0041	0.0038	0.0031
	D(32,38)	$\theta(40,38,51)$	D(60,26)	D(11,14)	D(18,22)	D(27,29)	D(13,15)	D(17,20)	$\theta(47,46,45)$	D(12,16)	
	1.155E-04	6.538E-05	4.465E-05	4.365E-05	4.302E-05	4.263E-05	9.909E-06	5.448E-06	5.448E-06	5.788E-07	

GF, geometric features; TF, texture features; FF, fusion features.

and PD patient number 0508 (0508PD). The FECF variables quantitatively describe the curvature changes in the human lips. As shown in *Figure 6*, the abscissa is the image sequence, and the ordinate represents the FECF in $\theta(33,32,48)$ [or $\theta(37,38,44)$]. The FECF graphs measure the distance between FEF and SEM, and retains the magnitude of the change, where its abscissa represents the SEM. The negative values below the abscissa indicate that the curvature of the mouth corner is reduce relative to the SEM. The positive values indicate that the curvature of the mouth corner is increased relative to the SEM. If the curve rises first and then falls, the mouth corner changes once

significantly. Three sharp peaks can be observed in FECF-0838, suggesting that the mouth angle changed greatly three times. These sharp peaks are mainly located below the abscissa, which suggested that the lips were stretched when the mouth was closed. The video of participant 0838HC showed that he smiled three times, in agreement with our interpretation. The visualized result was highly consistent with the diagnostic results. Interestingly, both sides of the mouth corners showed similar frequencies and amplitudes. Specifically, there were three obvious changes and their amplitudes were maintained at 15, 22, and 16, respectively, whether it was the left or right side of the mouth.

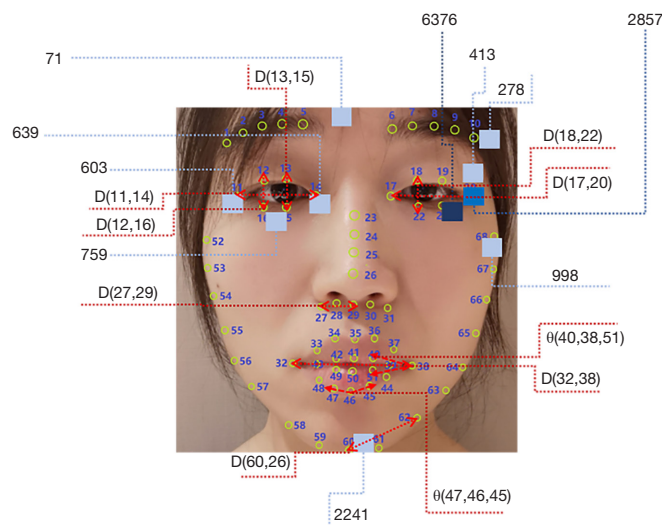


Figure 4 The area of interest for hypomimia. The blue square blocks represent the top 10 significant texture features, the darker blue squares indicate areas with time dimension (XT plane or YT plane). The light blue belongs to XY plane, the blue is XT plane, the darker blue is YT plane. The red arrows represent the top 10 significant facial expression change factors (FECFs) corresponding to the facial expression factors (FEFs). This image is published with the patient/participant's consent.

Therefore, the left side of the participant 0838HC was as flexible as the right side. The PD patient 0508PD, aged 64 years, was diagnosed with PD by the doctor. His onset side was on the right. In *Figure 6* (FECF-0508), the solid pink line represents the FECF maps of patient 0508PD in $\theta(33,32,48)$, corresponding to the left side of the mouth corner in the video. Similarly, the solid blue line represents $\theta(37,38,44)$, corresponding to the right mouth corner in the video. The direction of the participants is opposite to that presented in the video, due to the mirroring effect in the camera. Thus, the solid pink line is the right side of the 0508PD while the solid blue line is the left corner of the 0508PD. The solid blue line frequently exceeded the solid pink line by a large margin, which suggested that the patient's right muscles were not as flexible as his left muscles. On the other hand, compared to the FECF-0838, 0508PD had no apparent expression changes as there are no apparent sharp peaks in his FECF graph. In practical applications, the diagnostic result was consistent with the doctor's opinion. Moreover, this patient's PD showed unilateral onset.

In summary, the above analyses support the use of the features identified in this report for analyzing facial activities.

Physical meaning of the texture features

As mentioned above, the XT and YT planes are capable of characterizing the dynamic process of facial expression changes. We examined the XT plane in a case where the Y-axis coordinate was fixed at 40. In *Figure 7*, the first graph shows the XT plane ($Y=40$) of HC participant number 0842 (0842HC), and the second graph shows the XT plane ($Y=40$) of PD patient number 0451 (0451PD). After preprocessing, the facial images of the texture features were all scaled to 128×128 pixels. When $Y=40$, the XT plane represents the change of the eyes over time. As can be easily seen from *Figure 7*, 0842HC had more ripples and a larger ripple width. On the contrary, 0451PD had less ripples in the XT plane and a smaller ripple width. When participants blink once, the XT plane will generate a ripple. Thus, the number of ripples represents the number of blinks and the ripple width reflects the blink duration. The graph shows that the HC participant's eyes have more frequent and flexible movement, and their blink duration is larger. On the contrary, the PD patient had a lower blink rate and a shorter blink duration. This is because it is difficult for Parkinson's patients to maintain a blink state and they will unconsciously return to their original state. This suggested that the HC participant could flexibly use his eye muscles to generate movements. However, the PD patient experienced

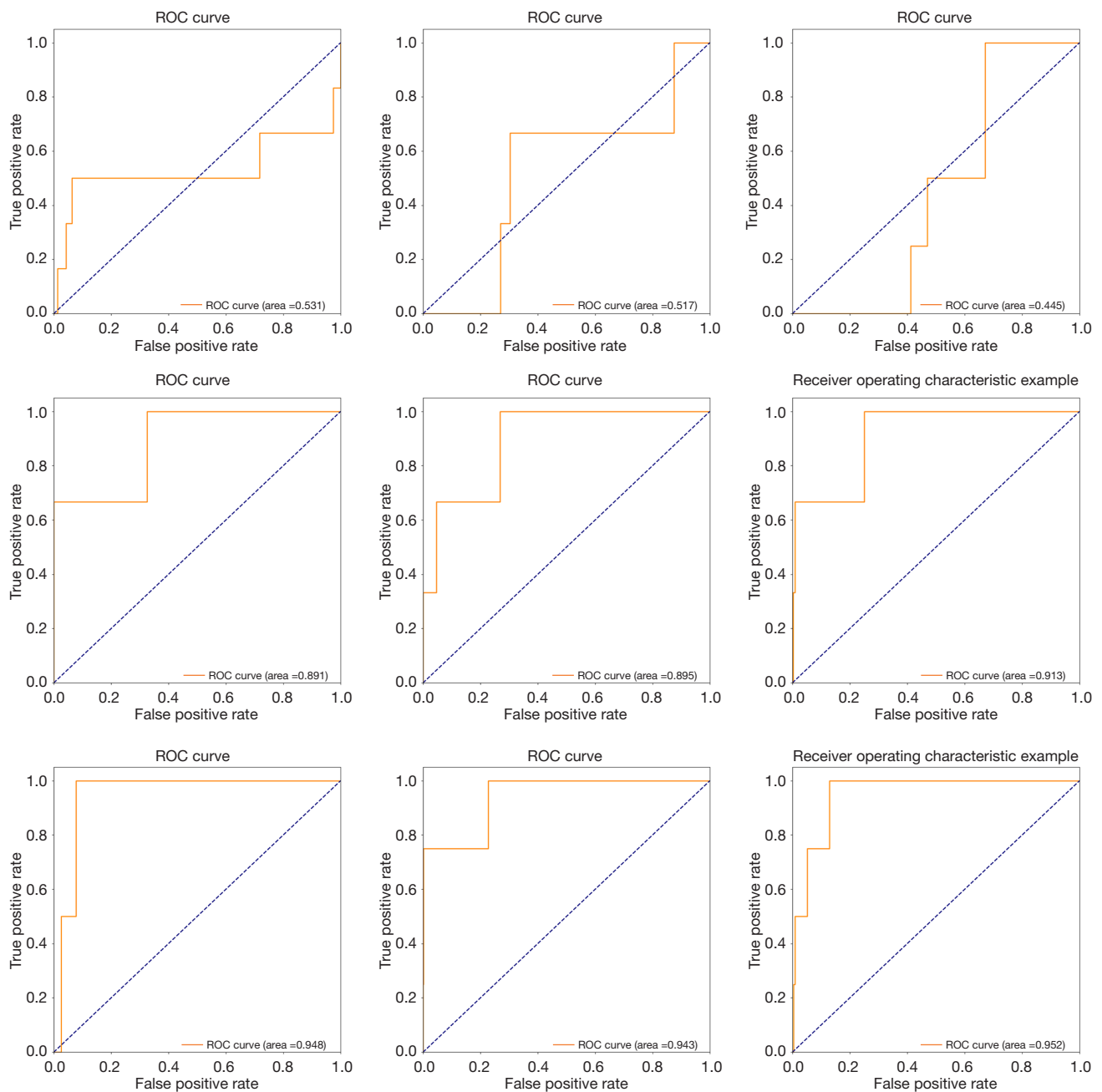


Figure 5 Feature adaptation on the model. The first row corresponds to the geometric features (GFs) in Table 4, the second row corresponds to texture features (TFs) in Table 4, and the third row corresponds to fusion features (FFs) in Table 4. Experiments were performed three times for each group of features.

difficulty in moving his eye muscles.

Conclusions

Hypomimia in PD can severely impact the patient's quality

of life. To date, most research related to Parkinson's hypomimia detection has primarily relied upon sensor-based wearable devices. However, it is difficult for patients to access this service due to the high costs and other factors. Thus, this paper proposed a remote and online system for

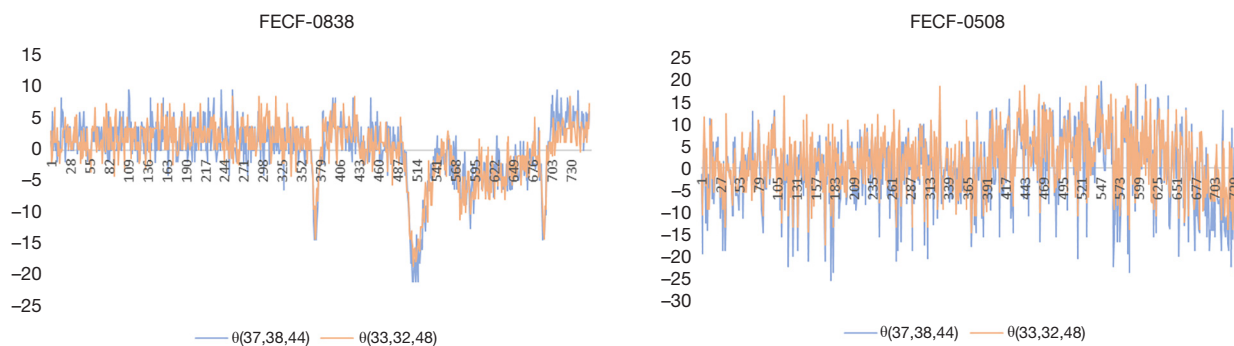


Figure 6 The facial expression change factor (FECF) map of 0838HC and 0508PD based on SEM.

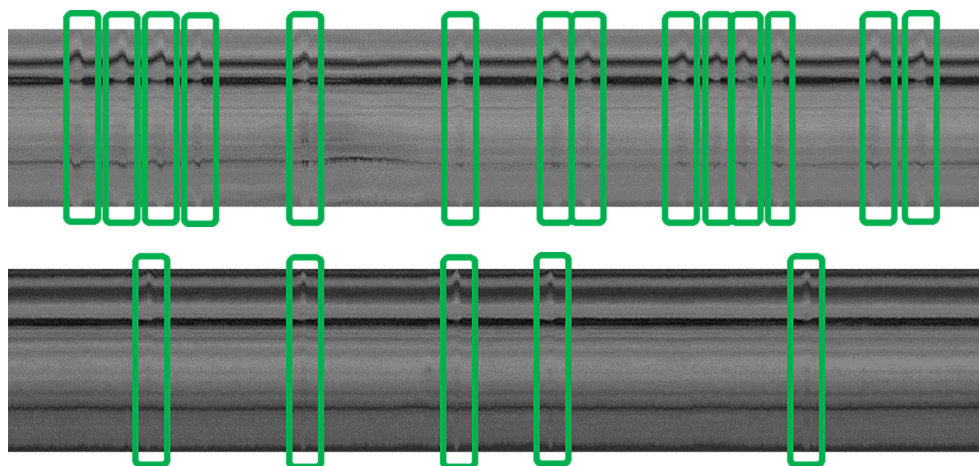


Figure 7 XT plane with the Y-axis coordinate fixed at 40, wherein, the first graph belongs to 0842HC, the second graph belongs to 0451PD.

detecting hypomimia in PD. Both geometric features and texture features were used in this detection system, which enabled the validation of pathological features and unilateral morbidity. In addition, these quantitative features can be used to evaluate the current facial state of Parkinson's disease patients with hypomimia during the rehabilitation program while the mapping module can judge whether patients' facial expressions are natural. Combined with the judgement result, engineered metrics monitor the status of patients in real time, which reflects the effectiveness of the treatment. In particular, Parkinson's hypomimia had a greater impact on the eyes and mouth during the expression of a smile. It should be noted that PD can also be evaluated by other types of expressions, e.g., angry, sadness, etc, and affects other body parts. This current work only focussed on the human face, and many other aspects should be considered in the future, including the limbs, hands, step

states, and even the brain. Further work exploring the relationship between other body parts and PD is warranted.

Acknowledgments

Funding: This work was supported by the National Key Research and Development Program of China (No. 2017YFB1400603), the National Natural Science Foundation of China (Grant No. 61825205, No. 61772459), and the National Science and Technology Major Project of China (No. 50-D36B02-9002-16/19).

Footnote

Reporting Checklist: The authors have completed the STARD reporting checklist. Available at <https://dx.doi.org/10.21037/atm-21-3457>

Data Sharing Statement: Available at <https://dx.doi.org/10.21037/atm-21-3457>

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <https://dx.doi.org/10.21037/atm-21-3457>). Dr. GS, Dr. BL, and Dr. JY report that this work was supported by the National Key Research and Development Program of China (No. 2017YFB1400603), the National Natural Science Foundation of China (Grant No. 61825205, No. 61772459), and the National Science and Technology Major Project of China (No.50-D36B02-9002-16/19). Dr. GS, Dr. BL, Dr. JY, and Dr. WL report patents pending of A Construction Method for Detecting Facial Bradykinesia based on Geometric Features and Texture Features, and report provision of study materials from Second Affiliated Hospital, Zhejiang University School of Medicine. Dr. KD is employed by Technical Department in Hangzhou Healink Technology Corporation Limited, which located at 188 Liyi Rd, Hangzhou, China. The other authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). This study was approved by the Second Affiliated Hospital, Zhejiang University School of Medicine, and the patient informed consent was waived as all data were anonymized.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

- Poewe W, Seppi K, Tanner CM, et al. Parkinson disease. *Nat Rev Dis Primers* 2017;3:17013.
- Sveinbjornsdottir S. The clinical symptoms of Parkinson's disease. *J Neurochem* 2016;139 Suppl 1:318-24.
- Ricciardi L, De Angelis A, Marsili L, et al. Hypomimia in Parkinson's disease: an axial sign responsive to levodopa. *Eur J Neurol* 2020;27:2422-9.
- Ho MW, Chien SH, Lu MK, et al. Impairments in face discrimination and emotion recognition are related to aging and cognitive dysfunctions in Parkinson's disease with dementia. *Sci Rep* 2020;10:4367.
- Ricciardi L, Visco-Comandini F, Erro R, et al. Facial Emotion Recognition and Expression in Parkinson's Disease: An Emotional Mirror Mechanism? *PLoS One* 2017;12:e0169110.
- Jankovic J. Parkinson's disease: clinical features and diagnosis. *J Neurol Neurosurg Psychiatry* 2008;79:368-76.
- Demrozi F, Bacchin R, Tamburin S, et al. Toward a Wearable System for Predicting Freezing of Gait in People Affected by Parkinson's Disease. *IEEE J Biomed Health Inform* 2020;24:2444-51.
- Ghoraani B, Hssayeni MD, Bruack MM, et al. Multilevel Features for Sensor-Based Assessment of Motor Fluctuation in Parkinson's Disease Subjects. *IEEE J Biomed Health Inform* 2020;24:1284-95.
- Lin B, Luo W, Luo Z, et al. Bradykinesia Recognition in Parkinson's Disease via Single RGB Video. *ACM Transactions on Knowledge Discovery from Data (TKDD)* 2020;14:1-19.
- Argaud S, Vérin M, Sauleau P, et al. Facial emotion recognition in Parkinson's disease: A review and new hypotheses. *Mov Disord* 2018;33:554-67.
- Aghanavasi S, Bergquist F, Nyholm D, et al. Motion Sensor-Based Assessment of Parkinson's Disease Motor Symptoms During Leg Agility Tests: Results From Levodopa Challenge. *IEEE J Biomed Health Inform* 2020;24:111-9.
- Bandini A, Orlandi S, Escalante HJ, et al. Analysis of facial expressions in parkinson's disease through video-based automatic methods. *J Neurosci Methods* 2017;281:7-20.
- Ventura MI, Baynes K, Sigvardt KA, et al. Hemispheric asymmetries and prosodic emotion recognition deficits in Parkinson's disease. *Neuropsychologia* 2012;50:1936-45.
- Argaud S, Delplanque S, Houvenaghel JF, et al. Does Facial Amimia Impact the Recognition of Facial Emotions? An EMG Study in Parkinson's Disease. *PLoS One* 2016;11:e0160329.
- Gray HM, Tickle-Degnen L. A meta-analysis of performance on emotion recognition tasks in Parkinson's disease. *Neuropsychology* 2010;24:176-91.
- Kan Y, Kawamura M, Hasegawa Y, et al. Recognition of emotion from facial, prosodic and written verbal stimuli in Parkinson's disease. *Cortex* 2002;38:623-30.

17. McIntosh LG, Mannava S, Camalier CR, et al. Emotion recognition in early Parkinson's disease patients undergoing deep brain stimulation or dopaminergic therapy: a comparison to healthy participants. *Front Aging Neurosci* 2014;6:349.
18. Wood A, Lupyan G, Sherrin S, et al. Altering sensorimotor feedback disrupts visual discrimination of facial expressions. *Psychon Bull Rev* 2016;23:1150-6.
19. Jin B, Qu Y, Zhang L, et al. Diagnosing Parkinson Disease Through Facial Expression Recognition: Video Analysis. *J Med Internet Res* 2020;22:e18697.
20. Seliverstov Y, Diagovchenko D, Kravchenko M, et al. Hypomimia detection with a smartphone camera as a possible self-screening tool for Parkinson disease (P3. 047). *Neurology* 2018.
21. Grammatikopoulou A, Grammalidis N, Bostantjopoulou S, et al. Detecting hypomimia symptoms by selfie photo analysis: for early Parkinson disease detection. *Proceedings of the 12th ACM International Conference on Pervasive Technologies Related to Assistive Environments* 2019;517-22.
22. Emrani S, McGuirk A, Xiao W. Prognosis and diagnosis of Parkinson's disease using multi-task learning. *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 2017;1457-66.
23. Vinokurov N, Arkadir D, Linetsky E, et al. Quantifying hypomimia in parkinson patients using a depth camera. *International Symposium on Pervasive Computing Paradigms for Mental Health*. Springer Cham 2015;63-71.
24. Hoehn MM, Yahr MD. Parkinsonism: onset, progression, and mortality. 1967. *Neurology* 1998;50:318 and 16 pages following.
25. Gibb WR, Lees AJ. The relevance of the Lewy body to the pathogenesis of idiopathic Parkinson's disease. *J Neurol Neurosurg Psychiatry* 1988;51:745-52.
26. Nasreddine ZS, Phillips NA, Bédirian V, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc* 2005;53:695-9.
27. Goetz CG, Tilley BC, Shaftman SR, et al. Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results. *Mov Disord* 2008;23:2129-70.
28. Zhao G, Pietikäinen M. Dynamic texture recognition using local binary patterns with an application to facial expressions. *IEEE Trans Pattern Anal Mach Intell* 2007;29:915-28.
29. Paulmann S, Pell MD. Dynamic emotion processing in Parkinson's disease as a function of channel availability. *J Clin Exp Neuropsychol* 2010;32:822-35.
30. Chen J, Chen Z, Chi Z, et al. Facial expression recognition in video with multiple feature fusion. *IEEE Transactions on Affective Computing* 2016;9:38-50.
31. Marsili L, Agostino R, Bologna M, et al. Bradykinesia of posed smiling and voluntary movement of the lower face in Parkinson's disease. *Parkinsonism Relat Disord* 2014;20:370-5.
32. Wen Y, Zhang K, Li Z, et al. A discriminative feature learning approach for deep face recognition. *European conference on computer vision Springer Cham* 2016;499-515.
33. Sun Y, Liang D, Wang X, et al. Deepid3: Face recognition with very deep neural networks. *arXiv preprint arXiv:1502.00873* 2015.
34. Liu W, Wen Y, Yu Z, et al. Sphereface: Deep hypersphere embedding for face recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition* 2017;212-20.
35. Ding C, Tao D. Robust face recognition via multimodal deep face representation. *IEEE Transactions on Multimedia* 2015;17:2049-58.
36. Yang J, Ren P, Zhang D, et al. Neural aggregation network for video face recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition* 2017;4362-71.
37. Rajnoha M, Mekyska J, Burget R, et al. Towards Identification of Hypomimia in Parkinson's Disease Based on Face Recognition Methods. 2018 10th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT) IEEE 2018;1-4.

(English Language Editor: J. Teoh)

Cite this article as: Su G, Lin B, Yin J, Luo W, Xu R, Xu J, Dong K. Detection of hypomimia in patients with Parkinson's disease via smile videos. *Ann Transl Med* 2021;9(16):1307. doi: 10.21037/atm-21-3457